11/29

Lecture 23: Approximate Message Passing

State evolution: $\forall$ nice $\psi : \mathbb{R}^2 \to \mathbb{R}$, $t \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(\underbrace{x_t^i}_{t\text{th iterate of AMP}}, \underbrace{X^i}_{\text{signal}}) = \mathbb{E}_{\substack{x\sim\{\pm 1\}\\ g\sim N(0,1)}}\left[\psi(\mu_t x + \sigma_t g, x)\right]$$

$$\mu_t = \sqrt{\lambda}\,\sigma_t^2, \quad \sigma_t^2 = \gamma_t/\lambda, \quad \gamma_{t+1} = \lambda(1 - \mathrm{mmse}(\gamma_t))$$

Lemma 1: $\mathrm{MSE}_{AMP}(t;\lambda) = 1 - \frac{\gamma_t^2}{\lambda^2}$.

Pf:

$$\mathrm{MSE}_{AMP}(t;\lambda,n) = \frac{1}{n^2}\mathbb{E}\left[\| XX^T - \hat{x}^t(\hat{x}^t)^T\|_F^2\right]$$

$$= \underbrace{\frac{\mathbb{E}[\|X\|_2^4]}{n^2}}_{1} - 2\cdot\underbrace{\frac{\mathbb{E}[\langle \hat{x}^t, X\rangle^2]}{n^2}}_{\text{I}} + \underbrace{\frac{\mathbb{E}\|x^t\|_2^4}{n^2}}_{\text{II}}$$

for I, we can use state evolution.

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i(\hat{x}^t)_i = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i f_{t-1}(x_i^t)$$

$$= \mathbb{E}_{X,\xi}\left[X f_{t-1}(\mu_{t-1}X + \sigma_{t-1}\xi)\right]$$

$$= \mu_t/\sqrt{\lambda} \quad (\text{by defn of } \mu_t)$$

$$= \gamma_t/\lambda,$$

So $\lim_{n\to\infty}\frac{1}{n^2}\mathbb{E}[\langle \hat{x}^t, X\rangle^2] = \frac{\gamma_t^2}{\lambda^2}$

Similar calculation for (II):

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\hat{x}_t)_i^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_{t-1}(x_i^t)^2$$

$$= \mathbb{E}\left[\mathbb{E}[X \mid \mu_{t-1} X + \sigma_{t-1} Z]^2 \right]$$

$$= \sigma_t^2 \qquad \text{(by defn of } \sigma_t^2\text{)}$$

$$= \gamma_t / \lambda,$$

so $$\lim_{n\to\infty} \frac{1}{n^2} \mathbb{E}\left[\|\hat{x}^t\|_2^4 \right] = \frac{\gamma_t^2}{\lambda^2}. \qquad \square$$

Lemma 2: $\mathrm{MMSE}(\lambda) = \lim_{t\to\infty} \mathrm{MSE}_{AMP}(t;\lambda)$

Pf: We will use the following:

Fact (I-MMSE relation):

$Y(\lambda) \doteq \sqrt{\frac{\lambda}{n}} XX^T + W$

$$\frac{1}{n} \cdot \frac{\partial}{\partial \lambda} \underbrace{I(XX^T; Y(\lambda))}_{\text{mutual information between signal and observation}} = \frac{1}{4} \mathrm{MMSE}(\lambda, n)$$

Note

when $\lambda = 0$, no information about signal present

$$\lim_{n\to\infty} \frac{1}{n} I(XX^T; Y(0)) = 0$$

when $\lambda = \infty$, there is full information about signal

$$\lim_{\lambda\to\infty} \lim_{n\to\infty} \frac{1}{n} I(XX^T; Y(\lambda)) = \lg 2, \quad \text{so}$$

$$\lg 2 = \lim_{\lambda\to\infty} \lim_{n\to\infty} \frac{1}{n} \left[I(XX^T; Y(\lambda)) - I(XX^T; Y(0)) \right]$$

(I-MMSE)

$$= \lim_{n\to\infty} \frac{1}{4} \int_0^\infty MMSE(\lambda, n)\, d\lambda$$

(Bayes optimality)

$$\leq \lim_{t\to\infty} \limsup_{n\to\infty} \frac{1}{4} \int_0^\infty MSE_{AMP}(t; \lambda, n)\, d\lambda$$

(dominated convergence thm).

$$= \frac{1}{4} \int_0^\infty \left(1 - \frac{\gamma_*(\lambda)^2}{\lambda^2}\right) d\lambda$$

If can show

$$\frac{1}{4} \int_0^\infty \left(1 - \frac{\gamma_*(\lambda)^2}{\lambda^2}\right) d\lambda = \lg 2,$$

then $\leq$ above is an equality, so AMP is Bayes-optimal as desired.

