**Marginal Polytope**

**Introduction**

- Want to compute:

\[ p(x_1, \ldots, x_5) = \sum_{x_1, x_2, x_3, x_4, x_5} p(x_1, x_2, x_3, x_4, x_5) \]

- Naïve method expensive: use good independence relationships to compute efficiently

**Sum Product Algorithm**

Message from factor node to variable node:

\[ M_{i \rightarrow a}(x_i) = \prod_{c \in \mathcal{N}(i) \setminus a} M_{c \rightarrow i}(x_c) \]

- \[ \sum \prod_{c \in \mathcal{N}(i) \setminus a} M_{c \rightarrow i}(x_c) \]

\[ p(x_i) = \frac{1}{Z_a} \sum_{x_a} f(x_a) \prod_{i \in \mathcal{E}(a)} M_{i \rightarrow a}(x_i) \]

\[ p(x_a) = \frac{1}{Z_a} \sum_{x_a} f(x_a) \prod_{i \in \mathcal{E}(a)} M_{i \rightarrow a}(x_i) \]
Loopy BP

- WHAT SHOULD WE DO WHEN THE GRAPH HAS LOOPS

**IDEA 1**: Anonymize variables until the graph has no loops & then run BP (JUNCTION TREE) (SLOW)

**IDEA 2**: Run BP & stop when converged. Empirically produces a good solution in loopy (but not densely connected graphs)

Why??

**FREE ENERGIES (FROM STATISTICAL PHYSICS)**

\[ p(x|\theta) = \frac{1}{Z(\theta)} \exp(-E(x, \theta)) \]

\[ \log Z(\theta) \geq -\langle E(x, \theta) \rangle_{q(x)} + H[q(x)] \]

with equality @

\[ q(x) = p(x|\theta) \]

1) Think about approximating \( A(\theta) = \log Z(\theta) \)

by approximating the Gibbs free energy & the form of optimisation \( q(x) \).

\( \Rightarrow \) Turns a marginalisation problem into an easy optimisation

**BP FOR TREES**

\[ p(x) = \prod_{v \in E} p(x_v) \prod_{i,j \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \]

\[ q(x|\theta) = \sum_{x_{\bar{i}}} \sum_{x_i} q(x_i) \prod_{i=1}^{M} \log \frac{f_i(x_i)}{q_i(x_i)} \]

\[ + \sum_{i,j \in E} q(x_i, x_j) \log \frac{q_i(x_i)q_j(x_j)}{q(x_i, x_j)} \]

\[ - \sum_{i \in E} q_i(x_i) \log q_i(x_i) + H(x) \]

\[ - \sum_{i,j \in E} q_i(x_i) \log q_i(x_i) \log q_j(x_i), q_i(x_i), q_j(x_i) \]
Some Facts

- Can do the same thing for loopy graphs

- Expression for the entropy now approximate (Bethe) no longer and lower (upper bound on $\sum q(x_i)$)

- Loopy BP fixed points = Local stationary point of the free energy

- Gibbs free energy was convex in \( q(x) \) but Bethe free energy is non-convex in \( q(x_i), q(x_i, x_j) \)

- Mean field, \( q(x) = \prod_i q(x_i) \), also a free energy optimise

\[
q(x) = \arg \max_q \sum_i q(x_i) \] \[
q(x_i) = \prod_i q(x_i)
\]

Also a non-convex optimisation

But mean field does return an upper bound on \( -\log Z(\theta) = A(\theta) \)

Maximise under constraints (Lagrangian)

= Recover BP updates

= Messages correspond to exponentiated Lagrange multipliers

\Rightarrow Convex optimisation for trees

Constraints

\[
\sum_{x_i, x_j} q(x_i, x_j) = 1
\]

\[
\sum_{x_i} q(x_i) = 1
\]

\[
\sum_{x_j} q(x_i, x_j) = q_1(x_i)
\]

\[
q(x_i) \geq 0 \quad q(x_i, x_j) \geq 0
\]

Optimise free energy & constraints (Lagrangian)

\Rightarrow Recover BP updates

\Rightarrow Messages correspond to exponentiated Lagrange multipliers

\Rightarrow Convex optimisation for trees

However
OUTLINE

- GOAL: USE ALL THE MACHINERY OF CONVEX ANALYSIS IN THESE PROBLEMS
  EXACT
  LOOP BP / MEAN FIELD
  TRW
  CONVEX
  NON CONVEX
  TRACTABLE
  INTRACTABLE
  NO BOUND / BOUND Z18
  BOUND Z18

1) Machinery: Exponential Families
   Mean Parameters & the marginal polytope
   Variational Formulation & Dual
   BP in this picture
   MF
   TRW Sum Product

---

EXPERIMENTAL FAMILY BASICS

\[ p(x|\theta) = \exp \left( \sum_i \phi_i(x) \theta_i - A(\theta) \right) \]

Sufficient statistics parameters \( \log Z(\theta) \)

Mean Parameters:
\[ M_i = \sum_x p(x|\theta) \phi_i(x); \quad A(\theta) = d \frac{A(\theta)}{d\theta_i} \]

