Probabilistic Graphical Models

Lecture 11: Convex Inference Relaxations. LP Relaxations

Volkan Cevher, Matthias Seeger
Ecole Polytechnique Fédérale de Lausanne

18/11/2011

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

Convex Variational Relaxations

- Inference: Convex variational problem, intractable

Convex Variational Relaxations

- Inference: Convex variational problem, intractable
- Relaxations lead to approximate solutions.
  Tractable, but may be non-convex
  - Structured mean field: $\mathcal{M}_{\mathsf{MF}}$ non-convex
  - Loopy belief propagation: $\mathrm{H}_{\mathsf{Bethe}}[\boldsymbol{\mu}]$ non-convex

Convex Variational Relaxations

- Inference: Convex variational problem, intractable
- Relaxations lead to approximate solutions.
  Tractable, but may be non-convex
  - Structured mean field: $\mathcal{M}_{\mathrm{MF}}$ non-convex
  - Loopy belief propagation: $\mathrm{H}_{\mathrm{Bethe}}[\mu]$ non-convex
- Why convex relaxation?
  - Unique solution. Know when you're done
  - Robust under small data changes (f.ex.: experimental design)
  - Statistical advantages

Wainwright, JMLR 7 (2006)

Convex Variational Relaxations

- Inference: Convex variational problem, intractable
- Relaxations lead to approximate solutions.
  Tractable, but may be non-convex
  - Structured mean field: $\mathcal{M}_{\text{MF}}$ non-convex
  - Loopy belief propagation: $\mathrm{H}_{\text{Bethe}}[\mu]$ non-convex
- Why convex relaxation?
  - Unique solution. Know when you're done
  - Robust under small data changes (f.ex.: experimental design)
  - Statistical advantages

Wainwright, JMLR 7 (2006)

- Exact inference convex. Are convex relaxations also tighter?
  $\Rightarrow$ Unfortunately not ((G)LBP hard to beat)

Convex Relaxations: The Recipe

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

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

$\mathcal{M}$ 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 F2

Convex Relaxations: The Recipe

$$\log Z \leq \max_{\boldsymbol{\tau}\in\mathcal{M}}\left\{\boldsymbol{\theta}^T\boldsymbol{\tau} + \tilde{\mathrm{H}}[\boldsymbol{\tau}]\right\}, \quad \boldsymbol{\mu} = \boldsymbol{\tau}_K,\ \boldsymbol{\theta}_{\setminus K} = \mathbf{0},$$

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

- Recipe for convex relaxation
  1. Concave upper bound $\tilde{\mathrm{H}}[\boldsymbol{\tau}]$ to entropy $\mathrm{H}[\boldsymbol{\mu}]$, tractable function of pseudomarginals $\boldsymbol{\tau}$ (small enough, but $\boldsymbol{\mu} = \boldsymbol{\tau}_K$) F2b

Convex Relaxations: The Recipe

$$\log Z \leq \max_{\boldsymbol{\tau} \in \tilde{\mathcal{M}}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\tau} + \tilde{\mathrm{H}}[\boldsymbol{\tau}] \right\},$$

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

- Recipe for convex relaxation
  1. Concave upper bound $\tilde{\mathrm{H}}[\boldsymbol{\tau}]$ to entropy $\mathrm{H}[\boldsymbol{\mu}]$, tractable function of pseudomarginals $\boldsymbol{\tau}$ (small enough, but $\boldsymbol{\mu} = \boldsymbol{\tau}_K$)
  2. Convex outer bound $\tilde{\mathcal{M}} = \{\boldsymbol{\tau}\} \supset \mathcal{M}$ (f.ex.: $\mathcal{M}_{\text{local}}$) F2c

Convex Relaxations: The Recipe

$$\log Z \leq \max_{\boldsymbol{\tau} \in \tilde{\mathcal{M}}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\tau} + \tilde{\mathrm{H}}[\boldsymbol{\tau}] \right\},$$

