The replica method and its applications in biomedical modelling and data analysis

Statistical Physics Approaches to Systems Biology, Havana, Feb 2019

ACC Coolen

King's College London and Saddle Point Science

replica method

A clever trick that enables the analytical calculation of averages
that are normally impossible to do, except numerically.

is particularly useful for

Complex heterogeneous systems composed of *many* interacting variables,
and with *many* parameters on which we have only statistical information.
(too large for numerical averages to be computationally feasible)

gives us

Analytical predictions for the behaviour of *macroscopic* quantities
in *typical* realisations of the systems under study.

note on biomedical applications

The 'large systems' could describe actual *biochemical processes*
(folding proteins, proteome, transcriptome, immune or neural networks, etc),
or *analysis algorithms* running on large biomedical data sets

1. The replica method
   - Exponential families and generating functions
   - The replica trick
   - The replica trick and algorithms
   - Alternative forms of the replica identity

2. Application: information storage in neural networks
   - Attractor neural networks
   - The replica calculation
   - Replica symmetry
   - Replica symmetric solution

3. Application: overfitting transition in linear separators
   - Linear separability of data – version space
   - The replica calculation
   - Gardner's replica symmetric theory

Exponential distributions

Often we study stochastic processes for $\mathbf{x} \in X \subseteq \mathbb{R}^N$,
that evolve to a stationary state, with prob distribution $p(\mathbf{x})$
many are of the following form:

- stationary state is *minimally informative*,
  subject to a number of constraints

$$\sum_{\mathbf{x}\in X} p(\mathbf{x})\omega_1(\mathbf{x}) = \Omega_1 \quad \ldots\ldots \quad \sum_{\mathbf{x}\in X} p(\mathbf{x})\omega_L(\mathbf{x}) = \Omega_L$$

This is enough to calculate $p(\mathbf{x})$:

- information content of $\mathbf{x}$: Shannon entropy
  hence

$$\textit{maximize} \quad S = -\sum_{\mathbf{x}\in X} p(\mathbf{x}) \log p(\mathbf{x})$$

$$\textit{subject to}: \begin{cases} p(\mathbf{x}) \geq 0 \ \ \forall \mathbf{x}, \ \ \sum_{\mathbf{x}\in X} p(\mathbf{x}) = 1 \\ \sum_{\mathbf{x}\in X} p(\mathbf{x})\omega_\ell(\mathbf{x}) = \Omega_\ell \ \ \textit{for all } \ell = 1 \ldots L \end{cases}$$

- solution using Lagrange's method:

$$\frac{\partial}{\partial p(\mathbf{x})}\left\{\lambda_0 \sum_{\mathbf{x}'\in X} p(\mathbf{x}') + \sum_{\ell=1}^{L} \lambda_\ell \sum_{\mathbf{x}'\in X} p(\mathbf{x}')\omega_\ell(\mathbf{x}') - \sum_{\mathbf{x}'\in X} p(\mathbf{x}') \log p(\mathbf{x}')\right\} = 0$$

$$\lambda_0 + \sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x}) - 1 - \log p(\mathbf{x}) = 0 \quad \Rightarrow \quad p(\mathbf{x}) = \mathrm{e}^{\lambda_0 - 1 + \sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}$$

$(p(\mathbf{x}) \geq 0$ *automatically satisfied*$)$

- 'exponential distribution':

$$p(\mathbf{x}) = \frac{\mathrm{e}^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}}{Z(\boldsymbol{\lambda})}, \qquad Z(\boldsymbol{\lambda}) = \sum_{\mathbf{x}\in X} \mathrm{e}^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}$$

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L): \quad \textit{solved from} \quad \sum_{\mathbf{x}\in X} p(\mathbf{x})\omega_\ell(\mathbf{x}) = \Omega_\ell \quad (\ell = 1 \ldots L)$$

example:
physical systems in thermal equilibrium
$L = 1$, $\omega(\mathbf{x}) = E(\mathbf{x})$ (energy), $\lambda = -1/k_B T$

$$p(\mathbf{x}) = \frac{\mathrm{e}^{-E(\mathbf{x})/k_B T}}{Z(T)}, \qquad Z(T) = \sum_{\mathbf{x}\in X} \mathrm{e}^{-E(\mathbf{x})/k_B T}$$

Generating functions

$$p(\mathbf{x}) = \frac{e^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}}{Z(\boldsymbol{\lambda})}, \qquad Z(\boldsymbol{\lambda}) = \sum_{\mathbf{x}\in X} e^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}, \qquad \langle f \rangle = \sum_{\mathbf{x}\in X} p(\mathbf{x}) f(\mathbf{x})$$

Idea behind generating functions:
reduce nr of state averages to be calculated ...

- define

  $F(\boldsymbol{\lambda}) = \log Z(\boldsymbol{\lambda})$

  $$\frac{\partial F(\boldsymbol{\lambda})}{\partial \lambda_k} = \frac{\sum_{\mathbf{x}\in X} \omega_k(\mathbf{x}) e^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}}{\sum_{\mathbf{x}\in X} e^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x})}} = \langle \omega_k(\mathbf{x}) \rangle$$

- how to calculate
  arbitrary state average $\langle \psi \rangle$?

  $$F(\boldsymbol{\lambda}, \mu) = \log \left[\sum_{\mathbf{x}\in X} e^{\mu\psi(\mathbf{x}) + \sum_\ell \lambda_\ell \omega_\ell(\mathbf{x})} \right]$$

  $$\langle \psi \rangle = \lim_{\mu \to 0} \frac{\partial F(\boldsymbol{\lambda}, \mu)}{\partial \mu}, \qquad \langle \omega_\ell \rangle = \lim_{\mu \to 0} \frac{\partial F(\boldsymbol{\lambda}, \mu)}{\partial \lambda_\ell}$$

The replica trick

first appearance: Marc Kac 1968
first application in physics: Sherrington & Kirkpatrick 1975
first application in biology: Amit, Gutfreund & Sompolinksy 1985

- Consider processes with many fixed (pseudo-)random parameters $\boldsymbol{\xi}$, distributed according to $\mathcal{P}(\boldsymbol{\xi})$

$$p(\mathbf{x}|\boldsymbol{\xi}) = \frac{e^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x},\boldsymbol{\xi})}}{Z(\boldsymbol{\lambda},\boldsymbol{\xi})}, \qquad Z(\boldsymbol{\lambda},\boldsymbol{\xi}) = \sum_{\mathbf{x}\in X} e^{\sum_{\ell=1}^{L} \lambda_\ell \omega_\ell(\mathbf{x},\boldsymbol{\xi})}$$

  - calculating state averages $\langle f \rangle_{\boldsymbol{\xi}}$ for each realisation of $\boldsymbol{\xi}$ is usually impossible
  - we are mostly interested in *typical* values of state averages
  - for $N \to \infty$ macroscopic averages will not depend on $\boldsymbol{\xi}$, only on $\mathcal{P}(\boldsymbol{\xi})$, 'self-averaging': $\lim_{N\to\infty} \langle f \rangle_{\boldsymbol{\xi}}$ indep of $\boldsymbol{\xi}$

  so focus on

$$\overline{\langle f \rangle_{\boldsymbol{\xi}}} = \sum_{\boldsymbol{\xi}} \mathcal{P}(\boldsymbol{\xi}) \langle f \rangle_{\boldsymbol{\xi}} = \sum_{\boldsymbol{\xi}} \mathcal{P}(\boldsymbol{\xi}) \Big\{ \sum_{\mathbf{x}\in X} p(\mathbf{x}|\boldsymbol{\xi}) f(\mathbf{x},\boldsymbol{\xi}) \Big\}$$

- new generating function:

$$\overline{F}(\boldsymbol{\lambda},\mu)=\sum_{\boldsymbol{\xi}}\mathcal{P}(\boldsymbol{\xi})\log Z(\boldsymbol{\lambda},\mu,\boldsymbol{\xi}),\qquad Z(\boldsymbol{\lambda},\mu,\boldsymbol{\xi})=\sum_{\mathbf{x}\in X}\mathrm{e}^{\mu\psi(\mathbf{x},\boldsymbol{\xi})+\sum_\ell\lambda_\ell\omega_\ell(\mathbf{x},\boldsymbol{\xi})}$$

$$\lim_{\mu\to0}\frac{\partial}{\partial\mu}\overline{F}(\boldsymbol{\lambda},\mu)=\lim_{\mu\to0}\sum_{\boldsymbol{\xi}}\mathcal{P}(\boldsymbol{\xi})\left\{\frac{\sum_{\mathbf{x}\in X}\psi(\mathbf{x},\boldsymbol{\xi})\mathrm{e}^{\mu\psi(\mathbf{x},\boldsymbol{\xi})+\sum_\ell\lambda_\ell\omega_\ell(\mathbf{x},\boldsymbol{\xi})}}{\sum_{\mathbf{x}\in X}\mathrm{e}^{\mu\psi(\mathbf{x},\boldsymbol{\xi})+\sum_\ell\lambda_\ell\omega_\ell(\mathbf{x},\boldsymbol{\xi})}}\right\}$$

$$=\sum_{\boldsymbol{\xi}}\mathcal{P}(\boldsymbol{\xi})\left\{\frac{\sum_{\mathbf{x}\in X}\psi(\mathbf{x},\boldsymbol{\xi})\mathrm{e}^{\sum_\ell\lambda_\ell\omega_\ell(\mathbf{x},\boldsymbol{\xi})}}{\sum_{\mathbf{x}\in X}\mathrm{e}^{\sum_\ell\lambda_\ell\omega_\ell(\mathbf{x},\boldsymbol{\xi})}}\right\}=\overline{\langle\psi\rangle_{\boldsymbol{\xi}}}$$

- main obstacle in calculating $\overline{F}$:
  the logarithm ...

  replica identity : $\overline{\log Z}=\lim_{n\to0}\frac{1}{n}\log\overline{Z^n}$

  proof:

$$\lim_{n\to0}\frac{1}{n}\log\overline{Z^n}=\lim_{n\to0}\frac{1}{n}\log\overline{[\mathrm{e}^{n\log Z}]}=\lim_{n\to0}\frac{1}{n}\log\overline{[1+n\log Z+\mathcal{O}(n^2)]}$$

$$=\lim_{n\to0}\frac{1}{n}\log[1+n\overline{\log Z}+\mathcal{O}(n^2)]=\overline{\log Z}$$

- apply $\overline{\log Z} = \lim_{n\to 0} \frac{1}{n} \log \overline{Z^n}$
  (simplest case $L = 1$)

$$\begin{aligned}
\overline{F}(\lambda) &= \sum_{\boldsymbol{\xi}} \mathcal{P}(\boldsymbol{\xi}) \log \Big[\sum_{\mathbf{x}\in X} \mathrm{e}^{\lambda\omega(\mathbf{x},\boldsymbol{\xi})} \Big] = \lim_{n\to 0} \frac{1}{n} \log \sum_{\boldsymbol{\xi}} \mathcal{P}(\boldsymbol{\xi}) \Big[\sum_{\mathbf{x}\in X} \mathrm{e}^{\lambda\omega(\mathbf{x},\boldsymbol{\xi})} \Big]^n \\
&= \lim_{n\to 0} \frac{1}{n} \log \sum_{\boldsymbol{\xi}} \mathcal{P}(\boldsymbol{\xi}) \Big[\sum_{\mathbf{x}^1\in X} \cdots \sum_{\mathbf{x}^n\in X} \mathrm{e}^{\lambda \sum_{\alpha=1}^n \omega(\mathbf{x}^\alpha,\boldsymbol{\xi})} \Big] \\
&= \lim_{n\to 0} \frac{1}{n} \log \Big[\sum_{\mathbf{x}^1\in X} \cdots \sum_{\mathbf{x}^n\in X} \sum_{\boldsymbol{\xi}} \mathcal{P}(\boldsymbol{\xi}) \mathrm{e}^{\lambda \sum_{\alpha=1}^n \omega(\mathbf{x}^\alpha,\boldsymbol{\xi})} \Big]
\end{aligned}$$

- notes:
  - impossible $\boldsymbol{\xi}$-average converted into simpler one ...
  - calculation involves n 'replicas' $\mathbf{x}^\alpha$ of original system
  - but $n \to 0$ at the end ... ?
  - penultimate step true only for *integer* n,
    so limit requires *analytical continuation* ...

since then: alternative (more tedious) routes,
these confirmed correctness of the replica method!

