

A perceptron version of the GREM, and some applications

Nicola Kistler, University of Bonn

joint with **Erwin Bolthausen**, University of Zurich

Outline

1. aDR: Additive Derrida-Ruelle cascades
2. Derrida's GREM
3. A perceptron version of the GREM
4. An application: GREM+Cavity

1. Additive Derrida-Ruelle cascades (two levels)

- Parameters $0 < m_1 < m_2 < \infty$
- $(x_{j_1}^1, j_1 \in \mathbb{N})$ Poisson Point Process $\text{PPP}(e^{-m_1 t})$
- Given j_1 , $(x_{j_1, l}^2, l \in \mathbb{N})$ $\text{PPP}(e^{-m_2 t} dt)$
- x^1, x^2 independent
- $(x_{j_1, l}^2, l \in \mathbb{N})$ independent for different j_1

$$\text{aDR}(m_1, m_2) \quad \text{PP} (x_j; j \in \mathbb{N}^2) \quad x_j = x_{j_1}^1 + x_{j_1, j_2}^2$$

A badly formulated conjecture the geometry of extremes

- Given a configuration space Σ_N , $\#\Sigma_N = 2^N$
- Random field $\{X_\alpha\}_{\alpha \in \Sigma_N}$, say gaussian
- $\text{cov}(X_\alpha, X_{\alpha'}) = Nf(\alpha, \alpha')$, $f : \Sigma_N \times \Sigma_N \rightarrow [0, 1]$
- $a_N \stackrel{\text{def}}{=} \mathbb{E} \max_{\sigma} X_\sigma + o(N)$

$$\lim_{N \rightarrow \infty} \sum_{\sigma} \delta_{X_\sigma - a_N} = \mathbf{aDR}$$

2. Derrida's GREM (two levels)

Hamiltonian $X_\alpha = X_{\alpha_1} + X_{\alpha_1, \alpha_2}$

- Configurations $\alpha = (\alpha_1, \alpha_2)$ $\#\alpha_i = 2^{N/2}$
- $X_{\alpha_1} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, Na_1)$ $X_{\alpha_1, \alpha_2} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, Na_2)$
- $a_1 + a_2 = 1$, (w.l.o.g. $0 < a_1 < a_2 < 1$)

Thm. [Bovier-Kurkova '04]

$$\lim_{N \rightarrow \infty} (X_\alpha - a_N)_\alpha = \text{aDR}(\beta_1, \beta_2)$$

Idea behind proof.

- Extreme value Theory: $\left(X_{\alpha_1} - a_N^{(1)} \right)_{\alpha_1} \rightarrow PPP(e^{-\beta_1 t} dt)$
- Also, for fixed α_1 : $\left(X_{\alpha_1, \alpha_2} - a_N^{(2)} \right)_{\alpha_2} \rightarrow PPP(e^{-\beta_2 t} dt)$

Thus, very plausible that

$$X_\alpha - a_N = \left(X_{\alpha_1} - a_N^{(1)} \right) + \left(X_{\alpha_1, \alpha_2} - a_N^{(2)} \right) \rightarrow x_{j_1}^1 + x_{j_1, j_2}^2 = \text{aDR}(\beta_1, \beta_2)$$

Crucial property: linearity

3. A perceptron version of the GREM

- Configurations $\alpha = (\alpha_1, \alpha_2)$, $\alpha_1, \alpha_2 = 1, \dots, 2^{N/2}$
- $(X_{\alpha_1, i}, X_{\alpha_1, \alpha_2, i}) \stackrel{\text{iid}}{\sim} \mu = \mu_1 \otimes \mu_2$ “well-behaved”
- $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ “well-behaved”

$$Y_\alpha = \sum_{i=1}^N \phi(X_{\alpha_1, i}; X_{\alpha_1, \alpha_2, i})$$

Rem. Linear $\phi(x_1; x_2) = x_1 + x_2$, μ gaussian \Leftrightarrow GREM

- Empirical measure $L_{N,\alpha} = \frac{1}{N} \sum_{i=1}^N \delta_{X_{\alpha_1,i}; X_{\alpha_1,\alpha_2,i}}$
- Free Energy $f_N(\phi) = \frac{1}{N} \log \sum_{\alpha} \left[\exp \int \phi(x) L_{N,\alpha}(dx) \right]$
- $\nu \in \mathcal{M}(\mathbb{R}^2)$ distribution on \mathbb{R}^2 , $\nu^{(1)}$ projection on 1st-coord
- $H(\cdot | \mu)$ relative entropy