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

- Recipe for convex relaxation
  1. Concave upper bound $\tilde{\mathrm{H}}[\boldsymbol{\tau}]$ to entropy $\mathrm{H}[\boldsymbol{\mu}]$, tractable function of pseudomarginals $\boldsymbol{\tau}$ (small enough, but $\boldsymbol{\mu} = \boldsymbol{\tau}_K$)
  2. Convex outer bound $\tilde{\mathcal{M}} = \{\boldsymbol{\tau}\} \supset \mathcal{M}$ (f.ex.: $\mathcal{M}_{\text{local}}$)
- Notes
  - Convex optimization. Upper bound to $\log Z$ (not lower, like MF)
  - $\boldsymbol{\tau}$ can have more components than $\boldsymbol{\mu}$ (clique marginals). Embedding: $\boldsymbol{\mu} = \boldsymbol{\tau}_K$, $\mathrm{H}[\boldsymbol{\tau}] = \mathrm{H}[\boldsymbol{\mu}]$, $\mathcal{M} = \{\boldsymbol{\tau} \,|\, \boldsymbol{\tau}_K = \boldsymbol{\mu} \in \mathcal{M}\}$
  - Must have $\mathrm{H}[\boldsymbol{\tau}] \leq \tilde{\mathrm{H}}[\boldsymbol{\tau}]$ for all $\boldsymbol{\tau} \in \mathcal{M}$ ($\Leftrightarrow \ \boldsymbol{\tau}_K \in \mathcal{M}$)

Examples for Convex Relaxations

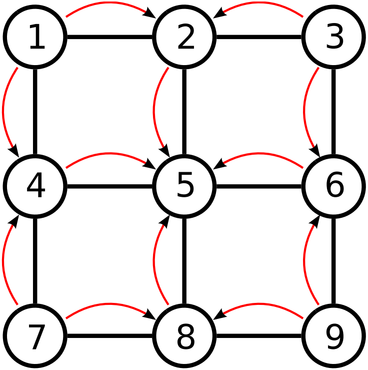
$$\log Z \leq \max_{\boldsymbol{\tau} \in \tilde{\mathcal{M}}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\tau} + \tilde{\mathrm{H}}[\boldsymbol{\tau}] \right\}$$

- Tree-reweighted belief propagation [Wainwright *et.al.*, UAI 02]
  $\Rightarrow$ Exercise sheet
- Conditional entropy decomposition [Globerson *et.al.*, AISTATS 07] F3

$$\mathrm{H}[P(\boldsymbol{x})] = \sum_i \mathrm{H}[P(x_i|\boldsymbol{x}_{<i})] \leq \sum_i \mathrm{H}[P(x_i|\boldsymbol{x}_{S_i})],$$
$$S_i \subset \{1, \dots, i-1\},\ S_i' = S_i \cup \{i\}$$

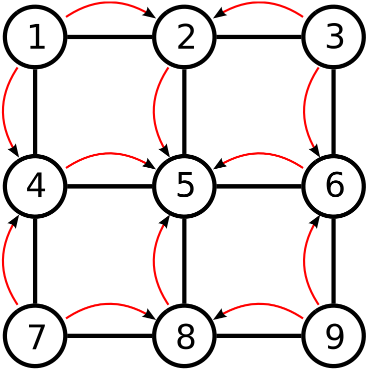
And: $P(\boldsymbol{x}_{S_i'}) \mapsto \mathrm{H}[P(x_i|\boldsymbol{x}_{S_i})] = \mathrm{H}[P(\boldsymbol{x}_{S_i'})] - \mathrm{H}[P(\boldsymbol{x}_{S_i})]$ concave!

Conditional Entropy Decomposition



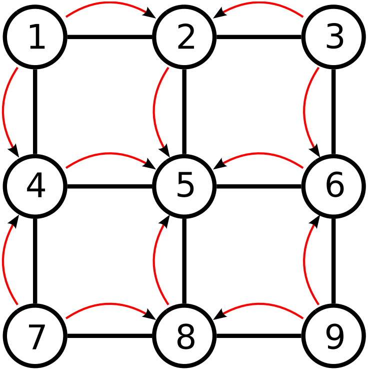
$$\mathrm{H}[\tau] = \mathrm{H}_1 + \mathrm{H}_{3|1} + \mathrm{H}_{7|13} + \mathrm{H}_{9|137} + \mathrm{H}_{2|1379} + \mathrm{H}_{4|13792} + \mathrm{H}_{6|137924} + \mathrm{H}_{8|1379246} + \mathrm{H}_{5|13792468}$$

