Mean Field Asymptotics in Statistical Learning. Apr 14

Lecture 22. AMP and state evolution.

[Bayati, Montanari, 2011]
[Berthier, Montanari, Nguyen, 2017].
[Rush Venket... 2017]

① LASSO AMP and State evolution.

The AMP algorithm: Fix $\{\eta_k : \mathbb{R} \to \mathbb{R}\}_{k\geq 1}$ [Chen, Lam, 2020]

$\eta_k = \eta(x; \lambda \zeta^k)$

$$
\boxed{
\begin{aligned}
x^{k+1} &= \eta_k(\theta^k) \in \mathbb{R}^d \qquad x^0 = 0, \quad z^0 = 0, \quad z^1 = 0. \\
\theta^k &= x^k + A^T z^k \quad \in \mathbb{R}^d \\
z^k &= y - Ax^k + \underbrace{\left[\frac{1}{n}\sum_j \eta_k'(\theta_j^k)\right] z^{k-1}}_{\text{Onsager.}} \quad \in \mathbb{R}^n
\end{aligned}
}
$$

$x \to -x$

$y = Ax + w$

$y \neq$

State evolution:

$$
\boxed{
\begin{aligned}
&\tau_{k+1}^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}[(\eta_k(X_0 + \tau_k G) - X_0)^2]. \\
&(X_0, G) \sim \mathbb{P}_0 \times N(0,1).
\end{aligned}
}
$$

Assumptions:

a) $A \in \mathbb{R}^{n\times d}$, $A_{ij} \sim_{iid} N(0, \frac{1}{n})$.

b) $x_0 \in \mathbb{R}^d$, $\frac{1}{d}\sum_{i=1}^d \delta_{x_{0,i}} \xrightarrow{weak} P_0$, $\frac{1}{d}\sum_{i=1}^d x_{0,i}^2 \longrightarrow \mathbb{E}_{X_0}[X_0^2] > 0$

c) $w \in \mathbb{R}^n$ $\frac{1}{n}\sum_{i=1}^n \delta_{w_i} \xrightarrow{weak} P_W$, $\frac{1}{n}\sum_{i=1}^n w_i^2 \longrightarrow \mathbb{E}_W[W^2] \equiv \sigma^2$.

d) $y = Ax_0 + w \in \mathbb{R}^n$.

e) $n/d \longrightarrow \delta$

Thm (Bayati. Montanari, 2011)

Let assumptions a) – e) hold. Let $\psi : \mathbb{R}^2 \to \mathbb{R}$ be any pseudo-Lipschitz function. Then a.s

$$
\lim_{\substack{d\to\infty \\ n/d \to \delta}} \frac{1}{d}\sum_{i=1}^d \psi(x_i^{k+1}, x_{0,i}) = \mathbb{E}[\psi(\eta_k(X_0 + \tau_k G), X_0)].
$$

$$
\frac{1}{d}\sum_{i=1}^d \psi(\theta_i^{k+1}, x_{0,i}) = \mathbb{E}[\psi(X_0 + \tau_k G, X_0)].
$$

② Intuition and heuristic derivation of SE.

Consider the auxilliary dynamics

Auxilliary dynamics (Not an algorithm)

$$
\begin{aligned}
z^k &= A_k(x_0 - x^k) + w \\
\theta^k &= x^k + A_k^T z^k \\
x^{k+1} &= \eta_k(\theta^k)
\end{aligned}
$$

AMP

$$
\begin{aligned}
z^k &= A(x_0 - x^k) + w + \underbrace{\left[\frac{1}{n}\sum_j \eta_k'(\theta_j^k)\right] z^{k-1}}_{\text{Onsager term}} \\
\theta^k &= x^k + A^T z^k \\
x^{k+1} &= \eta_k(\theta^k)
\end{aligned}
$$

Difference:

- ✲ No Onsager term in auxilliary dynamics
- ✲ $A_k \overset{d}{=} A$, A_k's are independent copies of A. Bolthausen 2012

Relationship:

- ✲ Approximately equivalent in distribution. (Vague),

Reason: Gaussian conditioning trick: Gaussian conditioning on linear equality constaint is still Gaussian.

$$\mathcal{F}_k = \sigma(x^0, z^0, x^1, z^1, \cdots, x^k, z^k, \theta^1, \cdots \theta^k)$$

$$A \mid \mathcal{F}_k = A \mid \{AU = U', A^T Q = Q'\} \overset{d}{=} P_Q^\perp A_k P_U^\perp + \text{low rank update}$$

where $A_k \overset{d}{=} A$.