The replica trick and algorithms

Suppose we have data D, with prob distr $\mathcal{P}(D)$
and an algorithm which minimises an error function $E(D,\boldsymbol{\theta})$
(maximum likelihood, Cox & Bayesian regression, SVM, perceptron, ...)

- algorithm outcome:

$$\boldsymbol{\theta}^\star(D) = \arg\min_{\boldsymbol{\theta}} E(D,\boldsymbol{\theta}), \qquad E_{\rm min}(D) = \min_{\boldsymbol{\theta}} E(D,\boldsymbol{\theta})$$

  typical performance:

$$\boldsymbol{\theta}^\star = \sum_D \mathcal{P}(D)\boldsymbol{\theta}^\star(D) = \overline{\boldsymbol{\theta}^\star(D)} \qquad E_{\rm min} = \sum_D \mathcal{P}(D) E_{\rm min}(D) = \overline{E_{\rm min}(D)}$$

- steepest descent identity & replica trick:

$$\begin{aligned}
E_{\rm min}(D) &= \min_{\boldsymbol{\theta}} E(D,\boldsymbol{\theta}) = -\lim_{\beta\to\infty}\frac{1}{\beta}\log\int {\rm d}\boldsymbol{\theta}\ {\rm e}^{-\beta E(D,\boldsymbol{\theta})} \\
E_{\rm min} &= \overline{E_{\rm min}(D)} = -\lim_{\beta\to\infty}\frac{1}{\beta}\overline{\log\int {\rm d}\boldsymbol{\theta}\ {\rm e}^{-\beta E(D,\boldsymbol{\theta})}} \\
&= -\lim_{\beta\to\infty}\lim_{n\to 0}\frac{1}{\beta n}\log\overline{\left[\int {\rm d}\boldsymbol{\theta}\ {\rm e}^{-\beta E(D,\boldsymbol{\theta})}\right]^n} \\
&= -\lim_{\beta\to\infty}\lim_{n\to 0}\frac{1}{\beta n}\log\int {\rm d}\boldsymbol{\theta}^1\ldots\boldsymbol{\theta}^n\ \overline{{\rm e}^{-\beta\sum_{\alpha=1}^n E(D,\boldsymbol{\theta}^\alpha)}}
\end{aligned}$$

Alternative forms of the replica identity

suppose we need averages, but for
a $p(\mathbf{x}|\boldsymbol{\xi})$ that is not of an exponential form?

or we need to average quantities that we
don't want in the exponent of $Z(\lambda\boldsymbol{\xi})$?

$$p(\mathbf{x}|\boldsymbol{\xi}) = \frac{W(\mathbf{x},\boldsymbol{\xi})}{\sum_{\mathbf{x}'\in X} W(\mathbf{x}',\boldsymbol{\xi})}, \qquad \overline{\langle f\rangle_{\boldsymbol{\xi}}} = \overline{\sum_{\mathbf{x}\in X} p(\mathbf{x}|\boldsymbol{\xi}) f(\mathbf{x},\boldsymbol{\xi})}$$

- main obstacle here:
  the fraction ...

$$\begin{aligned}
\overline{\langle f\rangle_{\boldsymbol{\xi}}} &= \overline{\Big[\frac{\sum_{\mathbf{x}\in X} W(\mathbf{x},\boldsymbol{\xi}) f(\mathbf{x},\boldsymbol{\xi})}{\sum_{\mathbf{x}\in X} W(\mathbf{x},\boldsymbol{\xi})}\Big]} = \overline{\Big[\sum_{\mathbf{x}\in X} W(\mathbf{x},\boldsymbol{\xi}) f(\mathbf{x},\boldsymbol{\xi})\Big]\Big[\sum_{\mathbf{x}\in X} W(\mathbf{x},\boldsymbol{\xi})\Big]^{-1}} \\
&= \lim_{n\to 0} \overline{\Big[\sum_{\mathbf{x}\in X} W(\mathbf{x},\boldsymbol{\xi}) f(\mathbf{x},\boldsymbol{\xi})\Big]\Big[\sum_{\mathbf{x}\in X} W(\mathbf{x},\boldsymbol{\xi})\Big]^{n-1}} \\
&= \lim_{n\to 0} \sum_{\mathbf{x}^1\in X} \ldots \sum_{\mathbf{x}^n\in X} \overline{f(\mathbf{x}^1,\boldsymbol{\xi}) W(\mathbf{x}^1,\boldsymbol{\xi}) \ldots W(\mathbf{x}^n,\boldsymbol{\xi})}
\end{aligned}$$

(again: used integer n, but $n \to 0$ ...)

- equivalence between two forms of replica identity, if

$$W(\mathbf{x},\boldsymbol{\xi}) = \mathrm{e}^{\sum_\ell \lambda_\ell \phi_\ell(\mathbf{x},\boldsymbol{\xi})}$$

proof:

$$\begin{aligned}
\overline{\langle f\rangle_{\boldsymbol{\xi}}} &= \lim_{n\to 0} \sum_{\mathbf{x}^1\in X}\cdots\sum_{\mathbf{x}^n\in X} \overline{f(\mathbf{x}^1,\boldsymbol{\xi})W(\mathbf{x}^1,\boldsymbol{\xi})\ldots W(\mathbf{x}^n,\boldsymbol{\xi})} \\
&= \lim_{n\to 0} \sum_{\mathbf{x}^1\in X}\cdots\sum_{\mathbf{x}^n\in X} \overline{f(\mathbf{x}^1,\boldsymbol{\xi})\,\mathrm{e}^{\sum_{\alpha=1}^n \sum_\ell \lambda_\ell \phi_\ell(\mathbf{x}^\alpha,\boldsymbol{\xi})}} \\
&= \lim_{n\to 0} \frac{1}{n} \sum_{\mathbf{x}^1\in X}\cdots\sum_{\mathbf{x}^n\in X} \overline{\Big[\sum_{\alpha=1}^n f(\mathbf{x}^\alpha,\boldsymbol{\xi})\Big]\,\mathrm{e}^{\sum_{\alpha=1}^n \sum_\ell \lambda_\ell \phi_\ell(\mathbf{x}^\alpha,\boldsymbol{\xi})}} \\
&= \lim_{n\to 0} \frac{1}{n} \lim_{\mu\to 0} \frac{\partial}{\partial\mu} \sum_{\mathbf{x}^1\in X}\cdots\sum_{\mathbf{x}^n\in X} \overline{\mathrm{e}^{\sum_{\alpha=1}^n \sum_\ell \lambda_\ell \phi_\ell(\mathbf{x}^\alpha,\boldsymbol{\xi}) + \mu \sum_{\alpha=1}^n f(\mathbf{x}^\alpha,\boldsymbol{\xi})}} \\
&= \lim_{\mu\to 0} \frac{\partial}{\partial\mu} \lim_{n\to 0} \frac{1}{n} \overline{\sum_{\mathbf{x}^1\in X}\cdots\sum_{\mathbf{x}^n\in X} \mathrm{e}^{\sum_{\alpha=1}^n \left[\sum_\ell \lambda_\ell \phi_\ell(\mathbf{x}^\alpha,\boldsymbol{\xi}) + \mu f(\mathbf{x}^\alpha,\boldsymbol{\xi})\right]}} \\
&= \lim_{\mu\to 0} \frac{\partial}{\partial\mu} \lim_{n\to 0} \frac{1}{n} \overline{Z^n(\boldsymbol{\lambda},\mu,\boldsymbol{\xi})}, \qquad Z(\boldsymbol{\lambda},\mu,\boldsymbol{\xi}) = \sum_{\mathbf{x}\in X} \mathrm{e}^{\sum_\ell \lambda_\ell \phi_\ell(\mathbf{x},\boldsymbol{\xi}) + \mu f(\mathbf{x},\boldsymbol{\xi})}
\end{aligned}$$

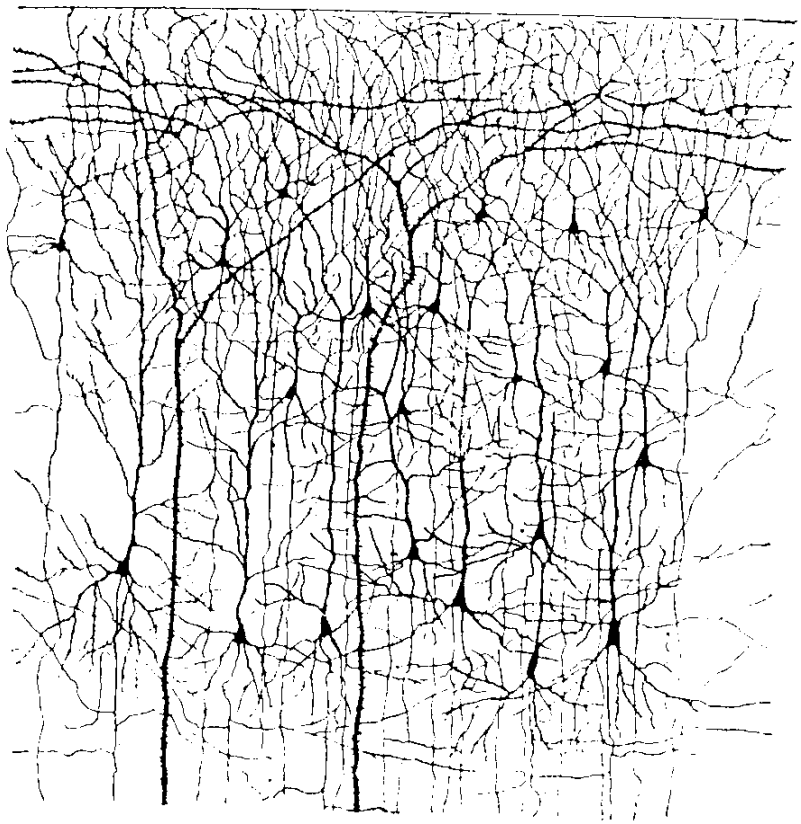
1. The replica method
   - Exponential families and generating functions
   - The replica trick
   - The replica trick and algorithms
   - Alternative forms of the replica identity

2. Application: information storage in neural networks
   - Attractor neural networks
   - The replica calculation
   - Replica symmetry
   - Replica symmetric solution

3. Application: overfitting transition in linear separators
   - Linear separability of data – version space
   - The replica calculation
   - Gardner's replica symmetric theory

