Mean Field Asymptotics in Statistical Learning. Mar 3.

Lecture 12: The LASSO risk.

① The LASSO problem.

Signal: $x_0 \in \mathbb{R}^d$, Noise: $w \in \mathbb{R}^n$, Sensing matrix: $A \in \mathbb{R}^{n \times d}$.

Response: $y = Ax_0 + w \in \mathbb{R}^n$.

Assumption 1: $A_{ij} \sim_{iid} N(0, \frac{1}{n})$, $w_i \sim_{iid} N(0, \sigma^2)$, $\frac{1}{d}\sum_{j=1}^d \delta_{x_{0,j}} \to \mathbb{P}_0$ (potentially sparse).

Goal: Observe y and A, estimate x_0.

LASSO estimator:

$$\hat{x} = \arg\min_x \frac{1}{2n}\|y - Ax\|_2^2 + \frac{\lambda}{n}\|x\|_1.$$

② Main result (to be shown using the replica method today).

Theorem: Consider the asymptotic regime $n/d \to \delta \in (0,\infty)$ as $d \to \infty$. Let (A, x_0, w, y) be a sequence of problem instances satisfy Assumption 1. Let $\hat{x}$ be the LASSO estimator.

(A) We have $\to$ $\mathbb{E}$ is w.r.t. A, w.

$$\lim_{\substack{d,n\to\infty \\ n/d \to \delta}} \mathbb{E}\|\hat{x} - x_0\|_2^2/d = \mathbb{E}_{(X_0,G)\sim \mathbb{P}_0 \times N(0,1)}\left[(\eta(X_0 + \tau_* G; \tau_*\alpha_*) - X_0)^2\right]$$

where $\eta(x;s) = \mathrm{sign}(x)\cdot(|x| - s)_+$ is the soft thresholding function. $\leftarrow$ shrink to 0.

Here, we define a function $\tau(\alpha)$ to be the largest solution of

$$\tau^2 = \sigma^2 + \delta^{-1}\mathbb{E}_{(X_0,G)\sim\mathbb{P}_0\times N(0,1)}\{[\eta(X_0 + \tau G; \alpha\tau) - X_0]^2\}.$$

and denote α_* by the unique non-negative solution of

$$\lambda = \alpha\cdot\tau(\alpha)\cdot\{1 - \delta^{-1}\mathbb{E}[\eta'(X_0 + \tau(\alpha)G; \alpha\cdot\tau(\alpha))]\}$$

Self consistent equation.

Hereby we define $\tau_* = \tau(\alpha_*)$.

$X_0 \sim \mathbb{P}_0$,
$Y = X_0 + \tau_* G$
$\hat{X} = \eta(Y; \alpha_*\tau_*)$

(B) Moreover, for any pseudo-Lipschitz function $\psi: \mathbb{R}^2 \to \mathbb{R}$, we have

$$\lim_{\substack{d\to\infty \\ n/d\to\delta}} \mathbb{E}\Big[\frac{1}{d}\sum_{j=1}^d \psi(\hat{x}_j, x_{0,j})\Big] = \mathbb{E}_{(X_0,G)\sim\mathbb{P}_0\times N(0,1)}[\psi(\underbrace{\eta(X_0 + \tau_* G; \alpha_*\tau_*)}_{\text{The solution of a one dim problem}}; X_0)]$$

Heuristic result: [Guo, Baron, Shamai, 2009], [Kabashima, Wadayama, Tanaka, 2009]

