STSCI 6940:

Lecture 8 Date: 03/04

Announcements:

→ Homework: Thursday April 8 (4pm)
return: Next Monday (5pm) 12th.

→ Month or so to read the paper.

one-on-one meetings March 22.

→ Presentations will start on the week of April 12. and end week of May 6.

→ Final week: no classes

→ Deadline for projects: Friday May 14. (midnight).

→ Project: write up to 8 pages in Neurips latex format.

- Approximate message passing :

$A \in \mathbb{R}^{n \times n}$ symmetric.

Initialize: $x^0 \in \mathbb{R}^n$, $x^{-1} = 0$ $f_0 \equiv 0$

Iteration: $x^{t+1} = A f_t(x^t) - b_t \cdot f_{t-1}(x^{t-1})$.

$f_t : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous, differentiable.

$$f_t(x^t) = \left(f_t(x_i^t)\right)_{i=1}^n$$

$$b_t = \frac{1}{n} \sum_{i=1}^n f_t'(x_i^t)$$

(remark: Can replace b_t by $\mathbb{E}[f_t'(\mu_t x_0 + \sigma_t \xi)]$)

- Non linear power iteration:

$$\begin{cases} x^{t+1} = A \hat{x}^t \\ \hat{x}^{t+1} = \dfrac{x^{t+1}}{\|x^{t+1}\|_2} \end{cases} \qquad f(x) = \frac{x}{\|x\|_2}.$$

Feature of the algorithm : presence of

the "memory" term : $- b_t f_{t-1}(x^{t-1})$.

will ensure the iterates are Gaussian.
under distributional assumptions on A.

Assumptions the matrix A:

$$Y = \sqrt{\frac{\lambda}{n}}\, x_0 x_0^T + W \qquad w_{ij} \sim N(0,1).$$

$$w_{ij} = w_{ji}.$$

$$A = \frac{1}{\sqrt{n}} Y. \qquad w_{ii} \sim N(0,2).$$

More generally: $w_{ij} = w_{ji}$ $\quad \mathbb{E}[w_{ij}] = 0$

$$\mathbb{E}[w_{ij}^2] = 1$$

w_{ij} is subgaussian

$$\left(\mathbb{E}[e^{\lambda w_{ij}}] \leq e^{\frac{\lambda \sigma^2}{2}} \ \forall \lambda \right).$$

- Exactly characterize the law of the iterates in the limit $n \to +\infty$.

- Theorem: If A is as above,

The signal x_0 and x^0_{init} have an empirical distribution of entries wich converges weakly to a distribution $P_0, \hat{P}_0$ (on $\mathbb{R}$) with bounded second moment.

Let $\hat{x}^t = f_t(x^t)$; and assume

$$\frac{1}{n} \|\hat{x}^0_{init}\|_2^2 \longrightarrow \rho \,, \quad \frac{1}{n} \hat{x}^{0T}_{init} x_0 \longrightarrow \tau$$

Then for every pseudo-Lipschitz function $\psi: \mathbb{R}^2 \to \mathbb{R}$ we have

$\forall t \geq 0$

$$\frac{1}{n} \sum_{i=1}^{n} \psi(x_i^t, x_{0i}) \longrightarrow \mathbb{E}\left[\psi(\mu_t x_0 + \sigma_t \xi, x_0)\right]$$

where $x_0 \sim P_0$, $\xi \sim N(0,1)$ independent.

- (μ_t, σ_t) satisfy the state evolution iteration.

$$\begin{cases} \mu_{t+1} = \sqrt{\lambda}\, \mathbb{E}[x_0 f_t(\mu_t x_0 + \sigma_t z)] . \\ \sigma_{t+1}^2 = \mathbb{E}[f_t(\mu_t x_0 + \sigma_t z)^2] . \\ \mu_0 = \sqrt{\lambda}\, \tau , \quad \sigma_0^2 = \rho . \end{cases}$$

Definitions:

- $x(n) \in \mathbb{R}^n$: The empirical distribution of the entries of $x(n)$ is

$$m_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i(n)} .$$

→ A a sequence of probability measures $(m_n)_n$.

$$m_n \overset{\text{weakly}}{\rightsquigarrow} p_0 .$$ has finite second moment.

- ψ is a pseudo-Lipschitz function if

$$|\psi(x) - \psi(y)| \leq L(1 + \|x\|_2 + \|y\|_2) \cdot \|x - y\|_2 .$$

An example is $\psi(x) = x^2$.