Conditional Entropy Decomposition



$$\mathrm{H}[\tau] \leq \mathrm{H}_1 + \mathrm{H}_3 + \mathrm{H}_7 + \mathrm{H}_9 + \mathrm{H}_{2|13} + \mathrm{H}_{4|17} + \mathrm{H}_{6|39} + \mathrm{H}_{8|79} + \mathrm{H}_{5|2468}$$

Conditional Entropy Decomposition



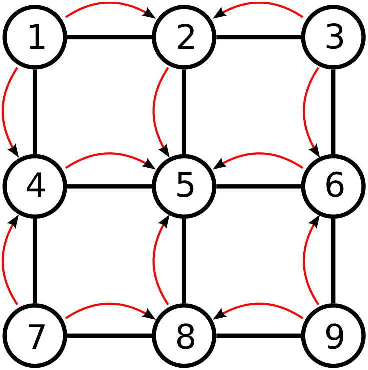
$$\mathrm{H}[\tau] \le \mathrm{H}_1 + \mathrm{H}_3 + \mathrm{H}_7 + \mathrm{H}_9 + \mathrm{H}_{213} - \mathrm{H}_{13} + \mathrm{H}_{417} - \mathrm{H}_{17} + \mathrm{H}_{639} - \mathrm{H}_{39} + \mathrm{H}_{879} - \mathrm{H}_{79} + \mathrm{H}_{52468} - \mathrm{H}_{2468}$$

Here, $\mathrm{H}_S := \mathrm{H}[\tau_S]$. Relaxation:
$\mathrm{H}_3 \ge \mathrm{H}_{3|1},\ \mathrm{H}_7 \ge \mathrm{H}_{7|13},\ \mathrm{H}_9 \ge \mathrm{H}_{9|137},\ \dots$

Index set for τ:

- Nodes, edges (μ)
- $\{13, 17, 39, 79\}$
- $\{213, 417, 639, 879\}$, $\{52468, 2468\}$

Conditional Entropy Decomposition



$$\mathrm{H}[\tau] \le \mathrm{H}_1 + \mathrm{H}_3 + \mathrm{H}_7 + \mathrm{H}_9 + \mathrm{H}_{213} - \mathrm{H}_{13} + \mathrm{H}_{417} - \mathrm{H}_{17} + \mathrm{H}_{639} - \mathrm{H}_{39} + \mathrm{H}_{879} - \mathrm{H}_{79} + \mathrm{H}_{52468} - \mathrm{H}_{2468}$$

Here, $\mathrm{H}_S := \mathrm{H}[\tau_S]$. Relaxation:
$\mathrm{H}_3 \ge \mathrm{H}_{3|1},\ \mathrm{H}_7 \ge \mathrm{H}_{7|13},\ \mathrm{H}_9 \ge \mathrm{H}_{9|137},\ \dots$

Index set for τ:

- Nodes, edges (μ)
- $\{13, 17, 39, 79\}$
- $\{213, 417, 639, 879\}$, $\{52468, 2468\}$

Note: In general ($\neq$ tree), relaxation cannot be avoided

Semidefinite Relaxations

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}, \quad \boldsymbol{x} \in \{0,1\}^m$$

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

- So far: Fencing in $\mathcal{M}$ by linear inequalities (larger polytope).

$$\{\boldsymbol{A} \mid \boldsymbol{A} \succeq \boldsymbol{0} \text{ (pos. semidef.)}\} \quad \Leftrightarrow \quad \{\boldsymbol{A} \mid \operatorname{tr} \boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T \geq 0 \ \ \forall \boldsymbol{x}\}$$

1 SD constraint: ∞ many linear constraints! Still tractable

Semidefinite Relaxations

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}, \quad \boldsymbol{x} \in \{0,1\}^m$$

Binary MRF. Pairwise and single node potentials ($|C_j| \le 2$)

- So far: Fencing in $\mathcal{M}$ by linear inequalities (larger polytope).

$$\{\boldsymbol{A} \mid \boldsymbol{A} \succeq \boldsymbol{0} \text{ (pos. semidef.)}\} \quad \Leftrightarrow \quad \{\boldsymbol{A} \mid \operatorname{tr} \boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T \ge 0 \;\; \forall \boldsymbol{x}\}$$