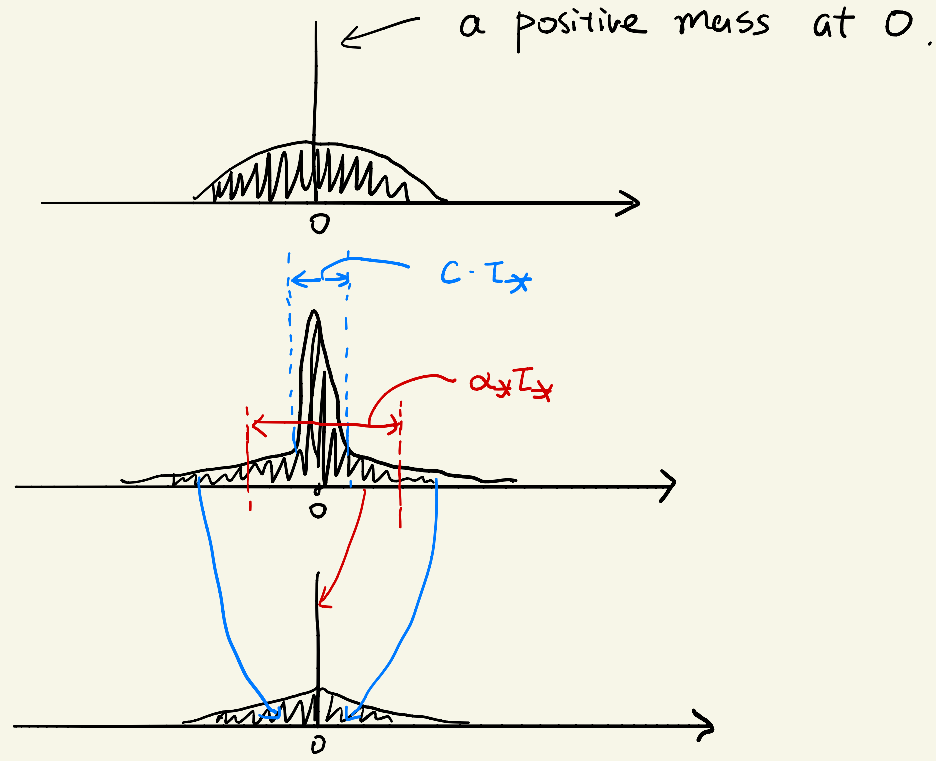
Rigorous proof: [Bayati, Montanari, 2010] $\leftarrow$ AMP

Alterative proof: CGMT, leave-one-out.

X_0 distribution

$Y = X_0 + \tau_* G$

$\hat{X} = \eta(Y; \alpha_* \tau_*)$



Application: Want to estimate the support of x_0

$$S = \{ j : x_{0,j} \neq 0 \},$$

using the support of Lasso solution

$$\hat{S}(\lambda) = \{ j : \hat{x}_{\lambda,j} \neq 0 \}.$$

False discovery proportion:

$$\mathrm{FDP}(\lambda) = \frac{|\{ j : \hat{x}_{\lambda,j} \neq 0,\ x_{0,j} = 0 \}|}{|\{ j : \hat{x}_{\lambda,j} \neq 0 \}|}$$

True positive proportion:

$$\mathrm{TPP}(\lambda) = \frac{|\{ j : \hat{x}_{\lambda,j} \neq 0,\ x_{0,j} \neq 0 \}|}{|\{ j : x_{0,j} \neq 0 \}|}$$

Goal: choose a λ, s.t $\underbrace{\mathbb{E}[\mathrm{FDP}(\lambda)]}_{\text{FDR}} \leq \alpha$, $\underbrace{\mathbb{E}[\mathrm{TPP}(\lambda)]}_{\text{Power}}$ as large as possible.

Under the model assumption ($A_{ij} \overset{iid}{\sim} N(0, \frac{1}{n})$, $w_i \overset{iid}{\sim} N(0, \sigma^2)$)

$$\frac{1}{d} |\{ j : \hat{x}_{\lambda,j} \neq 0 ,\ x_{0,j} = 0 \}| = \frac{1}{d} \sum_{j=1}^{d} \mathbb{1}\{ \hat{x}_{\lambda,j} \neq 0 ,\ x_{0,j} = 0 \}$$

$$\to \mathbb{P}(X_0 = 0,\ \eta(X_0 + \tau_* G ; \alpha_* \tau_*) \neq 0).$$

$X_0 \sim \mathbb{P}_0 \perp G \sim N(0,1)$

$$\frac{1}{d} |\{ j : \hat{x}_{\lambda,j} \neq 0 \}| = \frac{1}{d} \sum_{j=1}^{d} \mathbb{1}\{ \hat{x}_{\lambda,j} \neq 0 \}$$

$$\to \mathbb{P}(\eta(X_0 + \tau_* G ; \alpha_* \tau_*) \neq 0).$$

$$\Rightarrow \lim_{d \to \infty} \mathbb{E}[\mathrm{FDP}(\lambda)] = \frac{\mathbb{P}(X_0 = 0,\ \eta(X_0 + \tau_* G ; \alpha_* \tau_*) \neq 0)}{\mathbb{P}(\eta(X_0 + \tau_* G ; \alpha_* \tau_*) \neq 0)}.$$

Similarly for $\mathrm{TPP}(\lambda)$.

