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

aDR(m_1, m_2) PP $\left(x_{\boldsymbol{j}}; \boldsymbol{j} \in \mathbb{N}^2\right)$ $x_{\boldsymbol{j}} = x^1_{j_1} + x^2_{j_1,j_2}$

A badly formulated conjecture
the geometry of extremes

- Given a configuration space Σ_N, $\quad \sharp\Sigma_N = 2^N$

- Random field $\{X_\alpha\}_{\alpha \in \Sigma_N}$, say gaussian

- $\mathrm{cov}\,(X_\alpha, X_{\alpha'}) = N f(\alpha, \alpha')$, $\quad f : \Sigma_N \times \Sigma_N \to [0, 1]$

- $a_N \stackrel{\text{def}}{=} \mathbb{E}\,\mathrm{max}_\sigma\, X_\sigma + o(N)$

$$\lim_{N\to\infty} \sum_{\sigma} \delta_{X_\alpha - a_N} = \textbf{aDR}$$

2. Derrida's GREM (two levels)

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

- Configurations $\alpha = (\alpha_1, \alpha_2) \qquad \sharp\alpha_i = 2^{N/2}$

- $X_{\alpha_1} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, Na_1) \qquad X_{\alpha_1,\alpha_2} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, Na_2)$

- $a_1 + a_2 = 1$, (w.l.o.g. $0 < a_1 < a_2 < 1$)

Thm. **[Bovier-Kurkova '04]**

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

Idea behind proof.

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

Thus, very plausible that

$$X_\alpha - a_N = \left(X_{\alpha_1} - a_N^{(1)}\right) + \left(X_{\alpha_1,\alpha_2} - a_N^{(2)}\right) \to x^1_{j_1} + x^2_{j_1,j_2} = \mathrm{aDR}(\beta_1, \beta_2)$$

Crucial property: linearity

3. A perceptron version of the GREM

- Configurations $\alpha = (\alpha_1, \alpha_2),\ \alpha_1, \alpha_2 = 1, \ldots, 2^{N/2}$

- $\left(X_{\alpha_1,i}, X_{\alpha_1,\alpha_2,i}\right) \overset{\text{iid}}{\sim} \mu = \mu_1 \otimes \mu_2$ "well-behaved"

- $\phi : \mathbb{R}^2 \to \mathbb{R}$ "well-behaved"

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

Rem. Linear $\phi(x_1; x_2) = x_1 + x_2,\ \mu$ 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 $f_N(\phi) = \frac{1}{N}\log\sum_\alpha \left[\exp\int \phi(x)\boldsymbol{L_{N,\alpha}}(dx)\right]$

- $\nu \in \mathcal{M}(\mathbb{R}^2)$ distribution on $\mathbb{R}^2$, $\nu^{(1)}$ projection on 1st-coord

- $H(\cdot \mid \mu)$ relative entropy

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

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

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

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

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

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

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*

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

- $a_N = Nf(\phi) + O(\log N) = \mathbb{E}\max_\alpha Y_\alpha + O(\log N)$
- 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

$$f(\phi) = \sup_{\nu} \left\{ \int \phi d\nu - H(\nu \mid \mu) : \right.$$

$$\left. H\left(\nu^{(1)} \mid \mu^{(1)}\right) \leq \frac{\log 2}{2}, \quad H(\nu \mid \mu) \leq \log 2 \right\}$$

Candidate extremal measure: **Gibbs-like**

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

Link GIBBS $\leftrightarrow$ PARISI: with these measures

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

$$\leftrightarrow$$

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

(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

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

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

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

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

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

random temperatures

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

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

- Solution to *quenched* entropy condition

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

- Equivalent: minimizer of *quenched* Parisi Free Energy

$$\tilde{m}_2 = \operatorname{arg\,inf}_{0\le\varsigma\le 1}\left\{\frac{\log 2}{2\varsigma} + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\varsigma}\log \mathsf{E}_{\mu_2}\exp\varsigma\phi\left(X_{\alpha_1,i},\cdot\right)\right\}$$

$\Rightarrow$ state α_1 chooses random temperature for 2$^{\text{nd}}$-level through **MINIMUM Likelihood Estimation**

Thm. Self-averaging of random temperatures

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

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

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

Taylor+Edgeworth expansions ⇒ 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* $Y_\alpha = \Phi\left(L_{N,\alpha}\right)$ *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 ⇒ Parisi Solution for SK"

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

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

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

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

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

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*

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

Gibbs measure.

- $\mathcal{G}_{\beta,N}(\alpha,\sigma) = \frac{\exp \beta H_N(\alpha,\sigma)}{Z_N(\beta)}, \quad \langle\cdot\rangle_{\beta,N} = \sum_{\alpha,\sigma}(\cdot)\mathcal{G}_{\beta,N}(\alpha,\sigma)$

- $A_\alpha \stackrel{\text{def}}{=} \left\{(\alpha,\sigma) : \sigma \in \{\pm 1\}^N\right\}$

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

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

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

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

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

Thm. Ultrametricity and Chaos in temperature.

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

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

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

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