Attractor neural networks

$N \sim 10^{12-14}$ brain cells (neurons),
each connected with $\sim 10^{3-5}$ others



- **neurons**

  two states:
  $\sigma_i = 1$ (i fires electric pulses)
  $\sigma_i = -1$ (i is at rest)

- **dynamics of firing states**

$$\sigma_i(t+1) = sgn\Big[\overbrace{\sum_{j=1}^{N} J_{ij}\sigma_j(t)}^{activation\ signal} + \overbrace{\theta_i + z_i(t)}^{threshold,\ noise} \Big]$$

$\theta_i \in \mathbb{R}$: *firing threshold of neuron i*
$J_{ij} \in \mathbb{R}$: *synaptic connection $j \to i$*

learning = adaptation of $\{J_{ij}, \theta_i\}$

non-local 'distributed' storage of 'program' and 'data'

attractor neural networks

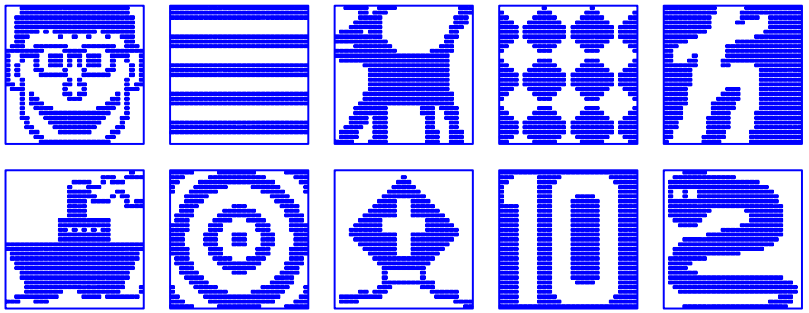
models for associative memory in the brain

- **the neural code**

  represent 'patterns' as
  micro-states $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)$

  $\circ$: $\sigma_i = -1$, $\bullet$: $\sigma_i = 1$

  e.g. $N = 400$,
  10 patterns:



- **information storage**

  modify synapses $\{J_{ij}\}$ such that $\boldsymbol{\xi}$ is
  stable state (attractor) of the neuronal dynamics

- **information recall**

  initial state $\boldsymbol{\sigma}(t = 0)$:
  evolution to nearest attractor

  if $\boldsymbol{\sigma}(0)$ close (i.e. similar) to $\boldsymbol{\xi}$:
  $\boldsymbol{\sigma}(t = \infty) = \boldsymbol{\xi}$

- learning rule: recipe for storing patterns via modification of $\{J_{ij}\}$
  Hebb (1949): $\Delta J_{ij} \propto \xi_i \xi_j$

  choose $J_{ij} = J_0 \xi_i \xi_j$, $\theta_i = 0$,
  update randomly drawn i at each step:

$$\sigma_i(t+1) = sgn\Big[\sum_{j=1}^{N} J_{ij}\sigma_j(t) + z_i(t)\Big] = sgn\Big[J_0 \xi_i \overbrace{\Big(\sum_{j=1}^{N} \xi_j \sigma_j(t)\Big)}^{\text{pattern overlap}} + z_i(t)\Big]$$

$$= \xi_i \; sgn\Big[J_0 \sum_{j=1}^{N} \xi_j \sigma_j(t) + \xi_i z_i(t)\Big]$$

  $M(t) = \sum_{j=1}^{N} \xi_j \sigma_j(t)$ *sufficiently large:* $\sigma_i(t+1) = \xi_i$
  now $M(t+1) \geq M(t)$ *...*
  will continue until $\boldsymbol{\sigma} = \boldsymbol{\xi}$

- proper analysis:

  noise: $P(z) = \frac{\beta}{2}[1 - \tanh^2(\beta z)]$,
  symmetric synapses: $J_{ij} = J_{ji}$, $J_{ii} = 0$
  sequential updates of σ_i

$$p(\boldsymbol{\sigma}) = \frac{e^{-\beta H(\boldsymbol{\sigma})}}{Z(\beta)}, \qquad H(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{i \neq j} \sigma_i J_{ij} \sigma_j - \sum_i \theta_i \sigma_i$$

a more realistic model,
solvable via the replica method

- **storage of a pattern $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N) \in \{-1, 1\}^N$**
  on background of zero-average Gaussian synapses

$$J_{ij} = \frac{J_0}{N}\xi_i\xi_j + \frac{J}{\sqrt{N}}z_{ij}, \qquad \overline{z}_{ij} = 0,\ \overline{z_{ij}}^2 = 1, \quad J, J_0 \geq 0, \quad \theta_i = 0$$

  to be averaged over: background synapses $\{z_{ij}\}$
  pattern overlap: $m(\boldsymbol{\sigma}) = \frac{1}{N}\sum_k \sigma_k\xi_k$

$$\begin{aligned} H(\boldsymbol{\sigma}) &= -\frac{1}{2}\sum_{i\neq j}\sigma_i\sigma_j\Big\{\frac{J_0}{N}\xi_i\xi_j + \frac{J}{\sqrt{N}}z_{ij}\Big\} \\ &= -\frac{J_0}{2N}\sum_{ij}\sigma_i\sigma_j\xi_i\xi_j + \frac{J_0}{2N}\sum_i 1 - \frac{J}{2\sqrt{N}}\sum_{i\neq j}\sigma_i\sigma_j z_{ij} \\ &= -\frac{1}{2}NJ_0m^2(\boldsymbol{\sigma}) + \frac{1}{2}J_0 - \frac{J}{\sqrt{N}}\sum_{i<j}\sigma_i\sigma_j z_{ij} \end{aligned}$$

- **generating function**

$$\overline{F} = \overline{\log Z(\beta)} = \lim_{n\to 0}\frac{1}{n}\log\overline{Z^n(\beta)} = \lim_{n\to 0}\frac{1}{n}\log\Big[\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n}\overline{e^{-\beta\sum_{\alpha=1}^n H(\boldsymbol{\sigma}^\alpha)}}\Big]$$

The replica calculation

short-hands: $m(\boldsymbol{\sigma}) = \frac{1}{N}\sum_i \xi_i \sigma_i, \quad \mathrm{D}z = (2\pi)^{-1/2}\mathrm{e}^{-z^2/2}\mathrm{d}z$

Gaussian integral: $\int \mathrm{D}z\, \mathrm{e}^{xz} = \mathrm{e}^{\frac{1}{2}x^2}$

- **average over random synapses**

$$
\begin{aligned}
\overline{Z^n(\beta)} &= \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \overline{\mathrm{e}^{-\beta \sum_{\alpha=1}^n H(\boldsymbol{\sigma}^\alpha)}} \\
&= \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \overline{\mathrm{e}^{-\beta \sum_{\alpha=1}^n \left[\frac{1}{2}J_0 - \frac{1}{2}NJ_0 m^2(\boldsymbol{\sigma}^\alpha) - \frac{J}{\sqrt{N}} \sum_{i<j} \sigma_i^\alpha \sigma_j^\alpha z_{ij} \right]}} \\
&= e^{-\frac{1}{2}n\beta J_0} \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \mathrm{e}^{\frac{1}{2}N\beta J_0 \sum_{\alpha=1}^n m^2(\boldsymbol{\sigma}^\alpha)} \overline{\mathrm{e}^{\frac{\beta J}{\sqrt{N}} \sum_{\alpha=1}^n \sum_{i<j} \sigma_i^\alpha \sigma_j^\alpha z_{ij}}} \\
&= e^{-\frac{1}{2}n\beta J_0} \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \mathrm{e}^{\frac{1}{2}N\beta J_0 \sum_{\alpha=1}^n m^2(\boldsymbol{\sigma}^\alpha)} \prod_{i<j} \int \mathrm{D}z\, \mathrm{e}^{\frac{\beta J}{\sqrt{N}} \sum_{\alpha=1}^n \sigma_i^\alpha \sigma_j^\alpha z} \\
&= e^{-\frac{1}{2}n\beta J_0} \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \mathrm{e}^{\frac{1}{2}N\beta J_0 \sum_{\alpha=1}^n m^2(\boldsymbol{\sigma}^\alpha)} \prod_{i<j} \mathrm{e}^{\frac{\beta^2 J^2}{2N} \left[\sum_{\alpha=1}^n \sigma_i^\alpha \sigma_j^\alpha \right]^2} \\
&= e^{-\frac{1}{2}n\beta J_0} \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \mathrm{e}^{N\left[\frac{1}{2}\beta J_0 \sum_{\alpha=1}^n m^2(\boldsymbol{\sigma}^\alpha) + \frac{1}{2}(\beta J)^2 \sum_{\alpha,\gamma=1}^n \left(N^{-2} \sum_{i<j} \sigma_i^\alpha \sigma_j^\alpha \sigma_i^\gamma \sigma_j^\gamma \right) \right]}
\end{aligned}
$$

- **complete square** in sums over neurons

$$\sum_{i<j}\sigma_i^\alpha\sigma_j^\alpha\sigma_i^\gamma\sigma_j^\gamma = \frac{1}{2}\sum_{i\neq j}\sigma_i^\alpha\sigma_j^\alpha\sigma_i^\gamma\sigma_j^\gamma = \frac{1}{2}\sum_{ij}\sigma_i^\alpha\sigma_j^\alpha\sigma_i^\gamma\sigma_j^\gamma - \frac{1}{2}\sum_i 1$$

$$= \frac{1}{2}\Big(\sum_i \sigma_i^\alpha\sigma_i^\gamma\Big)^2 - \frac{1}{2}N$$

hence

$$\overline{Z^n(\beta)} = e^{-\frac{1}{2}n\beta J_0}\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n} e^{N\left[\frac{1}{2}\beta J_0\sum_{\alpha=1}^n m^2(\boldsymbol{\sigma}^\alpha)+\frac{1}{4}(\beta J)^2\sum_{\alpha,\gamma=1}^n\left(\left(\frac{1}{N}\sum_i\sigma_i^\alpha\sigma_i^\gamma\right)^2-\frac{1}{N}\right)\right]}$$

$$= e^{-\frac{1}{2}n\beta J_0-\frac{1}{4}n(\beta J)^2}\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n} e^{N\left[\frac{1}{2}\beta J_0\sum_{\alpha=1}^n m^2(\boldsymbol{\sigma}^\alpha)+\frac{1}{4}(\beta J)^2\sum_{\alpha,\gamma=1}^n\left(\frac{1}{N}\sum_i\sigma_i^\alpha\sigma_i^\gamma\right)^2\right]}$$

- insert:

$$1 = \prod_{\alpha=1}^n\int \mathrm{d}m_\alpha\;\delta\Big(m_\alpha-\frac{1}{N}\sum_i\xi_i\sigma_i^\alpha\Big),\qquad 1 = \prod_{\alpha,\gamma=1}^n\int \mathrm{d}q_{\alpha\gamma}\;\delta\Big(q_{\alpha\gamma}-\frac{1}{N}\sum_i\sigma_i^\alpha\sigma_i^\gamma\Big)$$