(we will show this is the antiderivative of MSE_{AMP} w.r.t. λ)

Define $\psi(\gamma, \lambda) \triangleq \frac{\lambda}{4} + \frac{\gamma^2}{4\lambda} - \frac{\gamma}{2} + I(\gamma)$, where $I(\gamma) \triangleq I(X; \sqrt{\gamma}X + \xi)$.

Note that by design,

$$\frac{\partial}{\partial \lambda}\Psi(\gamma,\lambda) = \frac{1}{4}\left(1 - \frac{\gamma^2}{\lambda^2}\right)$$

(I-MMSE for scalar denoising problem)

$$\frac{\partial}{\partial \Psi}\Psi(\gamma,\lambda) = \frac{\gamma}{2\lambda} - \frac{1}{2} + \frac{1}{2}\mathrm{mmse}(\gamma)$$

$$= 0 \quad \text{if} \quad \gamma = \gamma_*(\lambda).$$

so $$\underline{\underline{\frac{\partial}{\partial \lambda}\Psi(\gamma_*(\lambda),\lambda) = \frac{1}{4}\left(1 - \frac{\gamma^2}{\lambda^2}\right)}}$$

Furthermore, $\lim_{\lambda\to 0}\Psi(\gamma_*(\lambda),\lambda) = 0$ as $\gamma_*(\lambda) \leq \lambda \to 0$ and $I(\gamma_*(\lambda)) \to 0$, so

$$\frac{1}{4}\int_0^\infty \left(1 - \frac{\gamma_*(\lambda)^2}{\lambda^2}\right) = \lim_{\lambda\to\infty}\Psi(\gamma_*(\lambda),\lambda)$$

$\gamma^*(\lambda) \to \lambda$ as $\lambda \to \infty$, so

$$\frac{\lambda}{4} + \frac{\gamma^*(\lambda)^2}{4\lambda} - \frac{\gamma^*(\lambda)}{2} \to 0.$$

Furthermore, $I(\gamma'(\lambda)) \to \lg 2$, so the proof is complete. $\square$

Proof of I-MMSE relation:

Let us consider a more general setting where $\mathbb{X}$ is some random vector in $\mathbb{R}^N$, and we observe $\mathbb{Y} = \sqrt{\lambda}\,\mathbb{X} + g$ for $g \sim N(0, \mathbb{I})$.

$$\mathrm{MMSE}(\lambda) \triangleq \mathbb{E}\left[\|\mathbb{X} - \mathbb{E}(\mathbb{X} \mid \mathbb{Y})\|^2\right] \quad \text{(unnormalized)}.$$

Mutual information

Joint dist. of $\mathbb{X}, \mathbb{Y}$

$$I(\mathbb{X}; \mathbb{Y}) \triangleq KL\left(P_{X,Y} \,\|\, P_X \otimes P_Y\right)$$

product of marginal dist.'s

$$= \mathbb{E}_{\mathbb{X},\mathbb{Y}}\left[\lg \frac{dP_{X,Y}}{dP_X \otimes P_Y}(\mathbb{X}, \mathbb{Y})\right]$$

$$= \mathbb{E}_{\mathbb{X},\mathbb{Y}}\left[\lg \left\{ \frac{\exp\left(-\frac{1}{2}\|\mathbb{Y} - \sqrt{\lambda}\,\mathbb{X}\|^2\right)}{\int dP_X(x)\, \exp\left(-\frac{1}{2}\|\mathbb{Y} - \sqrt{\lambda}\,x\|^2\right)} \right\}\right]$$

$$= -\mathbb{E}_{X,Y} \lg \int dP_X(x) \exp\left(\sqrt{\lambda}\langle x, Y\rangle - \sqrt{\lambda}\langle X, Y\rangle - \frac{\lambda}{2}\|x\|^2 + \frac{\lambda}{2}\|X\|^2\right)$$

$$= \frac{\lambda}{2}\mathbb{E}\|X\|^2 - \underbrace{\mathbb{E}_{X,Y} \lg \overbrace{\int dP_X(x) \exp\left(\sqrt{\lambda}\langle x, Y\rangle - \frac{\lambda}{2}\|x\|^2\right)}^{\text{partition function for Gibbs measure } \mu = \Pr[X = \cdot \mid Y]}}_{\triangleq F(\lambda) \text{ (free energy)}}$$

w.t.s $\frac{\partial}{\partial\lambda} F(\lambda) = \frac{1}{2}\left(\mathbb{E}\|X\|^2 - \mathrm{MMSE}(\lambda)\right)$