This gives a "simple" expression for the power.

③. The free energy trick and the replica trick

The configuration space: $\Omega = \mathbb{R}^d$, $\nu_0 =$ Lebesgue measure

Hamiltonian : $H_{\lambda,h}(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 + h\sum_{j=1}^{d}\psi(x_j, x_{0,j})$

Randomness : $y = \underline{A}\, x_0 + \underline{w}$ 2 source of randomness.

free entropy density: $\varphi(\beta,\lambda,h) \equiv \lim_{d\to\infty}\left[\frac{1}{d}\log\int_{\mathbb{R}^d}\exp\{-\beta H_{\lambda,h}(x)\}dx\right]$

Limiting ensemble average:

The free energy trick.

$$\lim_{d\to\infty}\mathbb{E}\left[\frac{1}{d}\sum_{j=1}^{d}\psi(\hat{x}_j, x_{0,j})\right] = \partial_h \lim_{\beta\to\infty}\left[-\frac{1}{\beta}\varphi(\beta,\lambda,h)\right]\Big|_{h=0}$$

(A) The $d\to\infty$ limit.

$$S(k,\beta,\lambda,h) \equiv \lim_{d\to\infty}\frac{1}{d}\log\mathbb{E}\left[\left(\int_{\mathbb{R}^d}\exp\{-\beta H_{\lambda,h}(x)\}dx\right)^k\right].$$

(B) The $k\to 0$ limit.

$$\varphi(\beta,\lambda,h) = \lim_{k\to 0}\left[\frac{1}{k}S(k,\beta,\lambda,h)\right].$$

(C) The h differentiation and $\beta\to\infty$ limit

$$m_*(\lambda) = \left\{\partial_h \lim_{\beta\to\infty}\left[-\frac{1}{\beta}\varphi(\beta,\lambda,h)\right]\right\}\Big|_{h=0}.$$

④ The calculation

(A) The $d\to\infty$ limit. $(n/d \to \delta)$.

$X^0 = x_0 \in \mathbb{R}^d$

w.r.t. A, w

$$\mathbb{E}[Z_n^k] = \mathbb{E}\int_{\mathbb{R}^{d\times k}}\exp\left\{-\beta\sum_{a=1}^{k}\|y - Ax^a\|_2^2/2 - \beta\lambda\sum_{a=1}^{k}\|x^a\|_1 - \beta h\sum_{j=1}^{d}\sum_{a=1}^{k}\psi(x_j^a, x_j^0)\right\}\prod_{a=1}^{k}dx^a$$

$$= \int_{\mathbb{R}^{d\times k}}\underbrace{\mathbb{E}\left[\exp\left\{-\beta\sum_{a=1}^{k}\|y - Ax^a\|_2^2/2\right\}\right]}_{E(x)}$$

w.r.t. A, w

$$\times\exp\left\{-\beta\lambda\sum_{a=1}^{k}\|x^a\|_1 - \beta h\sum_{j=1}^{d}\psi(x_j^a, x_j^0)\right\}\prod_{a=1}^{k}dx^a$$

w.r.t. A, w

$$E(x) = \mathbb{E}\left[\exp\left\{-\beta\sum_{a=1}^{k}\|y - Ax^a\|_2^2/2\right\}\right]$$

$$\mathbb{E}_G\left[\exp\{-G^T\Sigma^{-1}G/2\}\right] = \det(\Sigma)^{-\frac{1}{2}}$$

$$= \mathbb{E}\left[\exp\left\{-\beta\sum_{a=1}^{k}\|w - A(x^0 - x^a)\|_2^2/2\right\}\right]$$

$$= \mathbb{E}\left[\exp\left\{-\beta k\|w\|_2^2/2 + \beta\left\langle w, \sum_{a=1}^{k}A(x^0 - x^a)\right\rangle - \frac{\beta}{2}\sum_{a=1}^{k}\|A(x^0 - x^a)\|_2^2/2\right\}\right]$$

density of w

w.r.t. A

ε is dummy for w

$$= \mathbb{E}\int_{\mathbb{R}^n}\frac{1}{\sqrt{2\pi}^n\sigma^n}\exp\left\{-\beta k\|\varepsilon\|_2^2/2 - \frac{\|\varepsilon\|_2^2}{2\sigma^2} + \beta\left\langle\varepsilon, \sum_{a=1}^{k}A(x^0 - x^a)\right\rangle\right\}d\varepsilon$$