$\mathbf{m}\in\mathbb{R}^n,\ \mathbf{q}\in\mathbb{R}^{n^2}$:

$$\overline{Z^n(\beta)} = e^{-\frac{1}{2}n\beta J_0-\frac{1}{4}n(\beta J)^2}\int \mathrm{d}\mathbf{m}\mathrm{d}\mathbf{q}\;e^{N\left[\frac{1}{2}\beta J_0\sum_{\alpha=1}^n m_\alpha^2+\frac{1}{4}(\beta J)^2\sum_{\alpha,\gamma=1}^n q_{\alpha\gamma}^2\right]}$$
$$\times\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n}\Big[\prod_{\alpha=1}^n\delta\Big(m_\alpha-\frac{1}{N}\sum_i\xi_i\sigma_i^\alpha\Big)\Big]\Big[\prod_{\alpha,\gamma=1}^n\delta\Big(q_{\alpha\gamma}-\frac{1}{N}\sum_i\sigma_i^\alpha\sigma_i^\gamma\Big)\Big]$$

remember: $\delta(x) = (2\pi)^{-1} \int \mathrm{d}\hat{x}\, \mathrm{e}^{\mathrm{i}x\hat{x}}$

- the **sum over neuron state variables**

$$\sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \Big[\prod_{\alpha=1}^{n} \delta\Big(m_\alpha - \frac{1}{N}\sum_i \xi_i \sigma_i^\alpha\Big)\Big] \Big[\prod_{\alpha,\gamma=1}^{n} \delta\Big(q_{\alpha\gamma} - \frac{1}{N}\sum_i \sigma_i^\alpha \sigma_i^\gamma\Big)\Big]$$

$$= \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \int \frac{\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}}{(2\pi)^{n^2+n}} \mathrm{e}^{\mathrm{i}\sum_{\alpha=1}^n \hat{m}_\alpha \left[m_\alpha - \frac{1}{N}\sum_i \xi_i \sigma_i^\alpha\right] + \mathrm{i}\sum_{\alpha,\gamma=1}^n \hat{q}_{\alpha\gamma}\left[q_{\alpha\gamma} - \frac{1}{N}\sum_i \sigma_i^\alpha \sigma_i^\gamma\right]}$$

$$= \int \frac{\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}}{(2\pi)^{n(n+1)}} \mathrm{e}^{\mathrm{i}\sum_\alpha \hat{m}_\alpha m_\alpha + \mathrm{i}\sum_{\alpha\gamma}\hat{q}_{\alpha\gamma} q_{\alpha\gamma}} \sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \prod_{i=1}^{N} \mathrm{e}^{-\frac{\mathrm{i}}{N}\left[\sum_\alpha \hat{m}_\alpha \xi_i \sigma_i^\alpha + \sum_{\alpha\gamma}\hat{q}_{\alpha\gamma}\sigma_i^\alpha \sigma_i^\gamma\right]}$$

$$= \int \frac{\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}}{(2\pi)^{n(n+1)}} \mathrm{e}^{\mathrm{i}\sum_\alpha \hat{m}_\alpha m_\alpha + \mathrm{i}\sum_{\alpha\gamma}\hat{q}_{\alpha\gamma} q_{\alpha\gamma}} \prod_i \sum_{\sigma_1 \ldots \sigma_n} \mathrm{e}^{-\frac{\mathrm{i}}{N}\left[\sum_\alpha \hat{m}_\alpha \xi_i \sigma_\alpha + \sum_{\alpha\gamma}\hat{q}_{\alpha\gamma}\sigma_\alpha \sigma_\gamma\right]}$$

transform: $\hat{\mathbf{m}} \to N\hat{\mathbf{m}}$, $\hat{\mathbf{q}} \to N\hat{\mathbf{q}}$, $\sigma_\alpha \to \xi_i \sigma_\alpha$:

$$\sum_{\boldsymbol{\sigma}^1 \ldots \boldsymbol{\sigma}^n} \Big[\ldots\Big]\Big[\ldots\Big] = \int \frac{\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}}{(2\pi/N)^{n(n+1)}} \mathrm{e}^{\mathrm{i}N\left[\hat{\mathbf{m}}\cdot\mathbf{m} + \mathrm{Tr}(\hat{\mathbf{q}}\mathbf{q})\right]} \Big[\sum_{\boldsymbol{\sigma}\in\{-1,1\}^n} \mathrm{e}^{-\mathrm{i}\hat{\mathbf{m}}\cdot\boldsymbol{\sigma} - \mathrm{i}\boldsymbol{\sigma}\cdot\hat{\mathbf{q}}\boldsymbol{\sigma}}\Big]^N$$

$$= \int \frac{\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}}{(2\pi/N)^{n(n+1)}} \mathrm{e}^{\mathrm{i}N\hat{\mathbf{m}}\cdot\mathbf{m} + \mathrm{i}N\mathrm{Tr}(\hat{\mathbf{q}}\mathbf{q}) + N\log\sum_{\boldsymbol{\sigma}} \exp(-\mathrm{i}\hat{\mathbf{m}}\cdot\boldsymbol{\sigma} - \mathrm{i}\boldsymbol{\sigma}\cdot\hat{\mathbf{q}}\boldsymbol{\sigma})}$$

- combine everything ...

$$\overline{Z^n(\beta)} = e^{-\frac{1}{2}n\beta J_0 - \frac{1}{4}n(\beta J)^2 - n(n+1)\log(2\pi/N)} \int \mathrm{d}\mathbf{m}\mathrm{d}\mathbf{q}\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}\ \mathrm{e}^{N\Psi(\mathbf{m},\mathbf{q},\hat{\mathbf{m}},\hat{\mathbf{q}})}$$

$$\Psi(\ldots) = \frac{1}{2}\beta J_0 \mathbf{m}^2 + \frac{1}{4}(\beta J)^2 \mathrm{Tr}(\mathbf{q}^2) + \mathrm{i}\hat{\mathbf{m}}\cdot\mathbf{m} + \mathrm{i}\mathrm{Tr}(\hat{\mathbf{q}}\mathbf{q}) + \log \sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\mathrm{i}\hat{\mathbf{m}}\cdot\boldsymbol{\sigma} - \mathrm{i}\boldsymbol{\sigma}\cdot\hat{\mathbf{q}}\boldsymbol{\sigma}}$$

Hence

$$\begin{aligned} \overline{F} &= \lim_{n\to 0} \frac{1}{n} \log \overline{Z^n(\beta)} \\ &= -\frac{1}{2}\beta J_0 - \frac{1}{4}(\beta J)^2 - \log(\frac{2\pi}{N}) + \lim_{n\to 0} \frac{1}{n} \log \int \mathrm{d}\mathbf{m}\mathrm{d}\mathbf{q}\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}\ \mathrm{e}^{N\Psi(\mathbf{m},\mathbf{q},\hat{\mathbf{m}},\hat{\mathbf{q}})} \end{aligned}$$

- Since $\overline{F} = \mathcal{O}(N)$,
  large N behaviour follows from

$$\bar{f} = \lim_{N\to\infty} \overline{F}/N = \lim_{N\to\infty} \lim_{n\to 0} \frac{1}{nN} \log \int \mathrm{d}\mathbf{m}\mathrm{d}\mathbf{q}\mathrm{d}\hat{\mathbf{m}}\mathrm{d}\hat{\mathbf{q}}\ \mathrm{e}^{N\Psi(\mathbf{m},\mathbf{q},\hat{\mathbf{m}},\hat{\mathbf{q}})}$$

- assume limits commute,
  steepest descent integration:

$$\bar{f} = \lim_{n\to 0} \frac{1}{n} \mathrm{extr}_{\mathbf{m},\mathbf{q},\hat{\mathbf{m}},\hat{\mathbf{q}}} \Psi(\mathbf{m},\mathbf{q},\hat{\mathbf{m}},\hat{\mathbf{q}})$$

$$\Psi(\ldots) = \frac{1}{2}\beta J_0 \sum_\alpha m_\alpha^2 + \frac{1}{4}(\beta J)^2 \sum_{\alpha\gamma} q_{\alpha\gamma}^2 + \mathrm{i}\sum_\alpha \hat{m}_\alpha m_\alpha + \mathrm{i}\sum_{\alpha\gamma} \hat{q}_{\alpha\gamma} q_{\alpha\gamma}$$
$$+ \log \sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\mathrm{i}\sum_\lambda \hat{m}_\lambda \sigma_\lambda - \mathrm{i}\sum_{\lambda\zeta} \sigma_\lambda \hat{q}_{\lambda\zeta} \sigma_\zeta}$$

- **saddle-point eqns**

$$\frac{\partial \Psi}{\partial m_\alpha} = 0,\ \frac{\partial \Psi}{\partial q_{\alpha\gamma}} = 0: \quad \beta J_0 m_\alpha + \mathrm{i}\hat{m}_\alpha = 0,\quad \frac{1}{2}(\beta J)^2 q_{\alpha\gamma} + \mathrm{i}\hat{q}_{\alpha\gamma} = 0$$

$$\frac{\partial \Psi}{\partial \hat{m}_\alpha} = 0: \quad \mathrm{i}m_\alpha - \mathrm{i}\frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \mathrm{e}^{-\mathrm{i}\sum_\lambda \hat{m}_\lambda \sigma_\lambda - \mathrm{i}\sum_{\lambda\zeta} \sigma_\lambda \hat{q}_{\lambda\zeta} \sigma_\zeta}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\mathrm{i}\sum_\lambda \hat{m}_\lambda \sigma_\lambda - \mathrm{i}\sum_{\lambda\zeta} \sigma_\lambda \hat{q}_{\lambda\zeta} \sigma_\zeta}} = 0$$

$$\frac{\partial \Psi}{\partial \hat{q}_{\alpha\gamma}} = 0: \quad \mathrm{i}q_{\alpha\gamma} - \mathrm{i}\frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \sigma_\gamma \mathrm{e}^{-\mathrm{i}\sum_\lambda \hat{m}_\lambda \sigma_\lambda - \mathrm{i}\sum_{\lambda\zeta} \sigma_\lambda \hat{q}_{\lambda\zeta} \sigma_\zeta}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\mathrm{i}\sum_\lambda \hat{m}_\lambda \sigma_\lambda - \mathrm{i}\sum_{\lambda\zeta} \sigma_\lambda \hat{q}_{\lambda\zeta} \sigma_\zeta}} = 0$$

- **eliminate** $(\hat{\mathbf{m}}, \hat{\mathbf{q}})$

$$m_\alpha = \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \mathrm{e}^{\beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda\neq\zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda\neq\zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta}}$$

$$q_{\alpha\gamma} = \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \sigma_\gamma \mathrm{e}^{\beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda\neq\zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda\neq\zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta}}$$

trivial soln: $\mathbf{m} = \mathbf{q} = \mathbf{0}$, *any others?*

Replica symmetry

- $\beta = 0$ (infinite noise level):

$$m_\alpha = \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \mathrm{e}^0}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^0} = 0, \qquad q_{\alpha\gamma} = \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \sigma_\gamma \mathrm{e}^0}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^0} = 0 \qquad \mathbf{m = q = 0} \;\; \textit{if} \;\; \beta = 0$$