Thm. Gibbs Variational Principle - GVP $f(\phi) = \lim_{N \rightarrow \infty} f_N(\phi)$

$$= \sup_{\nu \in \mathcal{M}(\mathbb{R}^2)} \left\{ \int \phi d\nu - H(\nu | \mu) : H(\nu^{(1)} | \mu^{(1)}) \leq \frac{\log 2}{2}, H(\nu | \mu) \leq \log 2 \right\}$$

For $m_1, m_2 \in \mathbb{R}_+$, construct

$$\phi_1(x_1) = \frac{1}{m_2} \log \int e^{m_2 \phi(x_1, z)} \mu_2(dz)$$

$$\phi_0 = \frac{1}{m_1} \log \int e^{m_1 \phi_1(z)} \mu_1(dz)$$

$$\mathcal{P}(m_1, m_2) = \phi_0 + \frac{\log 2}{2m_1} + \frac{\log 2}{2m_2}$$

Thm. Parisi Variational Principle - PVP

$$f(\phi) = \inf_{0 < m_1 \leq m_2 \leq 1} \mathcal{P}(m_1, m_2)$$

Def. Low-temp if inf achieved in $0 < m_1 < m_2 < 1$ (might not exist, depends on the choice of ϕ, μ)

Recall

$$Y_\alpha = \sum_{i=1}^N \phi \left(X_{\alpha_1, i}; X_{\alpha_1, \alpha_2, i} \right)$$

Thm. **Emergence of aDR** *IF in low temp*

$$(Y_\alpha - a_N)_\alpha \rightarrow \text{aDR}(m_1, m_2)$$

- $a_N = Nf(\phi) + O(\log N) = \mathbb{E} \max_\alpha Y_\alpha + O(\log N)$
- m_1, m_2 *extremals in Parisi Variational Principle*

$$(m_1, m_2) = \arg \inf_{0 < \tilde{m}_1 < \tilde{m}_2 < 1} \mathcal{P}(\tilde{m}_1, \tilde{m}_2)$$

Sketch of the proofs. Free Energy. By Second Moment on empirical measures

$$f(\phi) = \sup_{\nu} \left\{ \int \phi d\nu - H(\nu | \mu) : \right. \\ \left. H(\nu^{(1)} | \mu^{(1)}) \leq \frac{\log 2}{2}, \quad H(\nu | \mu) \leq \log 2 \right\}$$

Candidate extremal measure: **Gibbs-like**

$$dG_{m_1, m_2}(x_1, x_2) \sim e^{m_1 \phi_1(x_1)} \times e^{m_2 \phi(x_1, x_2)} \mu(dx_1, dx_2)$$

Link GIBBS \leftrightarrow PARISI: with these measures

$$H \left(G_{m_1, m_2}^{(1)} \mid \mu^{(1)} \right) \leq \frac{\log 2}{2} \quad \partial_{m_1} \left\{ \phi_o + \frac{\log 2}{2m_1} + \frac{\log 2}{2m_2} \right\} \leq 0$$

\leftrightarrow

$$H \left(G_{m_1, m_2} \mid \mu \right) \leq \log 2 \quad \partial_{m_2} \left\{ \phi_o + \frac{\log 2}{2m_1} + \frac{\log 2}{2m_2} \right\} \leq 0$$

(In low temp, equality on both sides)

Corollary.