Gaussian integration

$$\times\exp\left\{-\frac{\beta}{2}\sum_{a=1}^{k}\|A(x^0 - x^a)\|_2^2\right\}\Big].$$

$$\boxed{\bar\sigma^2 = \frac{\sigma^2}{1+\beta k\sigma^2}.}$$

$$= \underbrace{(k\beta\sigma^2+1)^{-\frac{n}{2}}}_{C_n}\times\mathbb{E}\left[\exp\left\{\beta^2\left\|\sum_{a=1}^{k}A(x^0 - x^a)\right\|_2^2\bar\sigma^2/2 - \frac{\beta}{2}\sum_{a=1}^{k}\|A(x^0 - x^a)\|_2^2/2\right\}\right].$$

Note $A = \begin{bmatrix}u_1^T\\ \vdots \\ u_n^T\end{bmatrix}$, $u_i \sim_{iid} N(0, \frac{1}{n}I_d) \Rightarrow \|A(x^0 - x^a)\|_2^2 = \sum_{i=1}^{n}\left(u_i^T(x^0 - x^a)\right)^2$

w.r.t. $u \sim N(0, \frac{1}{n}I_d)$

$$= C_n\cdot\mathbb{E}\left[\exp\left\{\beta^2\left(\sum_{a=1}^{k}u^T(x^0 - x^a)\right)^2\bar\sigma^2/2 - \frac{\beta}{2}\sum_{a=1}^{k}\left[u^T(x^0 - x^a)\right]^2\right\}\right]^n$$

Define $G_a = u^T(x^0 - x^a)$, Then $\mathbb{E}[G_aG_b] = (x^0 - x^a)^T(x^0 - x^b)/n \equiv \Sigma_{ab}(x)$

$$= C_n\cdot\left[\int\frac{1}{(2\pi)^{\frac{k}{2}}\det(\Sigma)^{\frac{1}{2}}}\exp\left\{\beta^2\left(\sum_{a=1}^{k}G_a\right)^2\bar\sigma^2/2 - \frac{\beta}{2}\sum_{a=1}^{k}G_a^2 - \underbrace{G^T\Sigma^{-1}G/2}_{\text{density of } G}\right\}dG\right]^n$$

Gaussian integration

$$= C_n\cdot\left[\det\left(\Sigma^{-1} + \beta I_k - \beta^2\bar\sigma^2 1_k 1_k^T\right)\det(\Sigma)\right]^{-\frac{n}{2}}$$

$$= (1 + k\beta\sigma^2)^{-\frac{n}{2}}\left\{\det\left[I_k + \Sigma\left(\beta I_k - \beta^2\bar\sigma^2 1_k 1_k^T\right)\right]\right\}^{-\frac{n}{2}}$$

Define $\bar Q(x) = \left(\langle x^a, x^b\rangle/d\right)_{a,b\in[k]}$.

$\bar\mu(x) = \left(\langle x^a, x^0\rangle/d\right)_{a\in[k]}$.

$P = \|x^0\|_2^2/d$

$$\Rightarrow \Sigma(x) = \bar Q(x) - \bar\mu(x)1_k^T + 1_k\bar\mu(x)^T + P\,1_k1_k^T.$$

$$\Rightarrow E(x) = E\left(\bar Q(x), \bar\mu(x)\right).$$ ← abuse notation

introduce $1 = \int\delta(\bar Q - Q)\,\delta(\bar\mu - \mu)\,dQ\,d\mu.$

$$\Rightarrow \mathbb{E}[Z_n^k] = \int_{\mathbb{R}^{d\times k}} E\left(\bar Q(x), \bar\mu(x)\right)$$

