Constantine Cararmanias

Ph.D. Thesis of M. Winwright

Graphs with Cycles:
When can we solve this problem?

Why is this hard?

Given a distribution \( p(x) \) on \( \mathbb{R}^N \) variables, find marginals.

Main Problem and Approach
• How can we extend it further?

• How far does the idea extend?

• Why is it easy?

Trees: Marginalization is Easy
Outline

1. Geometry and Convex Analysis,

2. Reparameterization,

3. Extensions to Hypergraphs and Node Clustering,

Local Constraints $\Leftrightarrow$ Global Constraints.

Consequence of Tree Factorization:

$$\left(\frac{_{\exists} (s_r)^d}{_{\forall} (s_r)^d} \prod_{s_r} \right) (x)^s d \prod = (x)d$$

Tree Factorization:

$$\bigcup_{x \in \mathbb{R}} \prod_{s_r} (x)^s d = (x)d$$

Hammerly-Cillford Theorem:

Factorization
• Overcomplete Families: Also will be useful.

• Minimal Families: Linearly independent.

\[
\log \left( \left\{ (x)^{\phi \circ \theta} \right\}_{\text{exp } x} \right) = (\theta) \Phi
\]

\[
\left\{ (\theta) \Phi - (x)^{\phi \circ \theta} \right\}_{\text{exp } x} = (\theta'; x) d
\]

Exponential Families and Riemannian Geometry
Therefore \( \Phi(\theta) \) is convex.

\[
\begin{align*}
\left(\left[\left[\pmb{d}\phi\right]_{\theta} - \pmb{d}\phi\right]\left[\pmb{r}\phi\right]_{\theta} - \pmb{r}\phi\right]\right)_{\theta} = \{\pmb{d}\phi, \pmb{r}\phi\}_{\theta} \cap \Phi = (\theta)\frac{\pmb{d}\theta \pmb{r}\theta}{\Phi\theta}
\end{align*}
\]

\[
\{\phi\}_{\theta} = (\theta)\frac{\pmb{r}\theta \Phi}{\Phi\theta}
\]

\[
\left((\{x\} \pmb{r}\phi \left. \frac{\partial}{\partial \phi} \right|_{\phi}| x \right) \exp \{ x \}ight)_{\theta} = (\theta)\Phi
\]

Key properties:

Log Partition Function
\[
[x \phi]_{\theta} = \frac{x \theta \epsilon}{(\theta) \Phi \epsilon} = x[(\theta) \nu]
\]

This gives a map \( u \leftarrow \theta \) to dual coordinates:

\[
(d) H = (\theta) x u = x u
\]

and thus

\[
[x \phi]_{\theta} = (\theta) x u = x u
\]

By convexity, we have stationary conditions for optimunm:

\[
(x \phi - \theta L u)^{\theta} = (u) \nabla
\]

We define the Legendre dual of the log partition function:

**Legendre Transform and Duality**
\[(\theta : I = t_s x \cdot I = s x)d = [t_s x]_{\theta \text{d}} = t_s u \]
\[(\theta : I = s x)d = [s x]_{\theta \text{d}} = s u \]

The dual variables are given by:

\[\mathcal{H} \in (t, s) = \alpha \quad t_s x \quad s \quad s \quad x \quad x \quad \} = (x)^{x \phi} \]

The potential functions are

\[\{(\theta) \Phi - t_s x s x \theta \mathcal{H} \in (s, s) \bar{\mathcal{H}} + s x s \theta \Lambda \in (s, s) \bar{\mathcal{H}}\} \exp = (\theta : x)d \]

Example: Ising Model
\{0 = \mu_B \mid (\theta' x) d\} = \mathcal{M}

• An m-flat manifold is the image of an affine set of \(\mu\):

\{q = \theta \forall x \mid (\theta' x) d\} = \mathcal{M}

• An e-flat manifold is the image of an affine set of \(\theta\):

**Flat Manifolds**
is convex in second argument:

\[ (*\mu)\Phi - (\mu)\Phi + (\mu - *\mu) \mathcal{L}(\Phi) = (\theta)\Phi - (*\theta)\Phi + (*\theta - \theta) \mathcal{L}\mu = (*\theta||\theta)\mathcal{D} \]

becomes using the exponential and mean parameters,

\[ \left[(\star\theta, x) \mathcal{D} \log - (\theta, x) \mathcal{D} \log\right] \mathcal{D} = (*\theta||\theta)\mathcal{D} \]

Kullback–Leibler divergence:

I–Projections onto Flat Manifolds
\[ 0 = [\theta - \theta] L [\text{u} - * \text{u}] \]

Gradient condition: \( \star \text{u} - \text{u} = (\theta \| \star \theta) A^\theta \Delta \)

Convex objective, subject to linear constraints.