  1 SD constraint: ∞ many linear constraints! Still tractable

- SD outer bound: If $\boldsymbol{\mu} \in \mathcal{M}$, then

$$\mathrm{E}_Q\left[\begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix} (1 \; \boldsymbol{x}^T) \right] = \boldsymbol{M}_1(\boldsymbol{\mu}) \succeq \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{\mu} \in \mathcal{S}_1 = \{\boldsymbol{M}_1(\boldsymbol{\mu}) \succeq \boldsymbol{0}\}$$

  First-order SD outer bound (convex cone) F10

  - Can use intersection with $\mathcal{M}_{\text{local}}$
  - Higher-order possible ($\boldsymbol{g}(\boldsymbol{x})$ instead of $\boldsymbol{x}$)

Semidefinite Relaxations (II)

$$\mathcal{M} \subset \mathcal{S}_1 = \{\boldsymbol{\mu} \succeq \mathbf{0} \mid \boldsymbol{M}_1(\boldsymbol{\mu}) \succeq \mathbf{0}\}$$

How about the entropy $\mathrm{H}[\boldsymbol{\mu}]$ (need concave upper bound)?

- For continuous $\tilde{\boldsymbol{x}}$: Gaussian maximizes (differential) entropy

$$\mathrm{H}[P(\tilde{\boldsymbol{x}})] = -\int P(\tilde{\boldsymbol{x}}) \log P(\tilde{\boldsymbol{x}})\, d\tilde{\boldsymbol{x}} \leq (1/2) \log |2\pi\, e\, \mathrm{Cov}_P[\boldsymbol{x}]|$$

  [recall lecture 3]. Covariance is part of $\boldsymbol{M}_1(\boldsymbol{\mu})$

Semidefinite Relaxations (II)

$$\mathcal{M} \subset \mathcal{S}_1 = \{\boldsymbol{\mu} \succeq \mathbf{0} \mid \boldsymbol{M}_1(\boldsymbol{\mu}) \succeq \mathbf{0}\}$$

How about the entropy $\mathrm{H}[\boldsymbol{\mu}]$ (need concave upper bound)?

- For continuous $\tilde{\boldsymbol{x}}$: Gaussian maximizes (differential) entropy

$$\mathrm{H}[P(\tilde{\boldsymbol{x}})] = -\int P(\tilde{\boldsymbol{x}}) \log P(\tilde{\boldsymbol{x}})\, d\tilde{\boldsymbol{x}} \leq (1/2) \log |2\pi\, e\, \mathrm{Cov}_P[\boldsymbol{x}]|$$

  [recall lecture 3]. Covariance is part of $\boldsymbol{M}_1(\boldsymbol{\mu})$
- $\boldsymbol{x}$ discrete $\rightarrow$ $\tilde{\boldsymbol{x}}$ continuous by reverse quantization F11

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{u},\ u_i \sim U([-1/2, 1/2]), \quad \underbrace{P(\tilde{\boldsymbol{x}})}_{\text{density}} = \underbrace{P(\boldsymbol{x}(\tilde{\boldsymbol{x}}))}_{\text{distribution}}$$

Semidefinite Relaxations (II)

$$\mathcal{M} \subset \mathcal{S}_1 = \{\boldsymbol{\mu} \succeq \mathbf{0} \mid \boldsymbol{M}_1(\boldsymbol{\mu}) \succeq \mathbf{0}\}$$

How about the entropy $\mathrm{H}[\boldsymbol{\mu}]$ (need concave upper bound)?

- For continuous $\tilde{\boldsymbol{x}}$: Gaussian maximizes (differential) entropy

$$\mathrm{H}[P(\tilde{\boldsymbol{x}})] = -\int P(\tilde{\boldsymbol{x}}) \log P(\tilde{\boldsymbol{x}})\, d\tilde{\boldsymbol{x}} \le (1/2) \log |2\pi\, e \,\mathrm{Cov}_P[\boldsymbol{x}]|$$

  [recall lecture 3]. Covariance is part of $\boldsymbol{M}_1(\boldsymbol{\mu})$

- $\boldsymbol{x}$ discrete $\to$ $\tilde{\boldsymbol{x}}$ continuous by reverse quantization

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{u},\; u_i \sim U([-1/2, 1/2]), \quad \underbrace{P(\tilde{\boldsymbol{x}})}_{\text{density}} = \underbrace{P(\boldsymbol{x}(\tilde{\boldsymbol{x}}))}_{\text{distribution}}$$

- Convex semidefinite inference relaxation — Wainwright, Jordan, IEEE SP 54(6)
  - Tighter than LBP. Better marginal approximations
  - Much more expensive (no local belief propagation solver)

Convexity of (reweighted) Bethe

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

Recall Bethe approximation: Pretend $\mathcal{G}$ was a tree
$\Rightarrow$ When is this convex? Convexification?