Proof:
\[ \frac{d}{d\theta_i} \ln \sum_x \exp \left( \sum \phi_i(x) \theta_i \right) = \sum_x \phi_i(x) e^{\phi_i(x) \theta_i} \sum_x e^{\phi_i(x) \theta_i}; \]
\[ = \langle \phi_i(x) \rangle \]

\[ \frac{d^2}{d\theta_id\theta_j} A(\theta) = \langle \phi_i(x) \phi_j(x) \rangle - \langle \phi_i(x) \rangle \langle \phi_j(x) \rangle \geq 0 \]

\[ \Rightarrow A(\theta) \text{ is convex in the parameters} \]

Minimal REP vs Over-complete REP.
**Examples**

**Ising Model**

\[ p(x|\theta) = \exp\left( \sum_i x_i \theta_i + \sum_{i,j} \theta_{ij} x_i x_j - A|\theta| \right) \]

\[ M = \sum_x p(x|\theta) \varnothing(x) \]

Trace the set of allowable mean parameters as the distribution \( p(x|\theta) \) varies.

**Convex Set**

Because convex combinations can then generate interior.

\[ M(n+1) = \sum_x p(x|\theta) \varnothing(x) \quad M(n+2) = \sum_x p(x|\theta) \varnothing(x) \]

\[ M_0 = \sum_x \left( \lambda p^{(1)}(x \theta) + (1-\lambda) p^{(2)}(x \theta) \right) \varnothing(x) \]

\[ = \lambda M^{(1)} + (1-\lambda) M^{(2)} \]
MARGINAL POLYTOPE EXAMPLES

1D GAUSSIAN

\[ m_1 = <x_1> \quad m_2 = <x_2> \]

\[ <x_2^2> - <x_2>^2 \geq 0 \]

\[ m_1 \geq m_2 \]

2 NODE ISING MODEL

\[ m_1 = <x_1> \quad m_2 = <x_2> \quad m_{12} = <x_1 x_2> \]

\[ \begin{array}{c|c|c}
   m_1 & m_2 & m_{12} \\
   \hline
   0 & 0 & 0 \\
   1 & 0 & 0 \\
   0 & 1 & 0 \\
   1 & 1 & 1 \\
\end{array} \]

\[ 0 \leq m_{12} \leq 2m_1 m_2 \quad 1 + m_1 m_2 - m_{12} \leq 0 \]

VARIATIONAL PROBLEM & DUAL

PRIMAL

\[ A(\theta) = \sup_{m \in M} \left( \sum_i \theta_i m_i - A^*(m) \right) \] (Free energy)

Supremum achieved when \( m = <\phi(x)>p(x|\theta) \)

DUAL

\[ A^*(m) = \sup_{\theta \in \Theta} \left( \sum_i \theta_i m_i - A(\theta) \right) \]

Supremum achieved when \( \theta \) set s.t. \( m = <\phi(x)>p(x|\theta) \)

When \( A^*(m) = -H[p(x|\theta)] \)

\[ m \in M \]

Otherwise \( A^*(m) = +\infty \)
**Example of This**

\[ p(x|\theta) = \pi^x (1-\pi)^{1-x}, \quad x \in \{0, 1\} \]

\[ = \exp(\theta x - A(\theta)) \quad (\theta = \log \frac{\pi}{1-\pi}) \]

\[ A(\theta) = \log \sum_x \exp(\theta x) = \log(1 + e^\theta) \]

\[ A^*(\mu) = \sup_{\theta \in \mathbb{R}} \left[ \theta \mu - \log(1 + e^\theta) \right] \]

\[ \frac{dA^*(\mu)}{d\theta} = \mu - \frac{e^\theta}{1 + e^\theta} \Rightarrow \begin{cases} \mu = \frac{1}{1 + e^{-\theta}} & \text{for} \ \theta > 0 \\ \mu = \frac{1}{1 + e^{\theta}} & \text{for} \ \theta < 0 \end{cases} \]

\[ A^*(\mu) = \begin{cases} \mu \log \frac{\mu}{1-\mu} - \log (1+\frac{\mu}{1-\mu}) & \mu > 0 \\ \mu \log \frac{\mu}{1-\mu} - \log (1+\frac{1}{1-\mu}) = -H[p(x|\theta(\mu))] & \mu < 0 \end{cases} \]

Can also verify:

\[ A(\theta) = \max_{\mu \in [0, 1]} \left[ \theta \mu - \mu \log \mu - (1-\mu) \log(1-\mu) \right] \]

\[ = \log(1 + e^\theta) \circ \mu = \frac{1}{1 + e^{-\theta}} \]
Summary

- **THEORETICALLY**
  - WE CAN COMPUTE BOTH \( A(b) = \log 2(b) \) & \( H[p(x|b)] \)

**PROBLEM: LOG PARTITION FUNCTION**

**Entropy**

**USING VACATIONAL OPTIMISATION (CONVEX)**

- **PRACTICALLY**
  - COMPUTATIONALLY INTRACTABLE

1) Marginal polytope \( M \) very difficult to characterize

2) Negative entropy lacks an explicit form

**SOLUTION**

- Consider approximation to \( M \) & \( H[p(x|b)] \)