\( s.t. \quad \theta \in \mathcal{M}^a \)
\( (\theta \| \star \theta) \min^a D^\theta \)

\(-I\)-projection onto an \(-I\) flat manifold \( \mathcal{M}^a \)

\(-I\)-projections onto \(-I\) flat manifolds (cont’d)
\[(\theta \| \theta) a + (\theta \| \ast \theta) a = (\theta \| \ast \theta) a\]

Also, have Pythagorean relation:

\[0 = [\theta - \theta]_L [\nu - \ast \nu] = (\theta)^D \langle [\nu - \ast \nu] (\nu)_I - D, [\theta - \theta] \rangle\]

Meets orthogonally:

\[(\nu - \ast \nu) \ast + \nu) = (\ast \theta)\]

Consider an m-geodesic joining and \(\ast \theta\) and \(\theta\) \(\ast \theta\) m-geodesic joining and \(\theta\) and \(\ast \theta\)

I-Projections onto Flat Manifolds (cont'd)
\[
(\exists x \forall s x) \text{iff } \prod (\forall x \exists s x) \text{iff } \Lambda \exists s \prod Z \frac{1}{1} = (x) d
\]
\[
\frac{\forall x \exists s x \forall s d}{\forall x \forall s x \forall s d} \bigwedge \exists (s x) \forall s d \bigwedge = (x) d
\]
By Hammerley–Clifford we have the factorization:

\[
\mathcal{H}(t^s) \prod (s_x)^{s_{\phi}} \Lambda_{\Xi^s} \frac{Z}{I} = (x)d
\]

Loopy Graph Reparameterization
\[
\begin{align*}
\left( x \right)_{\mathcal{L}'} \left( x \right)_{\mathcal{L}} \frac{dz}{\Gamma} = \\
\prod_{i=1}^{\mathcal{L}'} \left( s x \right)^{s \phi} \prod_{9}^{\mathcal{L}} \left( s x \right)^{s \phi} \frac{Z}{\Gamma} = (x)d
\end{align*}
\]
\[ \text{and therefore, we have:} \]

\[ \prod_{s L} \prod_{I=s}^{\infty} (s x)^{s \phi} \prod_{9} (s x)^{s \phi} \prod_{9} = (x) \prod_{I}^{\infty} \]

Factor into Pseudomarginals
Repeat.
Next step: Find different subgraph with tree structure, and

Factor Tree into Pseudomarginals
\[(\gamma = \mathcal{E}(x)) \mathcal{G}(\mathcal{F} = s x) \theta_{\mathcal{H}} = \gamma_{s \mathcal{E} t} d = \gamma_{s \mathcal{E} t} u \]

\[(\mathcal{F} = s x) \mathcal{G}(\mathcal{E}) = s^t \mathcal{F} = s^t \mathcal{H} = s^t u \]

We have mean (dual) parameters:

\[(\gamma \mathcal{E}(s \mathcal{F} t)) = a \quad \text{for } \gamma = \mathcal{E}(x) \mathcal{G}(\mathcal{F} = s x) \mathcal{G} = (x)^{a \phi} \]

\[(\mathcal{F} \mathcal{E}(s \mathcal{F} t)) = a \quad \text{for } \gamma = \mathcal{E}(s x) \mathcal{G} = (x)^{a \phi} \]

Then we can define:

\[\{ w > \gamma, \mathcal{F} \supset \mathcal{H}, \mathcal{F} \supset \mathcal{E}, \exists \mathcal{E} \mathcal{F} t \mathcal{S}, t \in (s, t) (s, t) \mathcal{G} \} = \mathcal{A} \]