$$\times\exp\left\{-\beta\lambda\sum_{a=1}^{k}\|x^a\|_1 - \beta h\sum_{j=1}^{d}4(x_j^a, x_j^0)\right\}\prod_{a=1}^{k}dx^a$$

$$= \int dQ\,d\mu\;E(Q,\mu)\times \mathrm{Ent}(Q,\mu)$$

$$S(k,\beta,\lambda,h) \doteq \sup_{Q,\mu}\left[\frac{1}{d}\log E(Q,\mu)\right]\times\left[\frac{1}{d}\log\mathrm{Ent}(Q,\mu)\right].$$

where $\mathrm{Ent}(Q,\mu) = \int \delta(\bar{Q}(x) - Q)\,\delta(\bar{\mu}(x) - \mu)$

$$\times \exp\Big\{-\beta\lambda \sum_{a=1}^{k} \|x^a\|_1 - \beta h \sum_{j=1}^{d} \psi(x_j^a, x_j^0)\Big\} \prod_{a=1}^{k} dx^a$$

$\boxed{\delta = n/d}$

$$\underbrace{\frac{1}{d}\log E(Q,\mu)}_{e(Q,\mu)} = -\frac{\delta}{2}\log(1+k\beta\sigma^2) - \frac{\delta}{2}\log\det\Big[I_k + (Q - \mu 1^T + 1\mu^T + p\,1 1^T)(\beta I_k - \beta^2\tilde{\sigma}^2 1 1^T)\Big].$$

$$\tilde{\sigma}^2 = \frac{\sigma^2}{1+k\beta\tilde{\sigma}^2}$$

By delta identity formula + saddle point approx :

$$\mathrm{Ent}(Q,\mu) = \operatorname*{ext}_{r_{ab},\xi_a} \int_{\mathbb{R}^{dk}} \exp\Big\{-\sum_{ab} r_{ab}\,(d\cdot Q_{ab} - \langle x^a, x^b\rangle)/2 - \sum_a \xi_a\,(d\cdot\mu_a - \langle x^a, x^0\rangle)$$

$$-\beta\lambda\sum_{a=1}^{k}\|x^a\|_1 - \beta h \sum_{j=1}^{d}\psi(x_j^a, x_j^0)\Big\} \prod_{a=1}^{k} \underset{\in\mathbb{R}^d}{dx^a}$$

$$= \operatorname*{ext}_{r_{ab},\xi_a}\Bigg[\prod_{j=1}^{d}\Bigg(\int_{\mathbb{R}^k}\exp\Big\{-\sum_{ab} r_{ab}(Q_{ab} - x^a x^b)/2 - \sum_a \xi_a(\mu_a - x^a x_j^0)$$

$$-\beta\lambda\sum_{a=1}^{k}|x^a| - \beta h\,\psi(x^a, x_j^0)\Big\}\prod_{a=1}^{k}\overset{\in\mathbb{R}}{dx^a}\Bigg)\Bigg]$$

Law of large numbers, uniform in r_{ab}, ξ_a

$\frac{1}{d}\log \mathrm{Ent} = \operatorname*{ext}_{r} \frac{1}{d}\sum_{j=1}^{d}\log M(x_j^0, r) \to \operatorname*{ext}_{r} \mathbb{E}_{x^0\sim P_0}[\log M(x^0, r)].$

$$\Rightarrow \frac{1}{d}\log\mathrm{Ent} \to \operatorname*{ext}_{r_{ab},\xi_a}\mathbb{E}_{x^0}\Bigg[\log\int_{\mathbb{R}^k}\exp\Big\{-\sum_{ab} r_{ab}(Q_{ab} - x^a x^b)/2 - \sum_a \xi_a(\mu_a - x^a x^0)$$

$$-\beta\lambda\sum_{a=1}^{k}|x^a| - \beta h\,\psi(x^a, x^0)\Big\}\prod_{a=1}^{k} dx^a\Bigg]$$

$$\Rightarrow S(k,\beta,\lambda,h) = \operatorname*{ext}_{\substack{Q,\mu\\ r,\xi}}\Big[e(Q,\mu) + \mathrm{ent}(Q,\mu,r,\xi)\Big].$$