- (m_1, m_2) PVP-extremal $\Rightarrow G_{m_1, m_2}$ GVP-extremal
- Parisi order parameter $(m_1, m_2) \equiv$ inverse of temperatures for extremal measures in GVP

Energy levels. Under aDR-conjecture one expects

$$Y_\alpha = \sum_{i=1}^N \phi(X_{\alpha_1, i}; X_{\alpha_1, \alpha_2, i}) \approx Y_{\alpha_1}^{(1)} + Y_{\alpha_1, \alpha_2}^{(2)}$$

Natural guess: recover linearity through **telescopic sum**

$$Y_\alpha = \sum_{i=1}^N \int \phi(X_{\alpha_1, i}, y) G_{m_2}(dy | X_{\alpha_1, i}) + \left\{ \sum_{i=1}^N \phi(X_{\alpha_1, i}, X_{\alpha_1, \alpha_2, i}) - \sum_{i=1}^N \int \phi(X_{\alpha_1, i}, y) G_{m_2}(dy | X_{\alpha_1, i}) \right\}$$

where $G_{m_2}(dx_2 | X_{\alpha_1, i}) \sim e^{m_2 \phi(X_{\alpha_1, i}, x_2)} \mu_2(dx_2)$

...is **WRONG**: need to adjust through
random temperatures

$$Y_\alpha = \sum_{i=1}^N \int \phi(X_{\alpha_1, i}, y) G_{\tilde{m}_2}(dy | X_{\alpha_1, i}) +$$
$$+ \left\{ \sum_{i=1}^N \phi(X_{\alpha_1, i}, X_{\alpha_1, \alpha_2, i}) - \sum_{i=1}^N \int \phi(X_{\alpha_1, i}, y) G_{\tilde{m}_2}(dy | X_{\alpha_1, i}) \right\}$$

How to construct $\tilde{m}_2 = \tilde{m}_2(\alpha_1) = \tilde{m}_2(X_{\alpha_1, 1}, \dots, X_{\alpha_1, N})$?

- Solution to *quenched* entropy condition

$$\frac{1}{N} \sum_{i=1}^N H \left(G_{\tilde{m}_2} (\cdot | X_{\alpha_1, i}) \middle| \mu_2 \right) = \frac{\log 2}{2}$$

- Equivalent: minimizer of *quenched* Parisi Free Energy

$$\tilde{m}_2 = \arg \inf_{0 \leq \zeta \leq 1} \left\{ \frac{\log 2}{2\zeta} + \frac{1}{N} \sum_{i=1}^N \frac{1}{\zeta} \log E_{\mu_2} \exp \zeta \phi (X_{\alpha_1, i}, \cdot) \right\}$$

⇒ state α_1 chooses random temperature for 2nd-level through **MINIMUM Likelihood Estimation**

Thm. Self-averaging of random temperatures

$$\mathbb{P} \left[\exists \alpha_1 : \frac{1}{N} \sum_{i=1}^N \delta_{X_{\alpha_1, i}} \approx G_{m_1, m_2}^{(1)}, |\tilde{m}_2(\alpha_1) - m_2| \geq \sqrt{\frac{\log N}{N}} \right] = O\left(\frac{1}{N}\right)$$

Proof. CLT for MLE [Bhattacharya-Ghosh '75]

□

$$\begin{aligned} Y_\alpha - a_N &= \left\{ \sum_{i=1}^N \int \phi(X_{\alpha_1, i}, y) G_{\tilde{m}_2}(dy | X_{\alpha_1, i}) - a_N^{(1)} \right\} + \\ &+ \left\{ \sum_{i=1}^N \left(\phi(X_{\alpha_1, i}, X_{\alpha_1, \alpha_2, i}) - \int \phi(X_{\alpha_1, i}, y) G_{\tilde{m}_2}(dy | X_{\alpha_1, i}) \right) - a_N^{(2)} \right\} \end{aligned}$$

Taylor+Edgeworth expansions \Rightarrow convergence to aDR

aDR as universal objects: convergence holds for

- any *well behaved* function ϕ (say C_∞ , decaying sufficiently fast: $\int e^{\lambda\phi} d\mu < \infty$ for all λ)
- any *well behaved* underlying distribution μ (say non-lattice, C_∞ -density)
- any finite number of levels

Rem. *In fact, likely to hold also for hamiltonians $Y_\alpha = \Phi(L_{N,\alpha})$ with $L_{N,\alpha}$ empirical measures, under some smoothness condition on Φ , say Fréchet diff'ty. (Analysis very cumbersome - done only for free energy.)*

4. An application: GREM + Cavity

Thm. [Aizenman-Sims-Starr '03]

“aDR + cavity field \Rightarrow Parisi Solution for SK”

Natural (?) $\lim_{N \rightarrow \infty} [\text{GREM}(N) + \text{Cavity}(N)] = (?)$ i.e.

