Unsupervised learning of symmetry-breaking direction from examples

[Biehl 1993
Watkin, Nadal 1994
Reimann, Van den Broeck 1996

- direction $\vec{B}$, $\vec{B}^2 = N$, i.e. all components $B_i \sim O(1)$
- independent examples $\vec{\xi}^\mu$, $\mu = 1 \ldots P$

$$P_0(\vec{\xi}^\mu) \propto e^{-\frac{1}{2}(\vec{\xi}^\mu)^2}, \quad (\vec{\xi}^\mu)^2 \simeq N \text{ for large } N$$

bias: $$P(\vec{\xi}^\mu) \propto P_0(\vec{\xi}^\mu)\, e^{-V\left(\frac{\vec{\xi}^\mu \cdot \vec{B}}{\sqrt{N}}\right)}$$

weak bias, i.e. $|\vec{\xi}^\mu \cdot \vec{B}| \ll N$

example: $$V(\lambda) = -\frac{s}{2(1+s)}\lambda^2$$

more general V ~~can be~~ were considered

$$V(\lambda) = a\lambda^2 + b\lambda + c|\lambda| + \ldots$$

- inference of $\vec{J}$ (proxy for $\vec{B}$):

Bayes: $$P(\vec{J} \mid \{\vec{\xi}^\mu\}) \propto e^{-\sum_{\mu=1}^{P} V\left(\frac{\vec{\xi}^\mu \cdot \vec{J}}{\sqrt{N}}\right)}\, \delta(\vec{J}^2 - N)$$

Actually we can choose a different V for the inference!
Often we use $\beta V(\lambda)$ instead of $V(\lambda)$

$\beta \to \infty$: MAP decoding (gives PCA for $V(\lambda) = -\lambda^2$)
$\beta = 1$: Bayesian decoding

Normalization of $P(\vec{J} \mid \{\vec{\xi}^\mu\})$:

$$Z = \int d\vec{J}\, \delta(\vec{J}^2 - N)\, e^{-\beta \sum_\mu V\left(\frac{\vec{J} \cdot \vec{\xi}^\mu}{\sqrt{N}}\right)}$$

We want to compute the average of log Z over the examples. We use the replica approach

$$\left\langle Z^n \right\rangle_{\{\xi^\mu\}} = \frac{\mathcal{Z}(n)}{\mathcal{Z}(0)} \quad \text{with}$$

$$\mathcal{Z}(n) = \int \prod_{\mu i} d\xi_{i\mu}\, e^{-\frac{1}{2}\sum_{i\mu}\xi_{i\mu}^2 - \sum_\mu V\left(\frac{\sum_i \xi_{i\mu} B_i}{\sqrt{N}}\right)} \cdot \int \prod_{ai} dJ_{ia}\, \delta\left(\sum_i J_{ia}^2 - N\right) e^{-\beta\sum_{\mu a} V\left(\frac{\sum_i \xi_{i\mu} J_{ia}}{\sqrt{N}}\right)}$$

(1) introduction of stabilities:

$$1 = \int \prod_\mu \frac{d\lambda_\mu\, d\hat{\lambda}_\mu}{2\pi}\, e^{i\sum_\mu \hat{\lambda}_\mu\left(\lambda_\mu - \frac{1}{\sqrt{N}}\sum_i \xi_{i\mu} B_i\right)}$$

$$1 = \int \prod_{\mu a} \frac{dt_{\mu a}\, d\hat{t}_{\mu a}}{2\pi}\, e^{i\sum_{\mu a}\hat{t}_{\mu a}\left(t_{\mu a} - \frac{1}{\sqrt{N}}\sum_i \xi_{i\mu} J_{ia}\right)}$$

(2) Gaussian integration over examples:

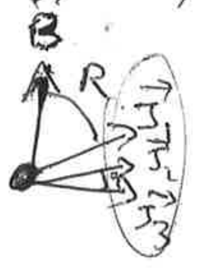
$$\int \prod_{i\mu} d\xi_{i\mu}\, e^{-\frac{1}{2}\sum_{i\mu}\xi_{i\mu}^2 - \frac{i}{\sqrt{N}}\sum_{i\mu}\xi_{i\mu}\left(B_i\hat{\lambda}_\mu + \sum_a J_{ia}\hat{t}_{\mu a}\right)}$$

$$= e^{-\frac{1}{2N}\sum_{i\mu}\left(B_i\hat{\lambda}_\mu + \sum_a J_{ia}\hat{t}_{\mu a}\right)^2} = e^{-\frac{1}{2}\sum_\mu\left(\hat{\lambda}_\mu^2 + \sum_a R_a\hat{\lambda}_\mu\hat{t}_{\mu a} + \sum_{ab} q_{ab}\hat{t}_{\mu a}\hat{t}_{\mu b}\right)}$$