Consider the auxilliary dynamics. Suppose SE holds up to time k

Then $\theta^{k-1} \overset{d}{\approx} x_0 + \tau_{k-1} \cdot g \quad \in \mathbb{R}^d$

$$x^k = \eta_{k-1}(\theta^{k-1}) \overset{d}{\approx} \eta_{k-1}(x_0 + \tau_{k-1} \cdot g). \in \mathbb{R}^d.$$

where $g \in \mathbb{R}^d$, $g_i \sim_{iid} N(0,1)$.

$$
\begin{aligned}
\theta^k - x_0 &= x^k + A_k^T(A_k(x_0 - x^k) + w) - x_0 \\
&= (A_k^T A_k - I_d)(x_0 - x^k) + A_k^T w.
\end{aligned}
$$

$$\theta^k - x_0 \overset{d}{\approx} N(0, (\tau^k)^2 I_d).$$

$$(\tau^k)^2 = \lim_{\substack{d\to\infty \\ n/d\to\delta}} \frac{1}{d} \|\theta^k - x_0\|_2^2$$

$$= \lim_{\substack{d\to\infty \\ n/d\to\delta}} \mathbb{E}\left[\|(A_k^T A_k - I)(x_0 - x^k) + A_k^T w\|_2^2 \right]/d .$$

$$\mathbb{E}\|A_k^T w\|_2^2/d = \|w\|_2^2/n \longrightarrow \sigma^2 \quad \text{as} \quad n\to\infty .$$

$$\mathbb{E}\left[\|(A_k^T A_k - I)(x_0 - x^k)\|_2^2 \right] \approx \frac{d}{n}\left(\|x^k - x_0\|_2^2/d\right)$$

$$\to \frac{1}{\delta} \mathbb{E}\left[\left(\eta_{k-1}(x_0 + \tau_{k-1} G) - x_0\right)^2 \right] .$$

$$\Rightarrow \quad (\tau^k)^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E}\left[\left(\eta_{k-1}(x_0 + \tau_{k-1} G) - x_0\right)^2 \right] ,$$

③ A more general theorem. $(N = n+d)$.

Standard form of AMP (Symmetric)

The LASSO AMP, $\mathbb{Z}_2$ sync AMP can be reduced to this form.

Fix $f_k : \mathbb{R}^N \to \mathbb{R}^N$, $\left(\text{separable. } f_k(u) = \begin{bmatrix} f_{k,1}(u_1) \\ \vdots \\ f_{k,N}(u_N) \end{bmatrix}\right)$

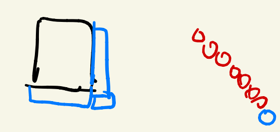
$W \sim GOE(N)$

$$\boxed{\begin{aligned} u^{k+1} &= W m^k \underbrace{- b_k m^{k-1}}_{\text{Onsager term}} && \in \mathbb{R}^N \\ m^k &= f_k(u^k) && \in \mathbb{R}^N \\ b_k &= \frac{1}{N}\sum_{i=1}^N \partial_i f_{k,i}(u^k) && \in \mathbb{R}. \end{aligned}}$$

SF-AMP-Symmetric

State evolution: $(K_{s,t})_{s,t \le k}$

K^k



$$\boxed{\begin{aligned} & K_{s+1,t+1} \equiv \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\langle f_s(Z^s), f_t(Z^t)\rangle\right] \quad ✡ \\ & Z^s, Z^t \in \mathbb{R}^N, \quad (Z^s_i, Z^t_i)_{i\in[N]} \sim_{iid} N\left(0, \begin{bmatrix} K_{ss} & K_{st} \\ K_{st} & K_{tt} \end{bmatrix}\right) \end{aligned}}$$

Assumptions:

a) $W \sim GOE(N)$, $W_{ij} \sim N(0, \frac{1}{N})$, $i<j$. $W_{ii} \sim N(0, \frac{2}{N})$.

b) $f_k : \mathbb{R}^N \to \mathbb{R}^N$ uniform Lipschitz.

c) $\frac{1}{N}\sum_{i=1}^N \delta_{u_i^0} \longrightarrow \mathbb{P}_0$

d) Assumptions gaurantees that ✡ exists.