Energy term $-\boldsymbol{\theta}^T\boldsymbol{\mu}$? Linear $\rightarrow$ convex

Convexity of (reweighted) Bethe

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

Recall Bethe approximation: Pretend $\mathcal{G}$ was a tree
$\Rightarrow$ When is this convex? Convexification?

Energy term $-\boldsymbol{\theta}^T\boldsymbol{\mu}$?	Linear $\rightarrow$ convex
Constraint set $\mathcal{M}_{\text{local}}$?	Convex polytope

Convexity of (reweighted) Bethe

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

Recall Bethe approximation: Pretend $\mathcal{G}$ was a tree
$\Rightarrow$ When is this convex? Convexification?

Energy term $-\boldsymbol{\theta}^T\boldsymbol{\mu}$?	Linear $\rightarrow$ convex
Constraint set $\mathcal{M}_{\text{local}}$?	Convex polytope
Bethe neg-entropy $-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}]$?	Not convex!

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

Convexity of (reweighted) Bethe

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

Recall Bethe approximation: Pretend $\mathcal{G}$ was a tree
$\Rightarrow$ When is this convex? Convexification?

Energy term $-\boldsymbol{\theta}^T \boldsymbol{\mu}$?	Linear $\rightarrow$ convex
Constraint set $\mathcal{M}_{\text{local}}$?	Convex polytope
Bethe neg-entropy $-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}]$?	Not convex!

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

- Above: Conditional neg-entropy is convex ($i \in C_j$)! F12

$$\mu_j(\boldsymbol{x}_{C_j}) \mapsto -\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j \setminus i}) \,|\, \mu_i(x_i)] = -\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + 1 \cdot \mathrm{H}[\mu_i(x_i)]$$

Convexity of (reweighted) Bethe (II)

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

- Resource allocation game: Assign $\mathrm{H}[\mu_i(x_i)]$ to "free" $-\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]$. Any concave terms left?

Convexity of (reweighted) Bethe (II)

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

- Resource allocation game: Assign $\mathrm{H}[\mu_i(x_i)]$ to "free" $-\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]$. Any concave terms left?
- $\mathcal{G}$ tree: Nothing left. Hey: Bethe relaxation convex for trees!

Convexity of (reweighted) Bethe (II)

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

- Resource allocation game: Assign $\mathrm{H}[\mu_i(x_i)]$ to "free" $-\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]$. Any concave terms left?
- $\mathcal{G}$ tree: Nothing left. Hey: Bethe relaxation convex for trees! Knew that already:
  - Not relaxation, but exact then
  - And exact variational inference is convex

Convexity of (reweighted) Bethe (II)

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum\nolimits_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum\nolimits_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

- Resource allocation game: Assign $\mathrm{H}[\mu_i(x_i)]$ to "free" $-\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]$. Any concave terms left?
- $\mathcal{G}$ has single cycle: Nothing left either [exercise sheet]
  - Bethe relaxation convex (solution usually not correct)
  - LBP always converges

Convexity of (reweighted) Bethe (II)

$$-\mathrm{H}_{\text{Bethe}}[\boldsymbol{\mu}] = \underbrace{-\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]}_{\text{convex (positive)}} + \underbrace{\sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{concave (negative)}}$$

- Resource allocation game: Assign $\mathrm{H}[\mu_i(x_i)]$ to "free" $-\mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})]$. Any concave terms left?
- Convexify Bethe relaxation by decreasing counting numbers n_i until "nothing left" Heskes, JAIR 26 (2006)
  - Reweighted Bethe relaxations solved by reweighted LBP
  - Somewhat lacks interpretability
  - TRW-BP (last exercise sheet) example with additional motivation

LP Relaxations

What about the mode (decoding)? Integer programming: F14

$$\text{MAP}: \ \boldsymbol{x}_* = \operatorname{argmax} \log \left[Z^{-1} e^{\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})} \right] = \operatorname{argmax} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})$$