- Choice of f_t : Suppose you're interested in estimating x_0 (signal).

$$\begin{cases} \text{maximize overlap between } x^t \text{ and } x_0 \\ \text{minimize MSE} \end{cases}$$

- Suppose we observe x_0 through the Gaussian additive channel $\bar{y} = \mu_t x_0 + \sigma_t z$.

Best estimator is posterior mean:

$$\mathbb{E}[x_0 | \tilde{y}] = \frac{\int x e^{\frac{\mu_t}{\sigma_t^2}\tilde{y}x - \frac{\mu_t^2}{2\sigma_t^2}x^2} dP_0(x)}{\int e^{\overline{\qquad\qquad}} dP_0(x)}.$$

let

$$g(y, q) = \frac{\int x e^{\sqrt{q}yx - \frac{q}{2}x^2} dP_0(x)}{\int e^{\sqrt{q}yx - \frac{q}{2}x^2} dP_0(x)}.$$

Then $\mathbb{E}[x_0 | \tilde{y}] = g\left(\frac{\tilde{y}}{\sigma_t}, \left(\frac{\mu_t}{\sigma_t}\right)^2\right)$

Let

$$\underline{f_t(y) = g\left(\frac{y}{\sigma_t}, \left(\frac{\mu_t}{\sigma_t}\right)^2\right) \quad \forall y \in \mathbb{R}.}$$

our estimator is

$$\hat{x}^t = f_t(x^t)$$

x^t: t^{th} iterate of AMP.

let $\tilde{y} = \mu_t x_0 + \sigma_t z$

$\hat{x}^t \cdot x_0$, $\|\hat{x}^t\|_2^2$?

$$\begin{aligned}
\mu_{t+1} &= \sqrt{\lambda}\, \mathbb{E}[x_0 f_t(\mu_t x_0 + \sigma_t z)]. \\
&= \sqrt{\lambda}\, \mathbb{E}\left[x_0\, g\left(\frac{\tilde{y}}{\sigma_t}, \left(\frac{\mu_t}{\sigma_t}\right)^2\right)\right] \\
&= \sqrt{\lambda}\, \mathbb{E}[x_0 \cdot \mathbb{E}[x_0 | \tilde{y}]] \\
&= \sqrt{\lambda}\, \mathbb{E}[\mathbb{E}[x_0 | \tilde{y}]^2] \quad \text{tower property.} \\
&= \sqrt{\lambda}\, \mathbb{E}[f_t(\mu_t x_0 + \sigma_t z)^2] \\
&= \sqrt{\lambda}\, \sigma_{t+1}^2.
\end{aligned}$$

Define $q_t = \frac{1}{\lambda} \left(\frac{\mu_t}{\sigma_t}\right)^2$.

$$q_{t+1} = \frac{1}{\lambda} \frac{\mu_{t+1}}{\sigma_{t+1}^2} \cdot \mu_{t+1}$$

$$= \frac{1}{\lambda} \sqrt{\lambda} \cdot \sqrt{\lambda} \cdot \mathbb{E}\left[f_t(\mu_t x_0, \sigma_t z)^2 \right].$$

$$= \mathbb{E}\left[g\left(\frac{\mu_t x_0 + \sigma_t z}{\sigma_t}, \left(\frac{\mu_t}{\sigma_t}\right)^2 \right)^2 \right]$$

$$= \mathbb{E}\left[g\left(\frac{\mu_t}{\sigma_t} x_0 + z, \left(\frac{\mu_t}{\sigma_t}\right)^2 \right)^2 \right]$$

$$= \mathbb{E}\left[g\left(\sqrt{\lambda q_t}\, x_0 + z, \lambda q_t \right)^2 \right].$$

$$\psi(r) = \mathbb{E} \log \int e^{\sqrt{r} z x + r x x_0 - \frac{r}{2} x^2} dP_0(x)$$

$$y = \sqrt{r}\, x_0 + z$$

$$\psi'(r) = \frac{1}{2} \mathbb{E}\left[\mathbb{E}[x_0 | y]^2 \right],$$

Therefore $\boxed{q_{t+1} = 2\psi'(\lambda q_t).}$

Let's understand $\hat{x}^t \cdot x_0$ and $\|\hat{x}^t\|_2^2$.

$$\hat{x}^t \cdot x_0 = \frac{1}{n} \sum_{i=1}^{n} f_t(x_i^t) \cdot x_{0i}$$

Now let $\psi(x, x_0) = f_t(x) \cdot x_0$.

