STSCI 6940:

Lecture 9 $\qquad\qquad$ Date: 03/11

Heuristic analysis / derivation of AMP.

- Explain the presence of the memory term.

$$\begin{cases} x^{t+1} = A m^t - b_t m^{t-1} \\ m^t = f_t(x^t), \quad b_t = \frac{1}{n}\sum_{i=1}^n f_t'(x_i^t). \end{cases}$$

- $A_{ji} = A_{ij} \sim N(0, \frac{1}{n})$
$A_{ii} = 0$.

$m^{-1} = 0, \quad m^0 \in \mathbb{R}^n$.

$t=0$: $\quad x_i^1 = (A m^0)_i \sim N(0, \frac{1}{n}\|m^0\|_2^2)$.

$t=1$:

$$x_i^2 = \sum_{j=1}^n a_{ij} m_j^1 - b_1 \cdot m_i^0$$

$$m_j^1 = f_1(x_j^1) = f_1\Big(\sum_{k=1}^n a_{jk} m_k^0\Big).$$

$$= f_1\Big(\underbrace{\sum_{k\neq i} a_{jk} m_k^0}_{\hat{x}_{j\to i}} + a_{ji} m_i^0\Big).$$

$a_{ij} = O(\frac{1}{\sqrt{n}})$,

$$= f_1(\hat{x}_{j\to i}) + f_1'(\hat{x}_{j\to i}) \cdot a_{ji} m_i^0 + O(\tfrac{1}{n}).$$

$$\Rightarrow x_i^2 = \sum_{j=1}^n a_{ij} f_1(\hat{x}_{j\to i}) + \sum_{j=1}^n f_1'(\hat{x}_{j\to i}) a_{ij}^2 m_i^0 - b_1 \cdot m_i^0 + o_p(\tfrac{1}{\sqrt{n}})$$

$$\sum_{j=1}^{n} a_{ij}^2 \, f_1'(\hat{x}_{j\to i}) = \frac{1}{n}\sum_{j=1}^{n} f_1'(\hat{x}_{j\to i}) + \sum_{j=1}^{n}\left(a_{ij}^2 - \frac{1}{n}\right) f_1'(\hat{x}_{j\to i})$$

independent

$$= \frac{1}{n}\sum_{j=1}^{n} f_1'(\hat{x}_{j\to i}) + o\left(\frac{1}{\sqrt{n}}\right).$$

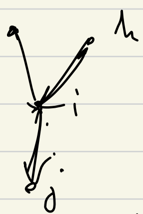
$$= \frac{1}{n}\sum_{j=1}^{n} f_1'(x_j^1) + o\left(\frac{1}{\sqrt{n}}\right).$$

$$b_1 = \frac{1}{n}\sum_{j=1}^{n} f_1'(x_j^1).$$

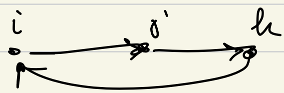
$$\Rightarrow \quad x_i^2 = \sum_{j=1}^{n} a_{ij} f_1(\hat{x}_{j\to i}) + o\left(\frac{1}{\sqrt{n}}\right).$$

$$\sim \mathcal{N}\left(0, \frac{1}{n}\|f_1(\hat{x}_{j\to i})\|_2^2\right).$$

$$\text{"}\sim\text{"} \ \mathcal{N}\left(0, \frac{1}{n}\|m^1\|_2^2\right).$$



$$x_i^1 = \sum_j a_{ij} m_j^0$$



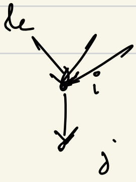
Backtracking term.

The role of the memory term is to cancell backtracking term (which are non Gaussian).

- Derivation of AMP from message passing on the complete graph:

MP:

- A graph $G = (V, E)$. $(m_{i\to j}^t)_{(i,j)\in \vec{E}}$: the messages



$$\begin{cases} x_{i\to j}^{t+1} = \sum_{k\in\partial i\setminus\{j\}} a_{ik}\, m_{k\to i}^t \\ m_{k\to i}^t = f_t(x_{k\to i}^t). \end{cases}$$

$$\mu_G(x) = \frac{1}{Z_G} \prod_{i\in V} \psi_i(x_i) \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j).$$

Belief propagation:

$$\mu^{t+1}_{i\to j}(x_i) = \psi_i(x_i) \prod_{k\in\partial i\setminus\{j\}} \left(\sum_{x_k} \psi(x_i, x_k)\, \mu^t_{k\to i}(x_k)\right)$$

MP maintains $O(n^2)$ messages at every iteration.

$$x^{t+1}_{i\to j} = \sum_{k=1}^{n} a_{ik}\, m^t_{k\to i} - \underbrace{a_{ij}\, m^t_{j\to i}}_{= O(\frac{1}{\sqrt{n}})} \quad ⊛$$

If remove ⊛ : we get power iteration.