We choose an overcomplete family:

Exponential Families
is an $e$-flat manifold.

\[
\{(0_\theta; \mathbf{x})_d = (\theta'; \mathbf{x})_d | \theta\} =: (0_\theta)(\mathcal{N})
\]

Given the distribution, \((0_\theta; \mathbf{x})_d(0_\theta; \mathbf{x})_d\), then

\[
\text{We will see this is easy for trees.}
\]

\[(\theta)(\mathcal{V}) = \mathcal{U}
\]

\text{We therefore want to compute:}

\text{Some Remarks}
therefore: Reparameterizing according to a tree correspond
to using the pseudomarginals \( \{ \frac{g}{\xi} u, \frac{g}{\xi} L \} \) we can compute it for graph structures.

Geometric Interpretation

We cannot compute \( v(\theta) \) exactly for loopy graphs.
\{x \in (E, s) \mid \bar{L} = \gamma_{\bar{L}}(\theta)p(0) \in \mathcal{L}_{| \mathcal{L}} \} =: \mathcal{G}_{\gamma} \text{ tree must be in the set:}

\text{The pseudo-marginals obtained from reparametrizing w.r.t.:}

\{\mathcal{H} \in (E, s) \mid \bar{L} = \gamma_{\bar{L}}(\theta)p(0) \in \mathcal{L}_{| \mathcal{L}} \} =: \mathcal{G} \text{ the marginals must satisfy local consistency constraints:}

\text{Geometric Interpretation}
that manifold, to the m-flat manifold EXPREE\textsuperscript{T}(G).

Reparameterizing according to tree $\mathcal{T}$, moves within the e-

distribution. Therefore we stay within the e-flat manifold.

EXPRE\textsuperscript{T}(G) and therefore EXPREE\textsuperscript{T}(G) are m-flat.

E-flat and M-flat Manifolds
Moving in E-Flat Manifold to M-Flat Manifolds
\[
\begin{align*}
\text{proj}_{\Lambda \Theta}(\theta^s) \\
\text{proj}_{\Lambda \Theta}(\theta^s) A \\
+ \text{proj}_{\Lambda \Theta}(\theta^s) B \\
= ((\theta) \forall \Theta) (\theta^s \forall \Theta) C \\
\end{align*}
\]

**Projection Interpretation.** The point in EXPRE\_TREE\_\(\Theta\) to which TRP moves has a

Projections onto M-Fat Manifolds
Like the Bethe approximation, $\mathcal{L}$ approximates $KL$ divergence.

- The Bethe approximations are not equivalent.
- For the full graph $G$ with cycles, the $KL$ divergence and $\mathcal{L}$ are not equivalent.
- Over the tree, minimizing $\mathcal{G}((\theta),\mathcal{L})$ can be seen to be equivalent.
- Minimizing $\mathcal{G}((\theta),\mathcal{L})$ yields correct marginals for the tree.
- Minimizing the $KL$ divergence $D(\mathcal{L} || ((\theta)\mathcal{L})) \Theta$ subject to $\mathcal{E}XPTREE$ yields correct marginals for the tree.

Variation Functional
\[ (u\theta \cdot ((I+u\theta) \nabla) \delta u) + ((I+u\theta \cdot \Omega) \delta = (I+u\theta \cdot \Omega) \delta \]

**Functional $\delta$**

Free Reparameterization is a projection with respect to the

**Variational Formulation**
34
5
6
7
8
9

\[ \forall [\mathbf{L} - \mathbf{1}] \theta \in \mathfrak{L} \mathfrak{e} \mathfrak{r} \mathfrak{e}(\mathfrak{c}) \mathfrak{r} \text{ that vector } \mathbf{L} \text{ satisfies stationary conditions w.r.t. } \mathfrak{g}: \]

\[ \frac{\partial \mathbf{L}}{\partial \mathfrak{g}} \exists \mathfrak{r} \mathfrak{e}(\mathfrak{c}) \mathfrak{r} \text{ A fixed point } \theta \text{ corresponds to a pseudo-marginal vector } \mathbf{L}. \]

\[ \text{Fixed Point Characterization} \]
A Blackboard Detour: Extensions to Hypergraphs
\[
\prod_{\mathcal{L}} \left( (\mathcal{L})^\theta \Phi (\mathcal{L})^{\eta'} \right) = \left( \left( (\mathcal{L})^\theta \Phi \right)^{\eta'} \right) \supseteq (\ast \theta) \Phi
\]