Apply the theorem:

$$\hat{x}^t \cdot x_0 \longrightarrow \mathbb{E}\left[x_0 f_t(\mu_t x_0 + \sigma_t z) \right].$$

$$= \frac{\mu_{t+1}}{\sqrt{\lambda}} = q_{t+1}.$$

Now $\frac{1}{n}\|\hat{x}^t\|_2^2$.

Apply the theorem to $\psi(x, x_0) = f_t(x)^2$.

$$\Rightarrow \quad \frac{1}{n}\|\hat{x}^t\|_2^2 \longrightarrow \mathbb{E}\left[f_t(\mu_t x_0 + \sigma_t z)^2\right]$$

$$= q_{t+1}.$$

The normalized overlap:

$$\frac{\langle \hat{x}^t, x_0\rangle}{\|\hat{x}^t\|_2 \cdot \|x_0\|_2} \longrightarrow \frac{q_{t+1}}{\sqrt{q_{t+1}} \cdot 1} = \sqrt{q_{t+1}}.$$

q_t will converge to a fixed point of the function $F(q) = 2\psi'(\lambda q)$.

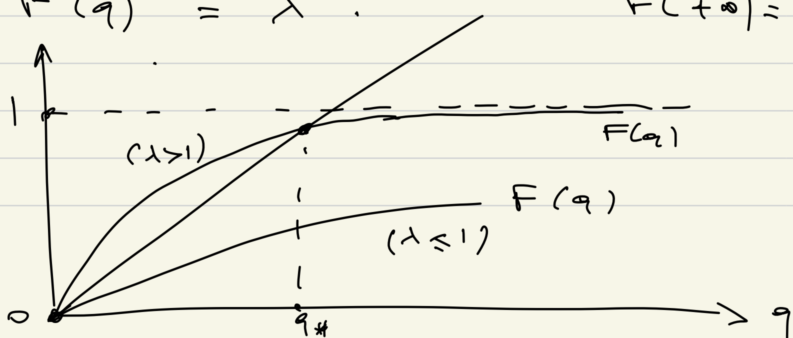
$$F(0) = \mathbb{E}[x_0]^2 = 0.$$

$$\psi''(r) = \frac{1}{2}\mathbb{E}\left[\mathrm{cov}(x|y)^2\right]$$

$$r = 0 \;;\quad \psi''(0) = \frac{1}{2}\mathbb{E}[x^2] = \frac{1}{2}.$$

$$\Rightarrow F'(q) = \lambda.$$

$$F(+\infty) = \mathbb{E}[x_0^2] = 1$$

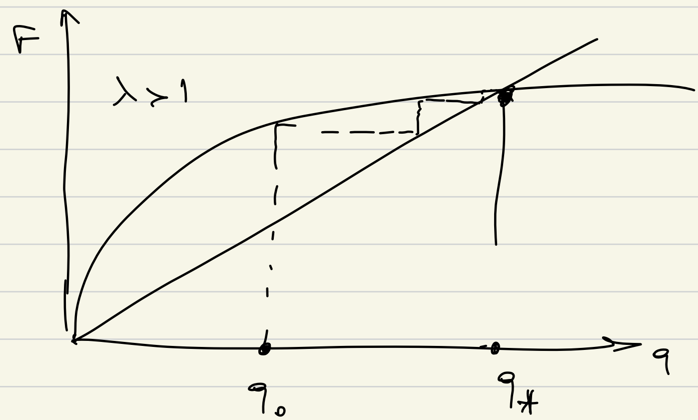


$\lambda \leq 1$: only one fixed point $q_* = 0$

$\lambda > 1$: Two fixed points 0 and $q_* > 0$.

AMP fails at estimating x_0 if $\lambda \leq 1$.

on the other hand if $\lambda > 1$ the it depends on initialization. q_0



Let's say $y_i = \sqrt{\varepsilon}\, x_{0i} + z_i$ $\quad \varepsilon > 0$

Initialize $\hat{x}_i^0 = \mathbb{E}[x_{0i} \mid y_i]$. $\forall\ 1 \leq i \leq n$.

$$q_0 = 2\psi'(\varepsilon) > 0.$$

overlap under AMP $\longrightarrow$ $q_* =$ Best overlap possible.

First remark : if $\lambda_c < 1$ then AMP fails in $[\lambda_c, 1]$

but it is information theoretically possible to estimate x_0 non trivially.

- Second remark: Initialize using first eigvec of Y. u: first eigvec. Montanari–Venkataramanan. 2019.

Let $x^0 = \sqrt{n(\lambda^2-1)} \cdot u$.

$$q_0 = \cancel{1} - \frac{1}{\lambda} > 0 \quad \text{if } \lambda > 1.$$