upto $(2\pi)^{\cdots}$ factor

with $R_a \equiv \frac{1}{N}\sum_i J_{ia} B_i$

$q_{ab} \equiv \frac{1}{N}\sum_i J_{ia} J_{ib}$



③ Introduction of order parameters

$$1 = \int \prod_{a<b} \frac{\underset{\mathbb{R}}{dq_{ab}}\, \underset{i\mathbb{R}}{d\hat{q}_{ab}}}{2\pi/N}\; e^{N \sum_{a<b} \hat{q}_{ab}\left(q_{ab} - \frac{1}{N}\sum_i J_{ia} J_{ib}\right)}$$

$$1 = \int \prod_a \frac{dR_a\, d\hat{R}_a}{2\pi/N}\; e^{N \sum_a \hat{R}_a \left(R_a - \frac{1}{N}\sum_i J_{ia} B_{ib}\right)}$$

$$1 = \int \prod_a \frac{d\hat{u}_a}{2\pi/N}\; e^{\frac{1}{2}\sum_a \hat{u}_a \left(N - \sum_i J_{ia}^2\right)} \Leftarrow \delta(\vec{J}_a^{\,2} - N)$$

Rewriting of $Z(n) =$

$$\int \prod_a d\hat{u}_a \int \prod_{a<b} d\hat{q}_{ab}\, dq_{ab} \int \prod_a d\hat{R}_a\, dR_a\; e^{N\left(\frac{1}{2}\sum_a \hat{u}_a + \sum_{a<b} \hat{q}_{ab} q_{ab} + \sum_a \hat{R}_a R_a\right)} \times$$

$$\left\{ \int \frac{d\lambda\, d\hat{\lambda}}{2\pi} \int \prod_a \frac{dt_a\, d\hat{t}_a}{2\pi}\; e^{\lambda\hat{\lambda} - V(\lambda) + \sum_a \left(\hat{t}_a t_a - \beta V(t_a)\right)}\; e^{-\frac{1}{2}\left(\hat{\lambda}^2 + \hat{\lambda}\sum_a R_a \hat{t}_a + \sum_{ab} q_{ab} \hat{t}_a \hat{t}_b\right)} \right\}^{P = \alpha N}$$

$$\left\{ \int \prod_a dJ_a\; e^{-\frac{1}{2}\sum_a \hat{u}_a J_a^2 - \sum_a \hat{R}_a \sum_a J_a - \sum_{a<b} \hat{q}_{ab} J_a J_b} \right\}^N$$

I chose $B_1 = B_2 = \cdots = B_N = 1$

④ Saddle point calculation:

as $N \to \infty$, integrals over $\hat{u}_a, \hat{q}_{ab}, q_{ab}, \hat{R}_a, R_a$ can be done by saddle-point method;

there exists the Replica Symmetric solution:

$$\boxed{\hat{u}_a = u,\quad \hat{q}_{ab} = \hat{q},\quad q_{ab} = q,\quad \hat{R}_a = R,\quad R_a = R}$$

(Stability can be checked a posteriori)

RS hypothesis allows us to have $\mathcal{N}(n)$ as an analytical function of n, so we can send $n \to 0$ and compute

$$\langle \log Z \rangle = \left(\frac{\mathcal{N}'(0)}{\mathcal{N}(0)} \right) = N \operatorname*{ext}_{q,R} G(q,R)$$

$$G(q,R) = \frac{1}{2} \log(1-q) + \frac{1}{2} \frac{1-R^2}{1-q} +$$

$$\alpha \int \underbrace{dt\, e^{-t^2/2 - V(t)}} \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \log\left(\int \frac{d\tau}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} e^{-\beta V(\sigma)} \right)$$