$$\left(= \frac{1}{2}\mathbb{E}\|\mathbb{E}[X \mid Y]\|^2\right)$$

$$\frac{\partial}{\partial\lambda} F(\lambda) = \underbrace{\mathbb{E}_{X,g}\mathbb{E}_{x\sim\mu}\left[\langle x, X\rangle\right.}_{\mathbb{E}\|\mathbb{E}[X \mid Y]\|^2} \left. + \frac{1}{2\sqrt{\lambda}}\langle x, g\rangle - \frac{1}{2}\|x\|^2\right]$$

$$\mathbb{E}_X \mathbb{E}_g \mathbb{E}_{x\sim\mu}\left[\frac{1}{2\sqrt{\lambda}}\langle x, g\rangle\right]$$

(Gaussian int. by parts) $= \mathbb{E}_X \mathbb{E}_g \operatorname{div}_g\left\{\mathbb{E}_{x\sim\mu}\left[\frac{1}{2\sqrt{\lambda}}x\right]\right\}$

Note, for any $i \in [N]$,

$$\frac{\partial}{\partial g_i} \mathbb{E}_{x \sim \mu}[x_i] = \sqrt{\lambda}\left(\mathbb{E}_{x\sim\mu}[x_i^2] - \mathbb{E}_{x\sim\mu}[x_i]^2\right)$$

(see prop. below)

so $\mathbb{E}_{\mathbb{X}, g} \operatorname{div}_g \left\{ \mathbb{E}_{x\sim\mu}\left(\frac{1}{2\sqrt{\lambda}} x\right) \right\}$

$$= \frac{1}{2} \mathbb{E}_{\mathbb{X}, g} \mathbb{E}_{x\sim\mu} \|x\|^2 - \frac{1}{2} \underbrace{\mathbb{E}_{\mathbb{X}, g} \|\mathbb{E}_{x\sim\mu}[x]\|^2}_{= \mathbb{E}\|\mathbb{E}[\mathbb{X} \mid \mathbb{Y}]\|^2}$$

so

$$\frac{\partial}{\partial \lambda} F(\lambda) = \frac{1}{2} \mathbb{E}\,\|\mathbb{E}[\mathbb{X} \mid \mathbb{Y}]\|^2 \text{ as}$$

desired

<u>Prop</u>: For $\mu = \Pr[\mathbb{X} = \cdot \mid \mathbb{Y}]$,

$$\frac{\partial}{\partial g_i} \mathbb{E}_{x\sim\mu}[x_i] = \sqrt{\lambda}\left(\mathbb{E}_{x\sim\mu}[x_i^2] - \mathbb{E}_{x\sim\mu}[x_i]^2\right)$$

f'

Pf: Note

$$\mathbb{E}_{x\sim\mu}\left[x_i\right] = \frac{\int e^{-\frac{\lambda}{2}\|x\|^2 + \sqrt{\lambda}\langle x, \mathbb{Y}\rangle} x_i \, dx}{\int e^{-\frac{\lambda}{2}\|x\|^2 + \sqrt{\lambda}\langle x, \mathbb{Y}\rangle} dx}$$

$$= \frac{\int e^{-\frac{\lambda}{2}\|x\|^2 + \langle x, \mathbb{X}\rangle + \sqrt{\lambda}\langle x, g\rangle} x_i \, dx}{\int e^{-\frac{\lambda}{2}\|x\|^2 + \langle x, \mathbb{X}\rangle + \sqrt{\lambda}\langle x, g\rangle} dx},$$

So

$$\frac{\partial}{\partial g_i}\mathbb{E}_{x\sim\mu}\left[x_i\right] = \frac{\sqrt{\lambda}\int e^{-\frac{\lambda}{2}\|x\|^2 + \langle x, \mathbb{X}\rangle + \sqrt{\lambda}\langle x, g\rangle} x_i^2 \, dx}{\int e^{-\frac{\lambda}{2}\|x\|^2 + \langle x, \mathbb{X}\rangle + \sqrt{\lambda}\langle x, g\rangle} dx} - \frac{\sqrt{\lambda}\left(\int e^{-\frac{\lambda}{2}\|x\|^2 + \langle x, \mathbb{X}\rangle + \sqrt{\lambda}\langle x, g\rangle} x_i \, dx\right)^2}{\left(\int e^{-\frac{\lambda}{2}\|x\|^2 + \langle x, \mathbb{X}\rangle + \sqrt{\lambda}\langle x, g\rangle} dx\right)^2}$$

$$= \sqrt{\lambda}\left(\mathbb{E}_{\mu}\left[x_i^2\right] - \mathbb{E}_{\mu}\left[x_i\right]^2\right). \qquad \square$$