---

**LOOPY BP & MARGINAL POLYTOPE**

- **MARGINAL POLYTOPE COMPLEX**

\[
\begin{align*}
\sum_{x_1, x_2} p(x_1, x_2, x_3) &= p(x_3) & \sum_{x_1} p(x_1, x_2) &= p(x_2) \\
&= p(x_2)
\end{align*}
\]

**JUST ENFORCE LOCAL PAIRWISE CONSTRAINTS**

\[
\begin{align*}
\sum_{x_1, x_2} p(x_1, x_2) &= p(x_2) \\
&= p(x_2)
\end{align*}
\]

**REDUCED SET OF LINEAR CONSTRAINTS ON MEAN PARAMETERS**

- **CONTRAINTS - STILL CONVEX**

**APPROXIMATE THE ENTROPY USING THE GROTHE ENTROPY**

**OBJECTIVE - NO LONGER**
Problems with Loopy BP

- Unrealisability of beliefs
- Negative Entropies (Conjecture @ MSD $H_{BEPG}$)
- Non Convergence?
  - No upper bound

Or lower

Mean Field & Maximal Polytope

- Restrict $g(x)$ to be fully factored $g(x) = \prod_i g_i(x_i)$
- Objective identical $\Rightarrow$ still convex
- Induces complex dependencies between the mean parameter

Eq. for Ising model $M_1 = \langle x_i \rangle$, $M_2 = \langle x_i^2 \rangle$

$M_2 - \langle x_i \rangle \langle x_i \rangle = \langle x_i \rangle \langle x_i \rangle$

Non convex constraint

Schematically

Diagram showing extremal values (dashed) and realizable values (solid). The set is non-convex.

Realizable

We know some mean parameters are not realizable

Mean field will have more vertices than facets
Tree Reweighted BP

Consider Ising model: \( p(x; \theta) = \exp\left( \sum_i \theta_i x_i + \sum_{ij} \theta_{ij} x_i x_j - A(\theta) \right) \)

1. A spanning tree provides an upper bound on the entropy:

\[
-H[\mathcal{T}] = \sup_{M \text{ rooted to tree}} (M \theta - A(\theta)) \leq \sup_M (M \theta - A(\theta)) = -H[M]
\]

\[\Rightarrow H[\mathcal{T}] \geq H[M]\]

2. Any convex combination of spanning trees also upper bounds \( H[M] \):

\[
\sum_T \rho_T H[\mathcal{T}] \geq H[M]
\]

\[
\sum_T \rho_T = 1
\]

Convex objective that just depends on marginal pairwise "mean" parameters.

Example

Spanning trees

\[
T_1 \quad p_1 = \frac{1}{4} \quad T_2 \quad p_2 = \frac{1}{4} \quad T_3 \quad p_3 = \frac{1}{4} \quad T_4 \quad p_4 = \frac{1}{4}
\]

So:

\[\Rightarrow p_3 = \frac{1}{2} \quad p_6 = 1\]

3. Use local consistency of pairwise/marginal beliefs to ensure consistency (like BP):

\[
\gamma(\theta, \rho) = \sup_M \left\{ \sum_T \theta_T + \sum_T \rho_T H(\mathcal{T}) \right\}
\]

If all spanning trees are used in the approximation.

4. \( H(\mathcal{T}) = \sum_{S \in V} H_s(x_S) - \sum_{S \in E} I_{S,T}(M_{\mathcal{T}}) \) for one tree

\[\Rightarrow H(M) \leq \sum_{S \in V} H_s(M_S) - \sum_{S \in E} I_{S,T}(M_{\mathcal{T}}) \] for convex comb.

NB fractional BP
COMMENTS ON TRW LBP

\[ q^{\text{TRW}}(\theta, \rho) = \max_{\theta \in \Theta} \left( \sum_{i} \sum_{s \in \mathcal{V}} \mathcal{E}_{i} + \sum_{s \in \mathcal{V}} H_s(s) \right) + \sum_{s \in \mathcal{E}} \rho_{st} \mathcal{E}_{s} \mathcal{E}_{t} \] - fractional BP forces

1. CONVEX IN \[ \theta \] (NB. \[ \rho \] st embodies constant of the pdf on Entropy & AIB)

2. UPPER BOUND ON \( \mathcal{E} \) - relaxing leads to non-convex problem

3. CAN MINIMIZE \( q^{\text{TRW}}(\theta, \rho) \) WRT. \( \rho \) TO GET TIGHTEST APPROXIMATION

\[ \Rightarrow \text{Hard to do exactly} \Rightarrow \text{Provide greedy, convex method} \]

- HMC/HC on chunks defined by the length-scale of the GP.

- Liam Paninski: invert of banded matrices fast

- Mean paper network (talk to John Cunningham)

- To check whether John can do any of this — we suspect not. Do he cannot (see his paper)

- SSOE \rightarrow Simultaneous diagonalization

- Friday 12th 1:30

\[ \text{11am - Feb} \]

\[ \text{Monash} \]