$\boxed{Q, r\in\mathbb{R}^{k\times k} \quad \mu,\xi\in\mathbb{R}^k}$

where

$$e(\cdot) = -\frac{\delta}{2}\log(1+k\beta\sigma^2) - \frac{\delta}{2}\log\det\Big[I_k + (Q - \mu 1^T + 1\mu^T + p\,1 1^T)(\beta I_k - \beta^2\tilde{\sigma}^2 1 1^T)\Big].$$

$$\tilde{\sigma}^2 = \frac{\sigma^2}{1+k\beta\tilde{\sigma}^2}$$

$$\mathrm{ent}(\cdot) = \mathbb{E}_{x^0}\Bigg[\log\int_{\mathbb{R}^k}\exp\Big\{-\sum_{ab} r_{ab}(Q_{ab} - x^a x^b)/2 - \sum_a \xi_a(\mu_a - x^a x^0)$$

$$-\beta\lambda\sum_{a=1}^{k}|x^a| - \beta h\,\psi(x^a, x^0)\Big\}\prod_{a=1}^{k} dx^a\Bigg]$$

(B) The $k\to 0$ limit.

Replica symmetric ansatzs.

$$Q = \begin{bmatrix} q_1 & & q_0 \\ & \ddots & \\ q_0 & & q_1\end{bmatrix},\quad r = \begin{bmatrix} r_1 & & r_0 \\ & \ddots & \\ r_0 & & r_1\end{bmatrix},\quad \mu = \begin{bmatrix}\mu\\ \vdots \\ \mu\end{bmatrix},\quad \xi = \begin{bmatrix}\xi\\ \vdots\\ \xi\end{bmatrix}.$$

Consider that later we'll take $\beta \to \infty$.

do reparametrization here: will be clear later.

$$q_0 = q, \qquad q_1 = q + \frac{w}{\beta}, \qquad r_0 = \beta^2 \rho, \qquad r_1 = \beta^2 \rho - \beta \nu$$

$$\mu = \mu, \qquad \xi = \beta \zeta$$

$$\Rightarrow \quad \lim_{k \to 0} \frac{1}{k} e = -\frac{\delta}{2} \log\left(1 + \frac{w}{\delta}\right) - \frac{\beta \delta}{2(\delta + w)} \left(\rho - 2\mu + q + \delta \sigma^2\right) \equiv \bar{e}$$

Need to use $\log\det\left(a I_k + b \mathbf{1}\mathbf{1}^T\right) = (k-1)\log b + \log(a + kb)$

and

$$\text{ent} = -\left[\beta^2 k^2 \rho q - \beta k \nu q + \beta k \rho w - k \nu w\right]/2 - k \beta \mu \zeta$$
$$+ \mathbb{E}_{x^0}\left[\log \int_{\mathbb{R}^k} \exp\left\{ \underbrace{\frac{\beta^2 \rho}{2} \cdot \left(\sum_{a=1}^k x^a\right)^2} - \frac{\beta \nu}{2} \sum_{a=1}^k (x^a)^2 + \beta \zeta \sum_{a=1}^k x^a x^0 \right.\right.$$
$$\left.\left. - \beta \lambda \sum_{a=1}^k |x_a| - \beta h \, \psi(x^a, x^0) \right\} \prod_{a=1}^k dx^a \right]$$

$$\boxed{\text{use} \quad \mathbb{E}\left[e^{\lambda\left(\sum_{a=1}^k x^a\right) G}\right] = \exp\left\{\frac{\lambda^2}{2}\left(\sum_{a=1}^k x^a\right)^2\right\}}$$

$$= -\left[\beta^2 k^2 \rho q - \beta k \nu q + \beta k \rho w - k \nu w\right]/2 - k \beta \mu \zeta$$
$$+ \mathbb{E}_{x^0}\left\{\log \mathbb{E}_G\left[\int_{\mathbb{R}^k} \exp\left\{ \underbrace{\beta \sqrt{\rho} \cdot \sum_{a=1}^k x^a G} - \beta \nu \sum_{a=1}^k (x^a)^2 / 2 + \beta \zeta \sum_{a=1}^k x^a x^0 \right.\right.\right.$$
$$\left.\left.\left. - \beta \lambda \sum_{a=1}^k |x_a| - \beta h \, \psi(x^a, x^0) \right\} \prod_{a=1}^k dx^a \right]\right\}$$