Hamiltonian
$$H_N(\alpha, \sigma) = X_\alpha + \sum_{i=1}^N g_{\alpha,i} \sigma_i \quad \sigma_i = \pm 1$$

- $X_\alpha = X_{\alpha_1} + \cdots + X_{\alpha_1, \dots, \alpha_K} \equiv K\text{-levels GREM}$

- $g_{\alpha,i} = g_{\alpha_1,i} + \cdots + g_{\alpha_1, \dots, \alpha_K,i} \quad g_{\alpha_1, \dots, \alpha_j,i} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, a_j)$

$$f_N(\beta) = \frac{1}{N} \log \sum_{\alpha, \sigma} e^{\beta H_N(\alpha, \sigma)} \sim \frac{1}{N} \log \sum_{\alpha} e^{\beta X_{\alpha} + \sum \log \cosh(\beta g_{\alpha, i})}$$

Perceptron GREM $\phi(\underline{x}, \underline{y}) = \sum_{i=1}^K \beta x_i + \log \cosh \left(\sum_{i=1}^K \beta y_i \right)$

Thm. Parisi Formula for free energy

$$f(\beta) = \lim f_N(\beta) = \min_{x \in \mathcal{X}} \left\{ GREM(\beta, x) + P(0, 0; \beta, x) \right\}$$

with $P = P(q, y; \beta, x)$ solution of

$$\partial_q P + \frac{1}{2} \partial_y^2 P + \frac{1}{2} x(q) (\partial_y P)^2 = 0, \quad P(1, y) = \log \cosh(\beta y)$$

Gibbs measure.

- $\mathcal{G}_{\beta,N}(\alpha, \sigma) = \frac{\exp \beta H_N(\alpha, \sigma)}{Z_N(\beta)}, \quad \langle \cdot \rangle_{\beta,N} = \sum_{\alpha, \sigma} (\cdot) \mathcal{G}_{\beta,N}(\alpha, \sigma)$
- $A_\alpha \stackrel{\text{def}}{=} \{(\alpha, \sigma) : \sigma \in \{\pm 1\}^N\}$

Thm. Pure States. $\exists \beta_{crit}$ s.t. for $\beta > \beta_{crit}$

$$\left\{ \mathcal{G}_{\beta,N}(A_\alpha) \right\}_\alpha \rightarrow \left(\frac{\exp[x_j]}{\sum_{j'} \exp[x_{j'}]} \right) \stackrel{(d)}{=} \text{Poisson-Dirichlet}$$

Ising spin overlap $q_N(\sigma, \sigma') = \frac{1}{N} \sum_i \sigma_i \sigma'_i$

GREM-overlap $d(\alpha, \alpha') = k$ iff $\alpha_i = \alpha'_i, i \leq k$

$\langle \cdot \rangle_{\beta, \beta', N} = \sum_{(\alpha, \sigma), (\alpha', \sigma')} (\cdot) \mathcal{G}_{\beta, N}(\alpha, \sigma) \mathcal{G}_{\beta', N}(\alpha', \sigma')$

Thm. Ultrametricity and Chaos in temperature.

- $\exists \{q_j, j = 1 \dots K\}$ s.t. for $\beta > \beta_{crit}$

$$\mathbb{E} \left\langle \mathbf{1}_{d(\alpha, \alpha')=j} \left(q_N(\sigma, \sigma') - q_j \right)^2 \right\rangle_{\beta, \beta, N} \rightarrow 0$$

- For $\beta \neq \beta', \beta, \beta' > \beta_{crit}$: $\mathbb{E} \langle \delta_{\alpha=\alpha'} \rangle_{\beta, \beta', N} \rightarrow 0$

Thus $\alpha = \alpha(\beta)$!