We're talking about Loopy BP as optimization

- 6 : send msg (i,j) for every directed edge $(i,j) \in E$

- $\nu^{(t+1)}_{i \to j} \propto \prod_{k \in \delta i \setminus j} \left\{ \sum_{x_k \in \mathcal{X}} \psi_{ik}(x_i, x_k) \nu^{(t)}_{k \to i}(x_k) \right\}$

BP is just $F : \underline{M(\mathcal{X})^{\vec{E}}} \to M(\mathcal{X})^{\vec{E}}$

$F(\nu^{(t)}) \to \nu^{t+1}$

- What does convergence mean?

$$\underbrace{F \ldots F(F(\nu_0))}_{t} = \underbrace{F(\ldots F(\nu_0))}_{t}$$

Found a fixed point of F.

Does F have a fpt? More than 1? When do we find them? Can we characterize them?

Existence

Any continuous fn mapping from $\mathcal{X}$ to $\mathcal{X}$ where $\mathcal{X}$ is closed + bounded + convex has a fixed pt.
↳ compact

Yes, BP has ≥ 1 fixed pt.

What is/are the fixed pts of F/BP?

<u>Variational methods</u> treat BP as optimization to analyze.

Accurate/Complex ⟵ BP ⟶ Simple/Inaccurate

Gibbs Free (2) Energy — Bethe Free Energy (3) — (1) Naive mean field

Gibbs variational Principle

- Start w hard optimization
- Add constraints, search smaller feasible set
- Compare to BP.

Can we est Z? given ψ_{ij}

- BP aims to find

$$\mu(x) = \frac{1}{Z} \prod_{(i,j) \in E} \psi_{ij}(x_i x_j) = \frac{1}{Z} \psi_{tot}(x)$$

- $b(x) \in M(\mathcal{X}^{|V|})$ can be treated as a prob. dist/<u>beliefs</u>

- We will analyze estimating the log partition fn

$$\underline{\Phi} \triangleq \log Z = \log\left(\sum_{x \in \mathcal{X}^n} \prod_{ij} \psi_{ij}(x_{ij})\right)$$

Def: The Gibbs free energy

$$G_{f_{total}}(b) = \underbrace{\sum_{x \in \mathcal{X}^n} b(x) \cdot \log(f_{total}(x))}_{\underbrace{E_b[-\log(f_{total}(X))]}_{\text{expected energy of } X \text{ w.r.t. } b}} - \underbrace{\sum b(x) \log(b(x))}_{\text{Entropy of } b}$$

$-\log(f_{total}(X))$: "energy" of state x

The variational characterization of Φ is $\Phi = \sup_{b \text{: PMFs over } \mathcal{X}^n} G(b)$

Claim:
- $G_{f_{total}}(b)$ is strictly concave
- $\sup_b G_{f_{total}}(b) = \Phi$
- $P(x) = \operatorname{argmax}_b G(b)$

Pf that $\Phi = \log Z = \sup_b G_\psi(b)$

$$G_\psi(b) = \underline{\sum_{x \in \mathcal{X}^{|V|}} b(x) \log \psi_{total}(x)} - \sum_{x \in \mathcal{X}^{|V|}} (b(x) \log(b(x))$$

$$= \sum_{x \in \mathcal{X}^{|V|}} b(x) \log\left(\frac{\psi_{total}(x)}{Z} \cdot Z\right) - \underline{\hphantom{xxxxxxxx}}$$

$$= \sum_x b(x) \left[\log Z + \log \frac{\psi_{total}(x)}{Z}\right] - \underline{\hphantom{xxxxxxxx}}$$

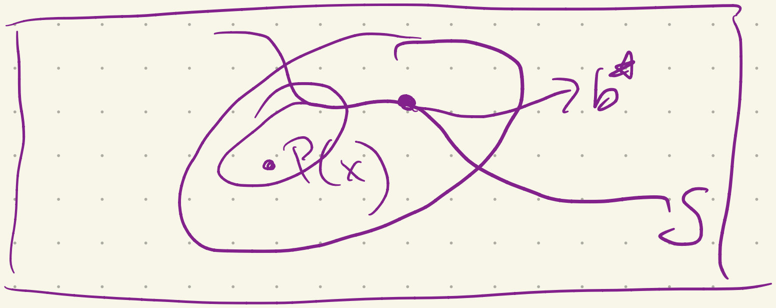
$$= \log(Z) + \sum_x b(x) \log \underbrace{\frac{\psi_{total}(x)}{Z}}_{} - \underline{\hphantom{xxxxxxxx}}$$

$$= \log(Z) + \sum_x b(x) \log(P(x)) - b(x) \log(b(x))$$

$$= \Phi - D_{KL}(b \| P) \leq \Phi$$

achieved for $b = P$, max

$$\Phi \geq \max_{b \in S} G_{f_{total}}(b)$$



Consider Naive MF fact

$$S_{MF} = \{ b \in \Delta_{|\mathcal{X}|^n - 1} : b(x) = b_1(x_1) \cdots b_n(x_n) \}$$

$$b = \{ b_i(\cdot) \}_{i=1}^{n}$$

Solve for b which maximizes GFE$(G(b))$
s.t. $b \in S_{MF}$

$$F_{MF}(b) = G_{f_{total}}(b_1 \times \cdots \times b_n)$$

$$= \sum_{x} b(x) \prod_{(i,j) \in E} \log(f_{ij}(x_i, x_j)) - \sum_{x} b(x) \log(b(x))$$

$$= \sum_{(i,j) \in E} \sum_{x_i, x_j} b_i(x_i) b_j(x_j) \log f_{ij}(x_i, x_j) - \sum_{i \in V} \sum_{x_i} b_i(x_i) \log(b_i(x_i))$$

Naive MF variational inference proble

$$\max_{b \in S_{MF}} \mathbb{F}_{MF}(b)$$

$$\text{s.t.} \quad \sum_{x_i} b_i(x_i) = 1 \quad \forall i \in V$$

Dimension $|\mathcal{X}| \cdot (n-1)$

Bilinear $b_i(x_i)\, b_j(x_j)$ (not concave)

Lagrangian of ↑

$$L(b,\lambda) = \mathbb{F}_{MF}(b) - \sum_{i \in V} \lambda_i \left(\sum_{x_i \in \mathcal{X}} b_i(x_i) - 1 \right)$$

$$\frac{\partial L(b,\lambda)}{\partial b_i(x_i)} = \sum_{j \in \partial i} \sum_{x_j} b_j(x_j) \log(\psi_{ij}(x_i, x_j)) + (1 + b_i \log(b_i)) - \lambda_i$$

b^* is a stationary pt of optimizin F_{MF} s.t. factorization

$$\forall i \quad \log b_i^*(x_i) = 1 + \lambda_i + \sum_{j \in \delta_i} \sum_{x_j} b_j^*(x_j) \log \psi_{ij}(x_i x_j)$$

$$b_i^*(x_i) \propto \exp\left[\qquad \right]$$