By convexity of \((\theta)^\Phi\), and Jensen's inequality:

\[
\{ \ast \theta = \left( (\mathcal{L})^\theta \right)^{\eta'} | (\eta', \theta) \} =: (\ast \theta) \forall
\]

\[
\{ z \in \mathcal{L} | (\mathcal{L})^\theta \} = \theta
\]

Bounds on Partition Function
\[
\{\eta \gamma + 1 \geq \eta \gamma + \eta \gamma \geq \eta \gamma \geq 0 \mid \gamma \} = (\mathbb{C})_\Pi
\]

where

\[
\eta \gamma + \max_{\gamma \in (\mathbb{C})_\Pi} \left\{ \left[ \left( \gamma, \mu \right) \Phi \right] \eta \right\}
\]

From the dual, via the Lagrangian:

\[
\cdot \gamma = \left[ \eta \right] \eta \mathbb{E} \quad \text{s.t. } \left\{ \left[ \left( \gamma, \mu \right) \Phi \right] \eta \right\}
\]

Fixing \( \eta \), we have a convex optimization problem:

\[
\text{Optimizing over Parameter } \gamma
\]
\[ \text{H} \in \{ t's \} \cdot (\chi)^{s} I^{s} \sum + (\chi)^{s} H \sum - = \left\{ (\chi)^{s} I \sum + (\chi)^{s} H \sum - \right\} (\ell)^{l} \sum = [((\chi)_{\ell}^{\sum})_{\eta}]^{l} \sum \]

Therefore, the expectation becomes

\[ \text{H} \in \{ t's \} \cdot (\chi)^{s} I \sum + (\chi)^{s} H \sum - = ((\chi)_{\ell}^{\sum})_{\eta} \]

We can decompose entropy over a tree distribution:

Optimizing Over Parameter $\eta$
\[(\star \theta, \varepsilon \eta \lambda)^H (\mathcal{G}) \mathbb{G} \text{argmax} \varepsilon \eta \lambda \supseteq (\star \theta) \phi \bullet \]

\[\{ (\star \theta, \varepsilon \eta \lambda \chi \xi \mathcal{F}) \} (\mathcal{G}) \mathbb{P} \varepsilon \chi \mu \varepsilon \chi \mu =: (\star \theta, \varepsilon \eta \lambda \chi) \mathcal{F} \bullet \]

\[\begin{aligned}
\end{aligned} \]

\[\{ \exists \varepsilon \in \mathcal{F} \text{ for some } \left[ \left( \xi \varepsilon \in \eta \right) \mathcal{I} \mathbb{H} = \varepsilon \eta \right] \text{ for } \mathcal{M} \mathbb{A} \mathcal{R} \mathbb{G} (\chi) \bullet \]

Optimal Bounds (\star \theta) \phi
Extensions to Hypergraphs.

- Fixed points always exist.

Geometric Interpretation: Reparameterization continues us to an e-flat manifold, and we “project” onto an m-flat manifolds.

- and exponential parameters via duality.

Convexity of Log partition function: Links mean parameters

- Tree Reparameterization is equivalent to BP.

Conclusion
• Upper bounds: solving a convex optimization gives upper bounds to $\Phi(\theta)$. 
\[
\begin{split}
\mathcal{L}_I \mathbb{H} + \mathcal{L}_H \mathbb{H} &= \text{Bethe} \\
\mathcal{L}_I \mathbb{H} + \mathcal{L}_H \mathbb{H} &= \text{Bethe} \\
\end{split}
\]

The Bethe free energy approximates the Gibbs free energy.

With a tree approximation:

\[ (x)q \log (x)q \mathbb{H} + (x)q \mathbb{H} = ((x)q)G \]

Minimizing the Gibbs free energy functional.

\[ \frac{1}{\mathcal{L}} / (x)q - \exp \frac{Z}{T} = (x)d \]

Boltzmann: BP and Bethe Free Energy