LP Relaxations

What about the mode (decoding)? Integer programming:

$$\mathrm{MAP}: \ \boldsymbol{x}_* = \operatorname{argmax} \log \left[Z^{-1} e^{\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})} \right] = \operatorname{argmax} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})$$

Linear criterion $\Rightarrow$ Linear programming transformation

$$\max_{\boldsymbol{x}} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x}) = \max_{Q(\boldsymbol{x})} \mathrm{E}_Q[\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})] = \max_{\boldsymbol{\mu} \in \mathcal{M}} \boldsymbol{\theta}^T \boldsymbol{\mu}$$

$\Rightarrow$ Variational inference without entropy term

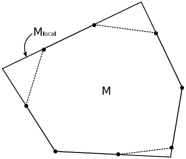
LP Relaxations

What about the mode (decoding)? Integer programming:

$$\text{MAP}: \ \boldsymbol{x}_* = \operatorname{argmax} \log \left[Z^{-1} e^{\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})} \right] = \operatorname{argmax} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})$$

Linear criterion $\Rightarrow$ Linear programming transformation

$$\max_{\boldsymbol{x}} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x}) = \max_{Q(\boldsymbol{x})} \mathrm{E}_Q[\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})] = \max_{\boldsymbol{\mu} \in \mathcal{M}} \boldsymbol{\theta}^T \boldsymbol{\mu}$$



$\Rightarrow$ Variational inference without entropy term

- LP relaxations by outer bounds: $\mathcal{M} \subset \mathcal{M}_{\text{local}}$
  $\Rightarrow$ Result always a tractable LP.
  Catch: $\mathcal{M}_{\text{local}}$ too many extreme points $\Rightarrow$ Fractional solutions

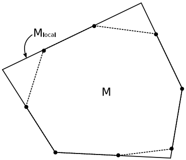
F14b

LP Relaxations

What about the mode (decoding)? Integer programming:

$$\mathrm{MAP}: \; \boldsymbol{x}_* = \operatorname{argmax} \log \left[Z^{-1} e^{\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})} \right] = \operatorname{argmax} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})$$

Linear criterion ⇒ Linear programming transformation



$$\max_{\boldsymbol{x}} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x}) = \max_{Q(\boldsymbol{x})} \mathrm{E}_Q[\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})] = \max_{\boldsymbol{\mu} \in \mathcal{M}} \boldsymbol{\theta}^T \boldsymbol{\mu}$$

⇒ Variational inference without entropy term

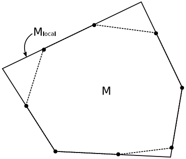
- LP relaxations by outer bounds: $\mathcal{M} \subset \mathcal{M}_{\text{local}}$
  ⇒ Result always a tractable LP.
  Catch: $\mathcal{M}_{\text{local}}$ too many extreme points ⇒ Fractional solutions
- Key question: Does max-product solve this LP?
  - For tree: Yes, and MP is a dual algorithm
  - In general: No.
    But (carefully) reweighted max-product does

LP Relaxations

What about the mode (decoding)? Integer programming:

$$\mathrm{MAP}: \ \boldsymbol{x}_* = \operatorname{argmax} \log \left[Z^{-1} e^{\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})} \right] = \operatorname{argmax} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})$$

Linear criterion ⇒ Linear programming transformation



$$\max_{\boldsymbol{x}} \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x}) = \max_{Q(\boldsymbol{x})} \mathrm{E}_Q[\boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x})] = \max_{\boldsymbol{\mu} \in \mathcal{M}} \boldsymbol{\theta}^T \boldsymbol{\mu}$$

⇒ Variational inference without entropy term

- LP relaxations by outer bounds: $\mathcal{M} \subset \mathcal{M}_{\text{local}}$
  ⇒ Result always a tractable LP.
  Catch: $\mathcal{M}_{\text{local}}$ too many extreme points ⇒ Fractional solutions
- Why should I care?
  - For several NP-hard problems: $\mathcal{M}_{\text{local}}$ "first order" LP relaxation frequently used
  - Reweighted MP: Fastest known algorithm in some cases

Wrap-Up

- Convex variational relaxations: Concave upper bound to entropy, convex outer bound to marginal polytope
- Bethe problem can be convex (graph with ≤ 1 cycle). It can be convexified
- First-order LP relaxations by fast convergent reweighted max-product