- bifurcations from trivial soln:

$$\begin{aligned} m_\alpha &= \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \left[1 + \beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda \neq \zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta\right]}{\sum_{\boldsymbol{\sigma}} \left[1 + \beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda \neq \zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta\right]} + \mathcal{O}(\mathbf{m}, \mathbf{q})^2 \\ &= \frac{2^n \beta J_0 m_\alpha}{2^n} + \ldots = \beta J_0 m_\alpha + \ldots \qquad \mathbf{m \neq 0} \;\; \textit{if} \;\; \beta J_0 > 1 \end{aligned}$$

$$\begin{aligned} q_{\alpha\gamma} &= \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \sigma_\gamma \left[1 + \beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda \neq \zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta\right]}{\sum_{\boldsymbol{\sigma}} \left[1 + \beta J_0 \sum_\lambda m_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 \sum_{\lambda \neq \zeta} \sigma_\lambda q_{\lambda\zeta} \sigma_\zeta\right]} + \mathcal{O}(\mathbf{m}, \mathbf{q})^2 \\ &= \frac{2^n (\beta J)^2 q_{\alpha\gamma} + \ldots}{2^n} + \ldots = (\beta J)^2 q_{\alpha\gamma} + \ldots \qquad \mathbf{q \neq 0} \;\; \textit{if} \;\; \beta J > 1 \end{aligned}$$

how to find form of nontrivial solns $\{m_\alpha, q_{\alpha\gamma}\}$?
need their physical interpretation!
use alternative form(s) of replica identity:

$$\overline{\langle f(\boldsymbol{\sigma})\rangle} = \lim_{n\to 0}\frac{1}{n}\sum_{\gamma=1}^{n}\sum_{\boldsymbol{\sigma}^1}\cdots\sum_{\boldsymbol{\sigma}^n}\overline{f(\boldsymbol{\sigma}^\gamma)\mathrm{e}^{-\beta\sum_{\alpha=1}^{n}H(\boldsymbol{\sigma}^\alpha)}}$$

$$\overline{\langle\langle f(\boldsymbol{\sigma},\boldsymbol{\sigma}')\rangle\rangle} = \lim_{n\to 0}\frac{1}{n(n-1)}\sum_{\alpha\neq\gamma=1}^{n}\sum_{\boldsymbol{\sigma}^1}\cdots\sum_{\boldsymbol{\sigma}^n}\overline{f(\boldsymbol{\sigma}^\alpha,\boldsymbol{\sigma}^\gamma)\mathrm{e}^{-\beta\sum_{\alpha=1}^{n}H(\boldsymbol{\sigma}^\alpha)}}$$

apply to $P(m|\boldsymbol{\sigma}) = \delta\Big[m - \frac{1}{N}\sum_{i=1}^{N}\xi_i\sigma_i\Big], \quad P(q|\boldsymbol{\sigma},\boldsymbol{\sigma}') = \delta\Big[q - \frac{1}{N}\sum_{i=1}^{N}\sigma_i\sigma_i'\Big]$

repeat steps of previous calculation,
gives expressions in terms of
saddle-point soln $\{m_\alpha, q_{\alpha\gamma}\}$:

$$\lim_{N\to\infty}\overline{\langle P(m|\boldsymbol{\sigma})\rangle} = \lim_{n\to 0}\frac{1}{n}\sum_{\alpha=1}^{n}\delta[m-m_\alpha]$$

$$\lim_{N\to\infty}\overline{\langle\langle P(q|\boldsymbol{\sigma},\boldsymbol{\sigma}')\rangle\rangle} = \lim_{n\to 0}\frac{1}{n(n-1)}\sum_{\alpha\neq\gamma=1}^{n}\delta[q-q_{\alpha\gamma}]$$

ergodic mean-field systems

fluctuations in quantities like $\frac{1}{N}\sum_{i=1}^{N}\xi_i\sigma_i$
or $\frac{1}{N}\sum_{i=1}^{N}\sigma_i\sigma_i'$ scale as $\mathcal{O}(N^{-1/2})$

hence

$$\lim_{N\to\infty}\langle P(m|\boldsymbol{\sigma})\rangle = \lim_{N\to\infty}\left\langle \delta\Big[m-\frac{1}{N}\sum_{i=1}^{N}\xi_i\sigma_i\Big]\right\rangle = \delta\Big[m-\frac{1}{N}\sum_{i=1}^{N}\xi_i\langle\sigma_i\rangle\Big]$$

$$\lim_{N\to\infty}\langle\langle P(q|\boldsymbol{\sigma},\boldsymbol{\sigma}')\rangle\rangle = \lim_{N\to\infty}\left\langle\left\langle \delta\Big[q-\frac{1}{N}\sum_{i=1}^{N}\sigma_i\sigma_i'\Big]\right\rangle\right\rangle = \delta\Big[q-\frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle^2\Big]$$

hence

$$\forall\alpha: \qquad m_\alpha = m = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\xi_i\overline{\langle\sigma_i\rangle}$$

$$\forall\alpha\neq\gamma: \qquad q_{\alpha\gamma} = q = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\overline{\langle\sigma_i\rangle^2}$$

replica-symmetric solution

Replica symmetric solution

$m_\alpha = m,\ q_{\alpha\neq\beta} = q,$ now find m and q ...

- **RS saddle-point eqns**
  insert RS form and use $\exp(\frac{1}{2}x^2) = \int \mathrm{D}z\ \mathrm{e}^{xz}$

$$m = \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \mathrm{e}^{\beta J_0 m \sum_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 q \sum_{\lambda\neq\zeta} \sigma_\lambda \sigma_\zeta}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\beta J_0 m \sum_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 q \sum_{\lambda\neq\zeta} \sigma_\lambda \sigma_\zeta}} = \frac{\sum_{\boldsymbol{\sigma}} \sigma_\alpha \mathrm{e}^{\beta J_0 m \sum_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 q [\sum_\lambda \sigma_\lambda]^2}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\beta J_0 m \sum_\lambda \sigma_\lambda + \frac{1}{2}(\beta J)^2 q [\sum_\lambda \sigma_\lambda]^2}}$$

$$= \frac{\int \mathrm{D}z \sum_{\boldsymbol{\sigma}} \sigma_\alpha \prod_{\lambda=1}^n \mathrm{e}^{\beta(J_0 m + Jz\sqrt{q})\sigma_\lambda}}{\int \mathrm{D}z \sum_{\boldsymbol{\sigma}} \prod_{\lambda=1}^n \mathrm{e}^{\beta(J_0 m + Jz\sqrt{q})\sigma_\lambda}}$$

$$= \frac{\int \mathrm{D}z\ \sinh[\beta(J_0 m + Jz\sqrt{q})] \cosh^{n-1}[\beta(J_0 m + Jz\sqrt{q})]}{\int \mathrm{D}z\ \cosh^n[\beta(J_0 m + Jz\sqrt{q})]}$$

similarly

$$q = \frac{\int \mathrm{D}z\ \sinh^2[\beta(J_0 m + Jz\sqrt{q})] \cosh^{n-2}[\beta(J_0 m + Jz\sqrt{q})]}{\int \mathrm{D}z\ \cosh^n[\beta(J_0 m + Jz\sqrt{q})]}$$

- **the limit** $n \to 0$

$$m = \int \mathrm{D}z\ \tanh[\beta(J_0 m + Jz\sqrt{q})], \qquad q = \int \mathrm{D}z\ \tanh^2[\beta(J_0 m + Jz\sqrt{q})]$$

RS equations for $m = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \xi_i \overline{\langle \sigma_i \rangle}$
and $q = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \overline{\langle \sigma_i \rangle^2}$

$$m = \int \mathrm{D}z \, \tanh[\beta(J_0 m + Jz\sqrt{q})], \qquad q = \int \mathrm{D}z \, \tanh^2[\beta(J_0 m + Jz\sqrt{q})]$$

- bifurcations away from $(m, q) = (0, 0)$:

$$\begin{aligned} m &= \int \mathrm{D}z \, \left[\beta J_0 m + \beta J z\sqrt{q} + \mathcal{O}(m, \sqrt{q})^3\right] = \beta J_0 m + \ldots \\ q &= \int \mathrm{D}z \, \left[\beta J_0 m + \beta J z\sqrt{q} + \mathcal{O}(m, \sqrt{q})^3\right]^2 = \int \mathrm{D}z \, (\beta J)^2 z^2 q + \ldots \\ &= (\beta J)^2 q + \ldots \end{aligned}$$

hence:
first continuous bifurcations away from $\mathbf{q} = \mathbf{m} = \mathbf{0}$,
as identified earlier, are the RS solutions

$$m = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \xi_i \overline{\langle \sigma_i \rangle}, \qquad q = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \overline{\langle \sigma_i \rangle^2}$$

$$m = \int \mathrm{D}z \, \tanh[\beta(J_0 m + Jz\sqrt{q})], \qquad q = \int \mathrm{D}z \, \tanh^2[\beta(J_0 m + Jz\sqrt{q})]$$

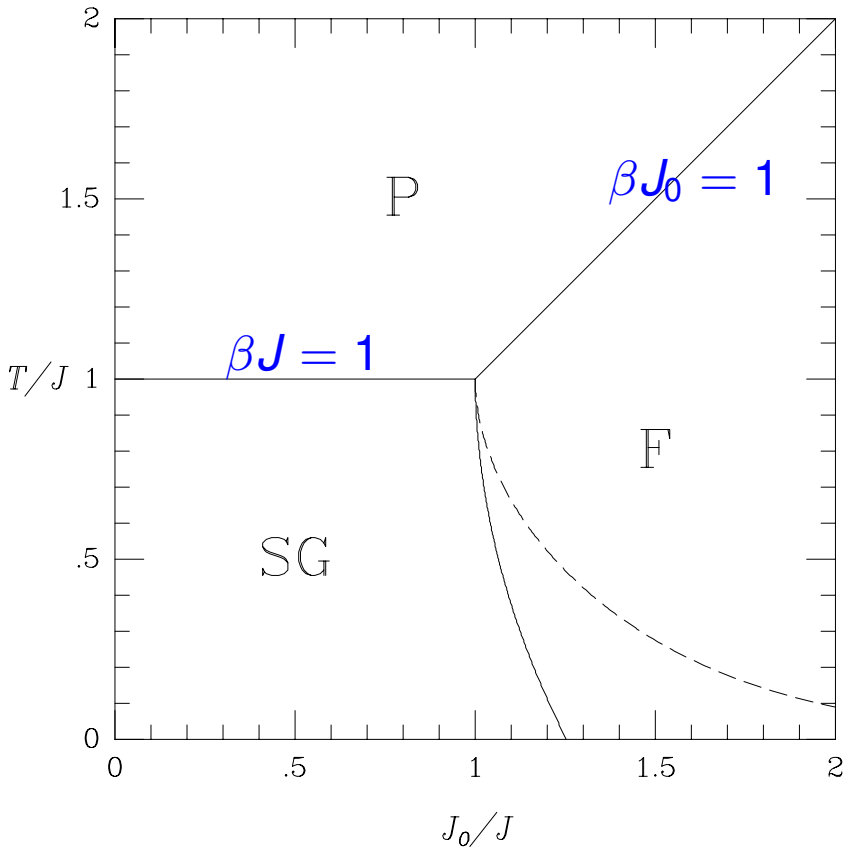
phase diagram

P: $m = q = 0$
random neuronal firing

SG: $m = 0$, $q > 0$
stable firing patterns, but
not related to stored pattern

F: $m, q > 0$
recall of stored information

$T = 1/\beta$ (noise strength)



1. The replica method
   - Exponential families and generating functions
   - The replica trick
   - The replica trick and algorithms
   - Alternative forms of the replica identity

2. Application: information storage in neural networks
   - Attractor neural networks
   - The replica calculation
   - Replica symmetry
   - Replica symmetric solution

3. Application: overfitting transition in linear separators
   - Linear separability of data – version space
   - The replica calculation
   - Gardner's replica symmetric theory

Linear separability of data and version space

Dimension mismatch and overfitting

two clinical outcomes (A,B),
4 patients, 60 expression levels ...