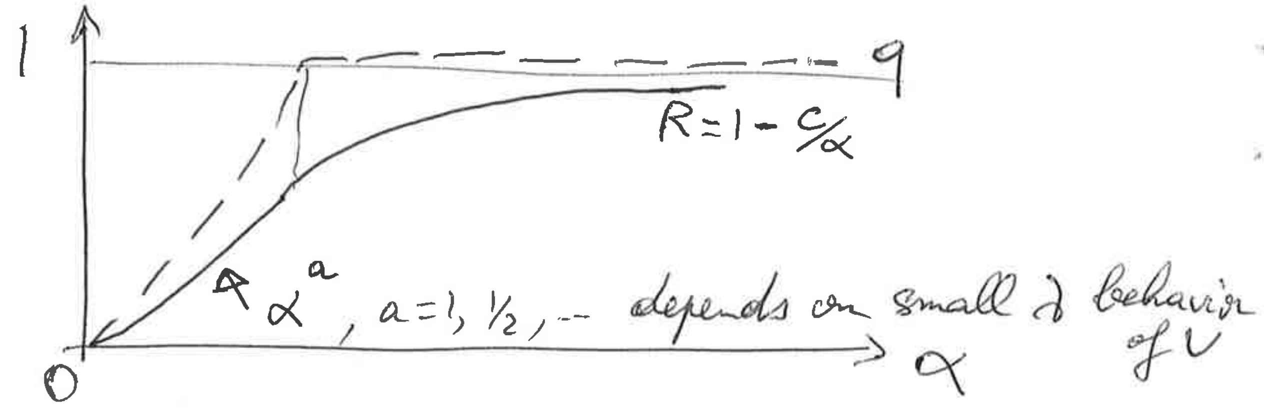
normalized i.e. divided by $\int dx\, e^{-x^2/2 - V(x)}$

with $\sigma = R t + \sqrt{q - R^2}\, z + \sqrt{1-q}\, \tau$

<u>NB:</u> $\hat{u}$, $\hat{R}$, $\hat{q}$ have been eliminated, i.e. expressed as functions of q and R.

5 <u>Results</u>:

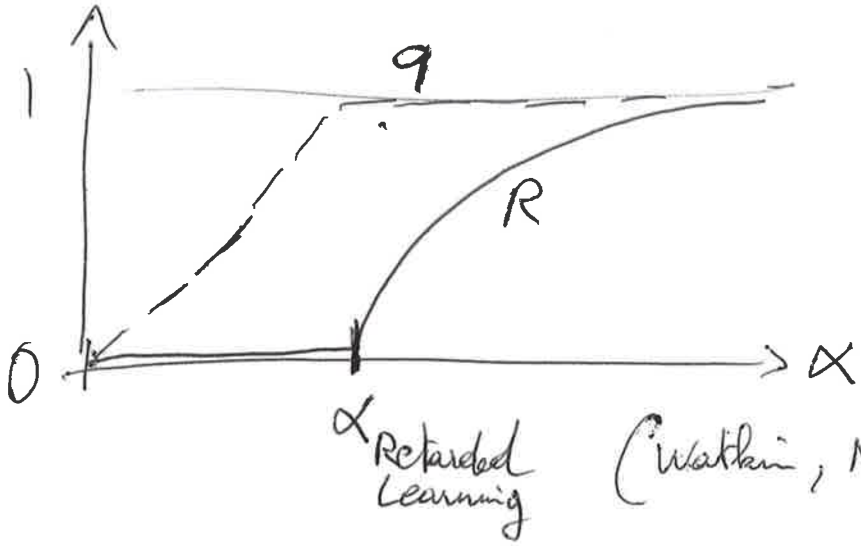
Two main scenarios:



This is the standard scenario, obtained if

$$\bar{\lambda} \equiv \int d\lambda \; e^{-\frac{\lambda^2}{2} - V(\lambda)} \lambda \neq 0$$

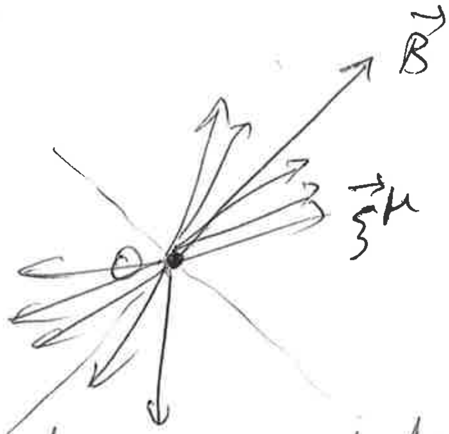
If the integral vanishes, we obtain:



(Watkin, Nadal 1994)

Intuition:

- if $\langle \vec{B} \cdot \vec{\xi}^{\mu} \rangle \neq 0$, then $\sum_{\mu} \vec{\xi}^{\mu}$ gives an estimate for $\vec{B}$,
  hence $\vec{B}$ can be inferred from <u>few</u> examples
- if $\langle \vec{B} \cdot \vec{\xi}^{\mu} \rangle = 0$, this is much harder



entropy of $\vec{J}$ decays quadratically with R

energy of $\vec{J}$ decreases ———— R, proportionally to α

NB: there may also be 1st order PT depending on the non-linearities of V.