$$= -\left[\beta^2 k^2 \rho q - \beta k \nu q + \beta k \rho w - k \nu w\right]/2 - k \beta \mu \zeta$$
$$+ \mathbb{E}_{x^0} \log \mathbb{E}_G \left[\left(\int_{\mathbb{R}} \exp\left\{\beta \sqrt{\rho} \cdot x \cdot G - \frac{\beta \nu}{2} \cdot x^2 + \beta \cdot \zeta \cdot x \cdot x^0 - \beta \lambda |x| - \beta h \, \psi(x, x^0)\right\} dx\right)^k\right]$$

$$\boxed{\lim_{k \to 0} \frac{1}{k} \log \mathbb{E}\left[Z(x)^k\right] = \mathbb{E}\left[\log Z(x)\right]}$$

$$\Rightarrow \quad \lim_{k \to 0} \frac{1}{k} \text{ent} = -\left[-\beta \chi q + \beta \rho w - \nu w\right]/2 - \beta \mu \zeta$$
$$+ \mathbb{E}_{G, x^0}\left[\log \int_{\mathbb{R}} \exp\left\{\overset{\sqrt{r_0}}{\boxed{\beta}\sqrt{\rho}} \cdot x \cdot G - \overset{(r_0 - r_1)}{\frac{\boxed{\beta}\nu}{2}} \cdot x^2 + \boxed{\beta} \zeta \cdot x \cdot x^0 - \boxed{\beta} \lambda |x| - \boxed{\beta} h \, \psi(x, x^0)\right\} dx\right] \equiv \overline{\text{ent}}$$

$$\varphi(\beta, \lambda, h) = \underset{\substack{q, w, \mu \\ \rho, \nu, \zeta}}{\text{ext}} \left[\bar{e} + \overline{\text{ent}}\right]$$

$$\boxed{\begin{aligned} &\lim_{\beta \to \infty} \frac{1}{\beta} \mathbb{E}\left[\log \int_{\mathbb{R}} \exp(\beta U(x))\, dx\right] \\ &= \mathbb{E}\left[\max_x U(x)\right] \end{aligned}}$$

(c) $f(\lambda,h) \equiv \lim_{\beta\to\infty} -\frac{1}{\beta}\varphi(\beta,\lambda,h)$

$$= \operatorname*{ext}_{\substack{q,w,\mu\\ \rho,\nu,\zeta}} \left\{ \lim_{\beta\to\infty}\left[-\frac{1}{\beta}\,\bar{e}\right] + \lim_{\beta\to\infty}\left[-\frac{1}{\beta}\,\overline{ext}\right]\right\}.$$

$$\lim_{\beta\to\infty}\left[-\frac{1}{\beta}\,\bar{e}\right] = \frac{\delta}{2(\delta+w)}\left(p-2\mu+q+\delta\sigma^2\right)$$

$$\boxed{\lim_{\beta\to\infty}\frac{1}{\beta}\log\int\exp\{\beta U(x)\}\,dx = \max_x U(x)}$$

$$\lim_{\beta\to\infty}\left[-\frac{1}{\beta}\,\overline{ext}\right] = (-\nu q+\rho w)/2 + \mu\zeta$$
$$-\mathbb{E}_{G,x^\circ}\left\{\max_x\left[\sqrt{\rho}\cdot x\cdot G - \frac{\nu}{2}\cdot x^2 + \zeta\cdot x\cdot x^\circ - \lambda|x| - h\psi(x,x^\circ)\right]\right\}$$

$$\Rightarrow\ f(\lambda,h) = \operatorname*{ext}_{q w\mu\rho\nu\zeta} f_{mf}(\lambda,h;q,w,\mu,\rho,\nu,\zeta).$$

$$f_{mf}(\lambda,h;q,w,\mu,\rho,\nu,\zeta) = \boxed{\frac{\delta}{2(\delta+w)}}\left(p-2\mu+\boxed{q}+\delta\sigma^2\right) + \boxed{(-\nu q+\rho w)/2} + \mu\zeta$$
$$-\mathbb{E}_{G,x^\circ}\left\{\max_x\left[\sqrt{\rho}\cdot x\cdot G - \frac{\nu}{2}\cdot x^2 + \zeta\cdot x\cdot x^\circ - \lambda|x| - h\psi(x,x^\circ)\right]\right\}$$