A : (100101001010010101010010001010111001001001001001001000011111)
A : (010001000010101001010101010010101000111100101001001010101000)
B : (001010001110101101100100100111001110010100101010101000101010)
B : (101011001010110010100100111100100101100111010111010001010010)

prognostic signature!

A : (100101001010010101010010010010101111001001001001001001000011111)
A : (010001000010101001010101010010101010001111001010010010101010000)
B : (001010001110101101100100100111001110010100101010101000101010)
B : (101011001010110010100100111100100101100111010111010001010010)

shuffle outcome labels ...

A : (100101001010010101010010001010111001001001001001001000011111)
B : (010001000010101001010101010010101010001111001010010010101010000)
A : (001010001110101101100100100111001110010100101010101000101010)
B : (101011001010110010100100111100100101100111010111010001010010)

overfitting, no reproducibility ...
how about overfitting in regression?

Suppose we have data D on N patients,
pairs of covariate vectors + clinical outcome labels

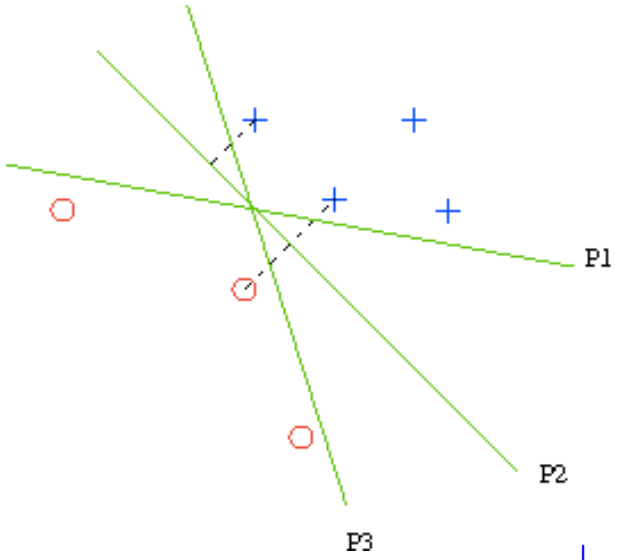
$$D = \{(\mathbf{x}^1, t^1), \ldots, (\mathbf{x}^N, t^N)\}, \qquad \mathbf{x}^i \in \{-1,1\}^p, \quad t^i \in \{-1,1\}, \qquad p, N \gg 1$$

e.g. $\mathbf{x}^i$= gene expressions of i (on/off)
t^i= treatment response of i (yes/no)

- assumed model:

$$t(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{\mu=1}^p \theta_\mu x_\mu > 0 \\ -1 & \text{if } \sum_{\mu=1}^p \theta_\mu x_\mu < 0 \end{cases}$$

$$= sgn\Big[\sum_{\mu=1}^p \theta_\mu x_\mu\Big]$$



$+$: $t^i = 1$
o : $t^i = -1$

- regression/classification task:
find parameters $\boldsymbol{\theta} = (\theta_1 \ldots, \theta_p)$
such that

$$\textit{for all } i = 1 \ldots N: \qquad t^i = sgn\Big[\sum_{\mu=1}^p \theta_\mu x_\mu^i\Big]$$

- data D explained perfectly by $\boldsymbol{\theta}$ if

$$\textit{for all } i = 1 \ldots N: \qquad t^i = sgn\left[\boldsymbol{\theta} \cdot \mathbf{x}^i\right], \quad \textit{i.e.} \quad t^i(\boldsymbol{\theta} \cdot \mathbf{x}^i) > 0$$

$$\textit{separating plane in input space}: \qquad \boldsymbol{\theta} \cdot \mathbf{x} = 0$$

$$\textit{distance } \Delta_i \textit{ between } \mathbf{x}^i \textit{ and separating plane}: \qquad d_i = t^i(\boldsymbol{\theta} \cdot \mathbf{x}^i)/|\boldsymbol{\theta}|$$

$$|\boldsymbol{\theta}| \textit{ irrelevant, so choose } |\boldsymbol{\theta}|^2 = p$$

- **version space**

  all $\boldsymbol{\theta}$ that solve above eqns
  with distances κ or larger

  volume of version space:

$$V(\kappa) = \int \mathrm{d}\boldsymbol{\theta}\ \delta(\boldsymbol{\theta}^2 - p) \prod_{i=1}^{N} \theta\left[\frac{t^i(\boldsymbol{\theta} \cdot \mathbf{x}^i)}{\sqrt{p}} > \kappa\right]$$

- high dimensional data: p large, $\alpha = N/p$
  $V(\kappa)$ scales exponentially with p, so

$$F \;=\; \frac{1}{p} \log V(\kappa) \;=\; \frac{1}{p} \log \int \mathrm{d}\boldsymbol{\theta}\ \delta(\boldsymbol{\theta}^2 - p) \prod_{i=1}^{N} \theta\left[\frac{t^i(\boldsymbol{\theta} \cdot \mathbf{x}^i)}{\sqrt{p}} > \kappa\right]$$

  $F = -\infty$: no solutions $\boldsymbol{\theta}$ exist, data D not linearly separable
  $F =$ finite: solutions $\boldsymbol{\theta}$ exist, data D linearly separable

overfitting: find parameters $\boldsymbol{\theta}$ that 'explain' random patterns
what if we choose **random data** D?

$$D = \{(\mathbf{x}^1, t^1), \ldots, (\mathbf{x}^N, t^N)\}, \qquad \mathbf{x}^i \in \{-1,1\}^p, \quad t^i \in \{-1,1\}, \quad \textit{fully random}$$

typical classification performance:

$$\overline{F} = \frac{1}{p}\overline{\log \int \mathrm{d}\boldsymbol{\theta}\, \delta(p-\boldsymbol{\theta}^2) \prod_{i=1}^{N} \theta\Big[\frac{t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)}{\sqrt{p}} > \kappa\Big]}$$

$$= \frac{1}{p}\overline{\log \int \frac{\mathrm{d}z}{2\pi} \mathrm{e}^{\mathrm{i}zp} \int \mathrm{d}\boldsymbol{\theta}\, \mathrm{e}^{-\mathrm{i}z\boldsymbol{\theta}^2} \prod_{i=1}^{N} \theta\Big[\frac{t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)}{\sqrt{p}} > \kappa\Big]}$$

transport data vars to harmless place,
using δ-functions, by inserting

$$1 = \int \mathrm{d}y_i\, \delta\Big[y_i - \frac{t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)}{\sqrt{p}}\Big] = \int \frac{\mathrm{d}y_i \mathrm{d}\hat{y}_i}{2\pi}\, \mathrm{e}^{\mathrm{i}\hat{y}_i y_i - \mathrm{i}\hat{y}_i t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)/\sqrt{p}}$$

gives

$$\overline{F} = \frac{1}{p}\overline{\log \int \frac{\mathrm{d}z\mathrm{d}\mathbf{y}\mathrm{d}\hat{\mathbf{y}}\mathrm{d}\boldsymbol{\theta}}{(2\pi)^{N+1}} \mathrm{e}^{\mathrm{i}zp + \mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y} - \mathrm{i}z\boldsymbol{\theta}^2} \Big(\prod_{i=1}^{N} \theta(y_i-\kappa)\mathrm{e}^{-\mathrm{i}\hat{y}_i t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)/\sqrt{p}}\Big)}$$

The replica calculation

large p, large N,
$N = \alpha p$:

$$\overline{F} = \lim_{p\to\infty} \frac{1}{p} \overline{\log \int \frac{\mathrm{d}z\mathrm{d}\mathbf{y}\mathrm{d}\hat{\mathbf{y}}\mathrm{d}\boldsymbol{\theta}}{(2\pi)^{N+1}} \mathrm{e}^{\mathrm{i}zp+\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}-\mathrm{i}z\boldsymbol{\theta}^2} \left(\prod_{i=1}^{N} \theta(y_i-\kappa)\mathrm{e}^{-\mathrm{i}\hat{y}_i t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)/\sqrt{p}} \right)}$$

- **replica identity**

  $\overline{\log Z} = \lim_{n\to 0} n^{-1} \log \overline{Z^n}$

$$\overline{F} = \lim_{p\to\infty} \lim_{n\to 0} \frac{1}{pn} \log \overline{\left[\int \frac{\mathrm{d}z\mathrm{d}\mathbf{y}\mathrm{d}\hat{\mathbf{y}}\mathrm{d}\boldsymbol{\theta}}{(2\pi)^{N+1}} \mathrm{e}^{\mathrm{i}zp+\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}-\mathrm{i}z\boldsymbol{\theta}^2} \left(\prod_{i=1}^{N} \theta(y_i-\kappa)\mathrm{e}^{-\mathrm{i}\hat{y}_i t^i(\boldsymbol{\theta}\cdot\mathbf{x}^i)/\sqrt{p}} \right) \right]^n}$$

$$= \lim_{p\to\infty} \lim_{n\to 0} \frac{1}{pn} \log \int \prod_{\alpha=1}^{n} \left[\frac{\mathrm{d}z^\alpha \mathrm{d}\mathbf{y}^\alpha \mathrm{d}\hat{\mathbf{y}}^\alpha \mathrm{d}\boldsymbol{\theta}^\alpha}{(2\pi)^{N+1}} \mathrm{e}^{\mathrm{i}pz^\alpha+\mathrm{i}\hat{\mathbf{y}}^\alpha\cdot\mathbf{y}^\alpha-\mathrm{i}z^\alpha(\boldsymbol{\theta}^\alpha)^2} \prod_{i=1}^{N} \theta[y_i^\alpha-\kappa] \right]$$

$$\times \overline{\mathrm{e}^{-\mathrm{i}\sum_{i=1}^{N}\sum_{\alpha=1}^{n} \hat{y}_i^\alpha t^i(\boldsymbol{\theta}^\alpha\cdot\mathbf{x}^i)/\sqrt{p}}}$$

- **average over data D:**

$$\Xi = \overline{e^{-i\sum_{i=1}^{N}\sum_{\alpha=1}^{n}\hat{y}_i^\alpha t^i(\boldsymbol{\theta}^\alpha\cdot\mathbf{x}^i)/\sqrt{p}}} = \overline{e^{-i\sum_{\mu=1}^{p}\sum_{i=1}^{N}t^i x_\mu^i\sum_{\alpha=1}^{n}\hat{y}_i^\alpha\theta_\mu^\alpha/\sqrt{p}}}$$

$$= \prod_{\mu=1}^{p}\prod_{i=1}^{N}\overline{e^{-it^i x_\mu^i\sum_{\alpha=1}^{n}\hat{y}_i^\alpha\theta_\mu^\alpha/\sqrt{p}}} = \prod_{\mu=1}^{p}\prod_{i=1}^{N}\cos\Big[\frac{1}{\sqrt{p}}\sum_{\alpha=1}^{n}\hat{y}_i^\alpha\theta_\mu^\alpha\Big]$$

$$= \prod_{\mu=1}^{p}\prod_{i=1}^{N}\Big\{1-\frac{1}{2p}\Big(\sum_{\alpha=1}^{n}\hat{y}_i^\alpha\theta_\mu^\alpha\Big)^2+\mathcal{O}(\frac{1}{p^2})\Big\} = e^{-\frac{1}{2p}\sum_{\mu=1}^{p}\sum_{i=1}^{N}\sum_{\alpha,\beta=1}^{n}\hat{y}_i^\alpha\hat{y}_i^\beta\theta_\mu^\alpha\theta_\mu^\beta+\mathcal{O}(p^0)}$$