Thm. (Simplified version in [Berthier, Montanari, Nguyen, 2017]).

Under asumptions above, consider AMP iterations $\{u^k, m^k \mid f_k, u^0\}$. Define $U_0 \sim \mathbb{P}_0$

$$(Z^1, \cdots, Z^{k+1}) \sim N(0, (K_{s,t})_{s,t\le k+1})$$

For any pseudo-Lipschitz function, $\{\psi : \mathbb{R}^{k+2} \to \mathbb{R}\}$.

$$\frac{1}{N}\sum_{i=1}^N \psi(u_i^0, u_i^1, \cdots, u_i^{k+1}) \longrightarrow \mathbb{E}[\psi(U_0, Z^1, \cdots, Z^{k+1})].$$

---

④ How to write LASSO AMP into standard form?

AMP: Standard form, non-symmetric.

$A \in \mathbb{R}^{n\times d}$

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

$$\begin{cases} p^{k+1} = A^T g_k(q^k) - \underline{d_k\, e_k(p^k)} \\ q^k = A\, e_k(p^k) - \underline{b_k\, g_{k-1}(q^{k-1})} \\ d_k = \frac{1}{n}\sum_{i=1}^{n} \partial_i g_{k,i}(q^k), \quad b_k = \frac{1}{n}\sum_{i=1}^{d} \partial_i e_{k,i}(p^k) \end{cases}$$

SF-AMP - non symmetric

$\frac{1}{n}\,\mathrm{div}\, e_k(p^k)$

Recall LASSO AMP gives:

$$\begin{aligned} x^{k+1} &= \eta_k(\theta^k) \\ \theta^k &= x^k + A^T z^k \\ z^k &= y - Ax^k + \underline{\left[\frac{1}{n}\sum_j \eta_k'(\theta_j^k)\right]} z^{k-1} \end{aligned}$$

LASSO-AMP.

Lemma: Given the following transformation, LASSO-AMP can be transformed to the SF-AMP-Nonsymmetric

$$\begin{cases} p^{k+1} = x_0 - (A^T z^k + x^k) \\ q^k = w - z^k \end{cases}$$

$$\begin{cases} g_k(q^k) = q^k - w \\ e_k(p^k) = \eta_k(x_0 - p^k) - x_0 \end{cases}$$

Lemma: Given the following transformation. SF-AMP-Nonsymmetric can be transformed into SF-AMP-symmetric.

$N = n+d$.

$$W = \sqrt{\frac{n}{n+d}}\begin{bmatrix} B & A \\ A^T & C \end{bmatrix}, \quad B \sim GOE(n), \quad \sqrt{\frac{n}{d}}\, C \sim GOE(d)$$

$$f_{2k+1}(u) = \sqrt{\frac{n+d}{n}}\begin{bmatrix} g_t(u_{1:n}) \\ 0 \end{bmatrix}, \quad f_{2k}(u) = \sqrt{\frac{n+d}{d}}\begin{bmatrix} 0 \\ e_k(u_{n+1:N}) \end{bmatrix}.$$

ⓑ Analysis of SF-AMP-Symmetric. — Two steps analysis

$$u^{k+1} = W m^k \underbrace{- b_k m^{k-1}}_{\text{Onsager term}} \in \mathbb{R}^N$$

$$m^k = f_k(u^k) \in \mathbb{R}^N$$

$$b_k = \frac{1}{N}\sum_{i=1}^{N} \partial_i f_{k,i}(u^k) \in \mathbb{R}.$$

Step 1: $u^1 = W m^0$ $\quad m^0 \quad m^{-1}=0$

$\left(b_1 = \frac{1}{N}\sum_{i=1}^{N} \partial_i f_{1,i}(u^1)\right)$

Step 2: $m^1 = f_1(u^1)$

$u^2 = W m^1 - b_1 \cdot m^0$

Step 1: $u^1 = W m^0$ $\quad W \sim GOE(N).$

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{u^1_i} \longrightarrow N(0, \|m^0\|_2^2/N).$$

$$W = (G + G^T)/\sqrt{2n} \qquad G_{ij} \sim_{iid} N(0,1).$$

$$W m^0 = (G m^0 + G^T m^0)/\sqrt{2n}$$

Step 2: Problem: W and m^1 are not independent

$m^1 = f_1(W m^0)$ $\quad u_2 = W f_1(W m^0) - b_1 m^0$

Trick: Gaussian conditioning trick.