Let $\quad \hat{x}^{t+1}_i = \sum_{k=1}^{n} a_{ik}\, m^t_{k\to i}$.

$$m^t_{i\to j} = f_t(x^t_{i\to j}) = f_t(\hat{x}^t_i - a_{ij}\, m^{t-1}_{j\to i}).$$

$$= f_t(\hat{x}^t_i) - \underbrace{a_{ij}\, m^{t-1}_{j\to i}\, f'_t(\hat{x}^t_i)}_{O(\frac{1}{\sqrt{n}})} + O(\tfrac{1}{n})$$

Let $\quad \hat{m}^t_i = f_t(\hat{x}^t_i)$.

$$\Rightarrow \hat{x}^{t+1}_i = \sum_{k=1}^{n} a_{ik}\left(\hat{m}^t_k - a_{ki}\, f'_t(\hat{x}^t_k)\, m^{t-1}_{i\to k}\right) + O(\tfrac{1}{\sqrt{n}}).$$

$$= \sum_{k=1}^{n} a_{ik}\, \hat{m}^t_k - \sum_{k=1}^{n} a_{ki}^2\, f'_t(\hat{x}^t_k)\, \hat{m}^{t-1}_i + O(\tfrac{1}{n})$$

$$= \underline{\qquad\qquad} - \left(\frac{1}{n}\sum_{k=1}^{n} f'_t(\hat{x}^t_k)\right) \hat{m}^{t-1}_i + O(\tfrac{1}{\sqrt{n}})$$

$$\begin{cases} \hat{x}^{t+1} = A\,\hat{m}^{t} - b_t \cdot \hat{m}^{t-1} \\ \hat{m}^{t} = f_t(\hat{x}^{t}) \end{cases}.$$

we reduced message passing iteration (involves $O(n^2)$ messages) to an iteration involving n terms.

$\Rightarrow$ "Approximate" MP.

the main assumption: although the graph is dense, the interactions between vertices are small: order $\frac{1}{\sqrt{n}}$.

Experiment 1: run MP on the complete graph $(x^0_{i\to j} = x^0_i)$

Experiment 2: run AMP. with init x^0_i.

Claim: $\forall t \geq 1: \frac{1}{n}\sum_{i=1}^{n} | x^t_{i\to j} - x^t_i |^2 = o(\frac{1}{n})$.

Heuristic derivation of state evolution:

$$A = \frac{\sqrt{\lambda}}{n} x_0 x_0^T + \frac{1}{\sqrt{n}} W.$$

$$x^{t+1} = A\, m^t - b_t\, m^{t-1}$$

$$= \frac{\sqrt{\lambda}}{n}(x_0^T m^t) \cdot x_0 + \frac{1}{\sqrt{n}} W m^t - b_t\, m^{t-1}.$$

Assumption 1: $\frac{1}{n} x_0^T \cdot m^t$ concentrates around $\mu_t \in \mathbb{R}$.

Assumption 2: Discard the memory term, and replace the noise W by a fresh copy at every iteration t.

$$x_i^{t+1} = \sqrt{\lambda}\, \mu_t \cdot x_{0i} + \tilde{\xi}_i$$

$$\mathbb{E}\left[\tilde{\xi}_i\right] = 0 \,, \qquad \mathbb{E}\left\{\tilde{\xi}_i^2\right\} = \frac{1}{n}\| m^t \|_2^2$$

Assumption 3: $\frac{1}{n}\| m^t \|_2^2$ concentrates around σ_t^2.

$$x_i^{t+1} = \sqrt{\lambda}\, \mu_t \, x_{0i} + \sigma_t \cdot \xi_i \qquad \xi_i \overset{iid}{\sim} N(0,1).$$

$$\frac{1}{n}\sum_{i=1}^{n} \phi(x_i^{t+1}, x_{0i}) = \frac{1}{n}\sum_{i=1}^{n} \phi(\sqrt{\lambda}\, \mu_t x_{0i} + \sigma_t \xi_i, x_{0i})$$

$$\longrightarrow \mathbb{E}\, \phi(\sqrt{\lambda}\, \mu_t x_0 + \sigma_t \xi, x_0).$$

$$\mu_{t+1} = \frac{1}{n} x_0^T \cdot m^{t+1} = \frac{1}{n} x_0^T \cdot f_{t+1}(x^{t+1})$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{0i} f_{t+1}(\sqrt{\lambda}\, \mu_t x_{0i} + \sigma_t \xi_i)$$

$$\longrightarrow \mathbb{E}\left[x_0 f_{t+1}(\sqrt{\lambda}\, \mu_t x_0 + \sigma_t \xi) \right].$$

Next σ_{t+1}:

$$\sigma_{t+1}^2 \simeq \frac{1}{n}\| m^{t+1} \|_2^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_{t+1}(x_i^{t+1})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_{t+1}(\sqrt{\lambda}\, \mu_t x_{0i} + \sigma_t \xi_i)^2$$

$$\longrightarrow \mathbb{E}\left\{ f_{t+1}(\sqrt{\lambda}\, \mu_t x_0 + \sigma_t)^2 \right\}$$