- giving

$$\overline{F} = \lim_{p\to\infty}\lim_{n\to 0}\frac{1}{pn}\log\int\prod_{\alpha=1}^{n}\Big[\frac{dz^\alpha d\mathbf{y}d\hat{\mathbf{y}}^\alpha d\boldsymbol{\theta}^\alpha}{(2\pi)^{N+1}}e^{ipz^\alpha+i\hat{\mathbf{y}}^\alpha\cdot\mathbf{y}^\alpha-iz(\boldsymbol{\theta}^\alpha)^2}\prod_{i=1}^{N}\theta(y_i^\alpha-\kappa)\Big]$$

$$\times\, e^{-\frac{1}{2p}\sum_{\mu=1}^{p}\sum_{i=1}^{N}\sum_{\alpha,\beta=1}^{n}\hat{y}_i^\alpha\hat{y}_i^\beta\theta_\mu^\alpha\theta_\mu^\beta+\mathcal{O}(p^0)}$$

$$= -\alpha\log(2\pi)+\lim_{p\to\infty}\lim_{n\to 0}\frac{1}{pn}\log\int\prod_{\alpha=1}^{n}\Big(dz^\alpha d\boldsymbol{\theta}^\alpha e^{ipz^\alpha-iz^\alpha(\boldsymbol{\theta}^\alpha)^2}\Big)$$

$$\times\prod_{i=1}^{N}\int\prod_{\alpha=1}^{n}\Big[dy_i^\alpha d\hat{y}_i^\alpha e^{i\sum_\alpha\hat{y}_i^\alpha y_i^\alpha}\theta[y_i^\alpha-\kappa]\Big]e^{-\frac{1}{2}\sum_{\alpha,\beta}\hat{y}_i^\alpha\hat{y}_i^\beta[\frac{1}{p}\sum_{\mu=1}^{p}\theta_\mu^\alpha\theta_\mu^\beta]}$$

- so, with $\mathbf{y} = (y_1, \ldots, y_n)$, $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_n)$, $\mathbf{z} = (z_1, \ldots, z_n)$:

$$\begin{aligned}\overline{F} &= -\alpha \log(2\pi) + \lim_{p\to\infty} \lim_{n\to 0} \frac{1}{pn} \log \int \mathrm{d}\mathbf{z} \Big(\prod_{\alpha=1}^{n} \mathrm{d}\boldsymbol{\theta}^{\alpha} \mathrm{e}^{\mathrm{i}pz^{\alpha} - \mathrm{i}z^{\alpha}(\boldsymbol{\theta}^{\alpha})^2} \Big) \\ &\quad \times \Big\{ \int \mathrm{d}\mathbf{y} \mathrm{d}\hat{\mathbf{y}}\ \mathrm{e}^{\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}} \prod_{\alpha=1}^{n} \theta[y^{\alpha} - \kappa]\ \mathrm{e}^{-\frac{1}{2}\sum_{\alpha,\beta} \hat{y}^{\alpha}\hat{y}^{\beta}[\frac{1}{p}\sum_{\mu=1}^{p} \theta_{\mu}^{\alpha}\theta_{\mu}^{\beta}]} \Big\}^{N}\end{aligned}$$

- insert

$$1 = \int \mathrm{d}q_{\alpha\beta}\ \delta\Big[q_{\alpha\beta} - \frac{1}{p}\sum_{\mu=1}^{p} \theta_{\mu}^{\alpha}\theta_{\mu}^{\beta}\Big] = \int \frac{\mathrm{d}q_{\alpha\beta}\mathrm{d}\hat{q}_{\alpha\beta}}{2\pi/p}\ \mathrm{e}^{\mathrm{i}p\hat{q}_{\alpha\beta}\left[q_{\alpha\beta} - \frac{1}{p}\sum_{\mu=1}^{p} \theta_{\mu}^{\alpha}\theta_{\mu}^{\beta}\right]}$$

to get

$$\begin{aligned}\overline{F} &= -\alpha \log(2\pi) + \lim_{p\to\infty} \lim_{n\to 0} \frac{1}{pn} \log \int \mathrm{d}\mathbf{z}\mathrm{d}\mathbf{q}\mathrm{d}\hat{\mathbf{q}}\ \mathrm{e}^{\mathrm{i}p\sum_{\alpha\beta=1}^{n} \hat{q}_{\alpha\beta}q_{\alpha\beta} + \mathrm{i}p\sum_{\alpha=1}^{n} z_{\alpha}} \\ &\quad \times \Big\{ \int \mathrm{d}\mathbf{y} \mathrm{d}\hat{\mathbf{y}}\ \mathrm{e}^{\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}} \prod_{\alpha=1}^{n} \theta[y^{\alpha} - \kappa]\ \mathrm{e}^{-\frac{1}{2}\hat{\mathbf{y}}\cdot\mathbf{q}\hat{\mathbf{y}}} \Big\}^{N} \int \prod_{\alpha=1}^{n} \Big(\mathrm{d}\boldsymbol{\theta}^{\alpha} \mathrm{e}^{-\mathrm{i}z^{\alpha}(\boldsymbol{\theta}^{\alpha})^2} \Big) \mathrm{e}^{-\mathrm{i}\sum_{\mu=1}^{p} \sum_{\alpha\beta} \hat{q}_{\alpha\beta}\theta_{\mu}^{\alpha}\theta_{\mu}^{\beta}}\end{aligned}$$

- so, with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$:
  (remember: $N = \alpha p$)

$$\begin{aligned}
\overline{F} &= -\alpha \log(2\pi) + \lim_{p\to\infty} \lim_{n\to 0} \frac{1}{pn} \log \int \mathrm{d}\mathbf{z}\mathrm{d}\mathbf{q}\mathrm{d}\hat{\mathbf{q}}\; \mathrm{e}^{\mathrm{i}p \sum_{\alpha\beta=1}^{n} \hat{q}_{\alpha\beta} q_{\alpha\beta} + \mathrm{i}p \sum_{\alpha=1}^{n} z_\alpha} \\
&\quad \times \Big\{ \int \mathrm{d}\mathbf{y}\mathrm{d}\hat{\mathbf{y}}\; \mathrm{e}^{\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}} \prod_{\alpha=1}^{n} \theta[y^\alpha - \kappa]\; \mathrm{e}^{-\frac{1}{2}\hat{\mathbf{y}}\cdot\mathbf{q}\hat{\mathbf{y}}} \Big\}^{\alpha p} \Big\{ \int \mathrm{d}\boldsymbol{\theta}\; \mathrm{e}^{-\mathrm{i}\sum_{\alpha=1}^{n} z^\alpha \theta_\alpha^2 - \mathrm{i}\boldsymbol{\theta}\cdot\hat{\mathbf{q}}\boldsymbol{\theta}} \Big\}^{p} \\
&= \lim_{p\to\infty} \lim_{n\to 0} \frac{1}{pn} \log \int \mathrm{d}\mathbf{z}\mathrm{d}\mathbf{q}\mathrm{d}\hat{\mathbf{q}}\; \mathrm{e}^{p\Psi(\mathbf{z},\mathbf{q},\hat{\mathbf{q}})}
\end{aligned}$$

$$\begin{aligned}
\Psi(\ldots) &= \mathrm{i} \sum_{\alpha\beta=1}^{n} \hat{q}_{\alpha\beta} q_{\alpha\beta} + \mathrm{i} \sum_{\alpha=1}^{n} z_\alpha + \log \int \mathrm{d}\boldsymbol{\theta}\; \mathrm{e}^{-\mathrm{i}\sum_{\alpha=1}^{n} z^\alpha \theta_\alpha^2 - \mathrm{i}\boldsymbol{\theta}\cdot\hat{\mathbf{q}}\boldsymbol{\theta}} \\
&\quad + \alpha \log \int \mathrm{d}\mathbf{y}\mathrm{d}\hat{\mathbf{y}}\; \mathrm{e}^{\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}} \prod_{\alpha=1}^{n} \theta[y^\alpha - \kappa]\; \mathrm{e}^{-\frac{1}{2}\hat{\mathbf{y}}\cdot\mathbf{q}\hat{\mathbf{y}}} - \alpha n \log(2\pi)
\end{aligned}$$

- assume limits $n \to 0$ and $p \to \infty$ commute,
  steepest descent integration

$$\overline{F} = \lim_{n\to 0} \frac{1}{n} \mathrm{extr}_{\mathbf{z},\mathbf{q},\hat{\mathbf{q}}} \Psi(\mathbf{z}, \mathbf{q}, \hat{\mathbf{q}})$$

$$\Psi(\mathbf{z},\mathbf{q},\hat{\mathbf{q}}) = \mathrm{i}\sum_{\alpha\beta=1}^{n}\hat{q}_{\alpha\beta}q_{\alpha\beta} + \mathrm{i}\sum_{\alpha=1}^{n} z_\alpha + \log\int \mathrm{d}\boldsymbol{\theta}\, \mathrm{e}^{-\mathrm{i}\sum_{\alpha=1}^{n} z^\alpha\theta_\alpha^2 - \mathrm{i}\boldsymbol{\theta}\cdot\hat{\mathbf{q}}\boldsymbol{\theta}}$$
$$+\alpha\log\int \mathrm{d}\mathbf{y}\mathrm{d}\hat{\mathbf{y}}\, \mathrm{e}^{\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}}\prod_{\alpha=1}^{n}\theta[y^\alpha-\kappa]\,\mathrm{e}^{-\frac{1}{2}\hat{\mathbf{y}}\cdot\mathbf{q}\hat{\mathbf{y}}} - \alpha n\log(2\pi)$$

- transform $\hat{q}_{\alpha\beta} = -\frac{1}{2}\mathrm{i}k_{\alpha\beta} - z_\alpha\delta_{\alpha\beta}$,
  and integrate over $\hat{\mathbf{y}}$:

$$\Psi(\mathbf{z},\mathbf{q},\mathbf{k}) = \frac{1}{2}\sum_{\alpha\beta=1}^{n}k_{\alpha\beta}q_{\alpha\beta} + \mathrm{i}\sum_{\alpha=1}^{n} z_\alpha(1-q_{\alpha\alpha}) + \log\int \mathrm{d}\boldsymbol{\theta}\,\mathrm{e}^{-\frac{1}{2}\boldsymbol{\theta}\cdot\mathbf{k}\boldsymbol{\theta}}$$
$$+\alpha\log\int \mathrm{d}\mathbf{y}\prod_{\alpha=1}^{n}\theta[y^\alpha-\kappa]\int \mathrm{d}\hat{\mathbf{y}}\,\mathrm{e}^{\mathrm{i}\hat{\mathbf{y}}\cdot\mathbf{y}-\frac{1}{2}\hat{\mathbf{y}}\cdot\mathbf{q}\hat{\mathbf{y}}} - \alpha n\log(2\pi)$$
$$= \frac{1}{2}\sum_{\alpha\beta=1}^{n}k_{\alpha\beta}q_{\alpha\beta} + \mathrm{i}\sum_{\alpha=1}^{n} z_\alpha(1-q_{\alpha\alpha}) + \log\frac{(2\pi)^{n/2}}{\sqrt{\mathrm{Det}\mathbf{k}}}$$
$$+\alpha\log\int \mathrm{d}\mathbf{y}\prod_{\alpha=1}^{n}\theta[y^\alpha-\kappa]\,\frac{(2\pi)^{n/2}}{\sqrt{\mathrm{Det}\mathbf{q}}}\mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{q}^{-1}\mathbf{y}} - \alpha n\log(2\pi)$$