$$\mathcal{F}_1 = \sigma(m^0, u^1).$$

$$W \mid \mathcal{F}_1 \overset{d}{=} W \mid \{u^1 = W m^0\}$$

Gaussian conditioning

$$= P_{m_0}^{\perp} W_1 P_{m_0}^{\perp} + \frac{\langle m^0, m^1\rangle}{\|m^0\|_2^2}\left[m^0 (u^1)^T + u^1 (m^0)^T\right] - \frac{\langle m^0, u^1\rangle \langle m^0, m^1\rangle}{\|m^0\|_2^4} m_0$$

$$W_1 \overset{d}{=} W$$

$$\frac{\langle m^0, W m^0\rangle}{\|m^0\|_2^2} \approx 0.$$

$$Wm^1 \,\Big|\, \mathcal{F}_1 \overset{d}{=} P_{m^0}^{\perp} W_1 P_{m^0}^{\perp} m^1 + \frac{\langle m^0, m^1\rangle}{\|m^0\|_2^2} u^1 + \underbrace{\frac{\langle u^1, m^1\rangle}{\|m^0\|_2^2} m^0}_{\approx\, b_1 m^0} - \underbrace{\frac{\langle m^0, u^1\rangle\langle m^0, m^1\rangle}{\|m^0\|_2^4} m^0}_{\approx\, 0}$$

Onsager correction term

$$\frac{\langle u^1, m^1\rangle}{\|m^0\|_2^2} \approx \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \tau_0 G \, f_{1,i}(\tau_0 G)\right] \Big/ \tau_0^2 = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \partial_i f_{1,i}(\tau_0 G)\right]$$

$$\approx \frac{1}{N}\sum_{i=1}^N \partial_i f_{1,i}(u^1) = b_1.$$

$$\Rightarrow \quad u^2 \,\Big|\, \mathcal{F}_1 \overset{d}{\approx} \underbrace{P_{m^0}^{\perp} W_1 P_{m^0}^{\perp} m^1}_{\text{approximately i.i.d. Gaussian coordinates.}} + \frac{\langle m^0, m^1\rangle}{\|m^0\|_2^2} u^1.$$

Define $h = P_{m^0}^{\perp} W_1 P_{m^0}^{\perp} m^1$,

$$\Rightarrow \quad \frac{1}{N}\sum_{i=1}^N \delta_{h_i} \longrightarrow N\left(0, \frac{\|P_{m^0}^{\perp} m^1\|_2^2}{N}\right).$$

$$\Rightarrow \quad \frac{1}{N}\sum_{i=1}^N \delta_{(u_i^1, u_i^2)} \longrightarrow N\left(0, \begin{bmatrix} \frac{\|m^0\|_2^2}{N}, & \langle m^0, m^1\rangle / N \\ \frac{\langle m^0, m^1\rangle}{N}, & \|m^1\|_2^2 / N \end{bmatrix}\right).$$

$$\begin{bmatrix} \frac{\|m^0\|_2^2}{N} & \frac{\langle m^0, m^1\rangle}{N} \\ \frac{\langle m^0, m^1\rangle}{N}, & \frac{\langle m^1, m^1\rangle}{N} \end{bmatrix} \approx \begin{bmatrix} \mathbb{E}[f_0(U_0)], & \mathbb{E}[f_0(U_0) f_1(Z^1)] \\ \mathbb{E}[f_0(U_0) f_1(Z^1)], & \mathbb{E}[f_1(Z^1)^2] \end{bmatrix}$$

$$= \begin{bmatrix} K_{00} & K_{01} \\ K_{01} & K_{11} \end{bmatrix}.$$