$$\frac{\partial f_{mf}}{\partial q} = \frac{\delta}{2(\delta+w)} - \frac{\nu}{2} = 0,\qquad \frac{\partial f_{mf}}{\partial \mu} = -\frac{\delta}{\delta+w} + \zeta = 0$$

$$\Rightarrow\ \nu = \zeta = \frac{\delta}{\delta+w} \qquad w = \left(\frac{1}{\nu}-1\right)\delta.$$

$$\Rightarrow\ f(\lambda,h) = \operatorname*{ext}_{\rho,\nu} f_{mf}(\lambda,h;\rho,\nu)$$

$$f_{mf}(\lambda,h;\rho,\nu) = \frac{\nu}{2}\left(p+\delta\sigma^2\right) + \frac{\rho\delta}{2}\left(\frac{1}{\nu}-1\right)$$
$$-\mathbb{E}_{G,x^\circ}\left\{\max_x\left[\sqrt{\rho}\cdot x\cdot G - \frac{\nu}{2}\cdot x^2 + \nu\cdot x\cdot x^\circ - \lambda|x| - h\psi(x,x^\circ)\right]\right\}$$

$$\partial_h f(\lambda,h)\big|_{h=0} = \partial_h f_{mf}(\lambda,h;\rho,\nu)\big|_{h=0} = \mathbb{E}_{G,x^\circ}\left[\psi(\hat{x},x^\circ)\right]$$

$$\hat{x} = \operatorname*{argmax}_x\left[-\frac{\nu}{2}x^2 + x(\nu x_0 + \sqrt{\rho}\,G) - \lambda|x|\right]$$
$$= \operatorname*{argmin}_x\left[\frac{1}{2}\left(x-\left(x_0+\frac{\sqrt{\rho}}{\nu}G\right)\right)^2 + \frac{\lambda}{\nu}|x|\right]$$
$$= \eta\left(x_0+\frac{\sqrt{\rho}}{\nu}G;\ \frac{\lambda}{\nu}\right).$$

soft-thresholding function c.f. the Theorem.

where ν, ρ is s.t. $f_{mf}(\lambda,0;\rho,\nu)$ is stationery.

$\eta(x_0 + \frac{\sqrt{\rho}}{\nu} G; \frac{\lambda}{\nu})$

$$\partial_\rho f_{mf} = \frac{\delta}{2}\left(\frac{1}{\nu} - 1\right) - \frac{1}{2\sqrt{\rho}} \mathbb{E}[\hat{x} G]$$

integration by part.

$$= \frac{\delta}{2}\left(\frac{1}{\nu} - 1\right) - \frac{1}{2\nu} \mathbb{E}\left[\partial \eta\left(x^0 + \frac{\sqrt{\rho}}{\nu} G; \frac{\lambda}{\nu}\right)\right] = 0$$

$$\partial_\nu f_{mf} = \frac{1}{2}(\rho + \delta \sigma^2) - \frac{\rho\delta}{2\nu^2} + \mathbb{E}\left[\frac{\hat{x}^2}{2} - x_0 \hat{x}\right]$$

$$= \frac{1}{2}\mathbb{E}[(\hat{x} - x_0)^2] + \frac{1}{2}\sigma^2 - \frac{\rho}{2\nu^2} = 0.$$

$$\Rightarrow \begin{cases} \frac{\rho}{\nu^2} = \sigma^2 + \delta^{-1} \mathbb{E}\left[\left(\eta\left(x_0 + \frac{\sqrt{\rho}}{\nu} G; \frac{\lambda}{\nu}\right) - x_0\right)^2\right] \\ \nu = 1 - \delta^{-1} \mathbb{E}\left[\partial \eta\left(x_0 + \frac{\sqrt{\rho}}{\nu} G; \frac{\lambda}{\nu}\right) - x_0\right]. \end{cases}$$

$$\begin{cases} \frac{\rho}{\nu^2} = \tau^2 \\ \nu = \frac{\lambda}{\alpha\tau}. \end{cases}$$

This recovers the self consistent equation of the theorem.