- re-organise:

$$\Psi(\mathbf{z},\mathbf{q},\mathbf{k}) = \frac{1}{2}\sum_{\alpha\beta=1}^{n} k_{\alpha\beta}q_{\alpha\beta} + \mathrm{i}\sum_{\alpha=1}^{n} z_\alpha(1-q_{\alpha\alpha}) - \frac{1}{2}\log\mathrm{Det}\,\mathbf{k} - \frac{1}{2}\alpha\log\mathrm{Det}\,\mathbf{q}$$
$$+\alpha\log\int \mathrm{d}\mathbf{y}\ \prod_{\alpha=1}^{n}\theta[y^\alpha-\kappa]\mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{q}^{-1}\mathbf{y}} + \frac{1}{2}n(1-\alpha)\log(2\pi)$$

- extremise with respect to $\mathbf{z}$:
  $\partial\Psi/\partial z_\alpha = 0: \quad q_{\alpha\alpha} = 0$ for all α

$$\Psi(\mathbf{q},\mathbf{k}) = \frac{1}{2}n(1-\alpha)\log(2\pi) + \frac{1}{2}\sum_{\alpha\beta=1}^{n} k_{\alpha\beta}q_{\alpha\beta} - \frac{1}{2}\log\mathrm{Det}\,\mathbf{k} - \frac{1}{2}\alpha\log\mathrm{Det}\,\mathbf{q}$$
$$+\alpha\log\int \mathrm{d}\mathbf{y}\ \prod_{\alpha=1}^{n}\theta[y^\alpha-\kappa]\mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{q}^{-1}\mathbf{y}}$$

next: ergodicity assumption,
replica-symmetric form for $\mathbf{q}$ and $\mathbf{k}$ ...

1. The replica method
   - Exponential families and generating functions
   - The replica trick
   - The replica trick and algorithms
   - Alternative forms of the replica identity

2. Application: information storage in neural networks
   - Attractor neural networks
   - The replica calculation
   - Replica symmetry
   - Replica symmetric solution

3. Application: overfitting transition in linear separators
   - Linear separability of data – version space
   - The replica calculation
   - Gardner's replica symmetric theory

Gardner's replica symmetric theory

$$\Psi(\mathbf{q},\mathbf{k}) = \frac{1}{2}n(1-\alpha)\log(2\pi) + \frac{1}{2}\sum_{\alpha\beta=1}^{n} k_{\alpha\beta}q_{\alpha\beta} - \frac{1}{2}\log \mathrm{Det}\,\mathbf{k} - \frac{1}{2}\alpha \log \mathrm{Det}\,\mathbf{q}$$
$$+\alpha \log \int \mathrm{d}\mathbf{y} \prod_{\alpha=1}^{n} \theta[y^{\alpha}-\kappa]\mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{q}^{-1}\mathbf{y}}$$

RS saddle-points

$$q_{\alpha\beta} = \delta_{\alpha\beta} + (1-\delta_{\alpha\beta})q, \qquad k_{\alpha\beta} = K\delta_{\alpha\beta} + (1-\delta_{\alpha\beta})k$$

- eigenvalues:

$$\mathbf{x} = (1,\ldots,1): \qquad (\mathbf{kx})_{\alpha} = \sum_{\beta=1}^{n}[k+(K-k)\delta_{\alpha\beta}]x_{\beta} = nk+K-k$$
$$\textit{eigenvalue}: \quad \lambda = nk+K-k$$
$$\sum_{\alpha=1}^{n} x_{\alpha} = 0: \qquad (\mathbf{kx})_{\alpha} = \sum_{\beta=1}^{n}[k+(K-k)\delta_{\alpha\beta}]x_{\beta} = (K-k)x_{\alpha}$$
$$\textit{eigenvalue}: \quad \lambda = K-k \quad (n-1\ \textit{fold})$$

hence

$$\mathrm{Det}\,\mathbf{k} = (nk+K-k)(K-k)^{n-1}, \qquad \mathrm{Det}\,\mathbf{q} = (nq+1-q)(1-q)^{n-1}$$

- invert $\mathbf{q}$, try $(\mathbf{q}^{-1})_{\alpha\beta} = r + (R-r)\delta_{\alpha\beta}$, demand:

$$\begin{aligned}\delta_{\alpha\beta} &= (\mathbf{q}\mathbf{q}^{-1})_{\alpha\beta} = \sum_{\gamma}(q+(1-q)\delta_{\alpha\gamma})(r+(R-r)\delta_{\gamma\beta}) \\ &= nqr + q(R-r) + r(1-q) + (R-r)(1-q)\delta_{\alpha\beta}\end{aligned}$$

so $$nqr + q(R-r) + r(1-q) = 0, \qquad (R-r)(1-q) = 1$$

$$R = r + \frac{1}{1-q}, \qquad r = -\frac{q}{(1-q)(1-q+nq)}$$

- hence, using $\exp[\frac{1}{2}x^2] = \int \mathrm{D}z\, \mathrm{e}^{xz}$

$$\log \int \mathrm{d}\mathbf{y} \prod_{\alpha=1}^{n} \theta[y^\alpha - \kappa] \mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{q}^{-1}\mathbf{y}} = \log \int \mathrm{d}\mathbf{y} \prod_{\alpha=1}^{n} \theta[y^\alpha - \kappa] \mathrm{e}^{-\frac{1}{2}\sum_{\alpha\beta} y_\alpha [r+(R-r)\delta_{\alpha\beta}] y_\beta}$$

$$\begin{aligned}&= \log \int \mathrm{d}\mathbf{y} \prod_{\alpha=1}^{n} \theta[y^\alpha - \kappa] \mathrm{e}^{-\frac{1}{2}r[\sum_{\alpha\beta} y_\alpha]^2 - \frac{1}{2}(R-r)\sum_\alpha y_\alpha^2} \\ &= \log \int \mathrm{D}z \int \mathrm{d}\mathbf{y} \prod_{\alpha=1}^{n} \theta[y^\alpha - \kappa] \mathrm{e}^{z\sqrt{-r}\sum_{\alpha\beta} y_\alpha - \frac{1}{2}(R-r)\sum_\alpha y_\alpha^2} \\ &= \log \int \mathrm{D}z \Big[\int_\kappa^\infty \mathrm{d}y\, \mathrm{e}^{z\sqrt{-r}y - \frac{1}{2}(R-r)y^2} \Big]^n\end{aligned}$$

put everything together ...

$$
\begin{aligned}
\frac{1}{n}\Psi(\mathbf{q},\mathbf{k}) &= \frac{1}{2}(1-\alpha)\log(2\pi)+\frac{1}{2}K+\frac{1}{2}(n-1)qk-\frac{1}{2n}\log[(nk+K-k)(K-k)^{n-1}] \\
&\quad -\frac{\alpha}{2n}\log[(nq+1-q)(1-q)^{n-1}]+\frac{\alpha}{n}\log\int \mathrm{D}z\Big[\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{z\sqrt{-r}y-\frac{1}{2}(R-r)y^2}\Big]^n \\
&= \frac{1}{2}(1-\alpha)\log(2\pi)+\frac{1}{2}(K-qk)-\frac{1}{2n}\log(1+\frac{nk}{K-k})-\frac{1}{2}\log(K-k) \\
&\quad -\frac{\alpha}{2n}\log(1+\frac{nq}{1-q})-\frac{\alpha}{2}\log(1-q)+\mathcal{O}(n) \\
&\quad +\frac{\alpha}{n}\log\int \mathrm{D}z\Big[1+n\log\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{zy\sqrt{q}/(1-q)-\frac{1}{2(1-q)}y^2}+\mathcal{O}(n^2)\Big]
\end{aligned}
$$

take limit $n \to 0$:

$$
\begin{aligned}
2\overline{F} &= (1-\alpha)\log(2\pi)+\mathrm{extr}_{K,k,q}\Big\{K-qk-\frac{k}{K-k}-\log(K-k) \\
&\quad -\frac{\alpha q}{1-q}-\alpha\log(1-q)+2\alpha\int \mathrm{D}z\ \log\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{zy\sqrt{q}/(1-q)-\frac{1}{2(1-q)}y^2}\Big\}
\end{aligned}
$$

$$2\overline{F} = (1-\alpha)\log(2\pi) + \mathrm{extr}_{K,k,q}\Big\{K - qk - \frac{k}{K-k} - \log(K-k) - \frac{\alpha q}{1-q} - \alpha\log(1-q) + 2\alpha\int \mathrm{D}z\ \log\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{zy\sqrt{q}/(1-q)-\frac{1}{2(1-q)}y^2}\Big\}$$

- **extremise over K and k**

$$\begin{cases} \frac{\partial}{\partial K} = 0: & 1 + \frac{k}{(K-k)^2} - \frac{1}{K-k} = 0 \\ \frac{\partial}{\partial k} = 0: & -q - \frac{1}{K-k} - \frac{k}{(K-k)^2} + \frac{1}{K-k} = 0 \end{cases} \Rightarrow \quad K = \frac{1-2q}{(1-q)^2}, \quad k = -\frac{q}{(1-q)^2}$$

result:
$$2\overline{F} = (1-\alpha)\log(2\pi) + \mathrm{extr}_{q}\Big\{\frac{1}{1-q} - \frac{\alpha q}{1-q} + (1-\alpha)\log(1-q) + 2\alpha\int \mathrm{D}z\ \log\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{zy\sqrt{q}/(1-q)-\frac{1}{2(1-q)}y^2}\Big\}$$

- write y-integral in terms of error function $\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{d}t\ \mathrm{e}^{-t^2}$:

$$\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{zy\sqrt{q}/(1-q)-\frac{1}{2(1-q)}y^2} = \mathrm{e}^{\frac{qz^2}{2(1-q)}}\int_{\kappa}^{\infty}\mathrm{d}y\ \mathrm{e}^{-\frac{[y-z\sqrt{q}]^2}{2(1-q)}} = \sqrt{2(1-q)}\ \mathrm{e}^{\frac{qz^2}{2(1-q)}}\frac{\sqrt{\pi}}{2}\Big\{1-\mathrm{Erf}\Big[\frac{K-z\sqrt{q}}{\sqrt{2(1-q)}}\Big]\Big\}$$

- insert previous integral:

$$2\overline{F} \;=\; \log\pi + (1-2\alpha)\log 2 + \mathrm{extr}_q\Big\{\frac{1}{1-q} + \log(1-q) + 2\alpha\int \mathrm{D}z\ \log\Big[1-\mathrm{Erf}\Big(\frac{K-z\sqrt{q}}{\sqrt{2(1-q)}}\Big)\Big]\Big\}$$

- **extremisation with respect to q**

short-hand $u(z,q) = (\kappa - z\sqrt{q})/\sqrt{2(1-q)}$,
use $\mathrm{Erf}'(x) = \frac{2}{\sqrt{\pi}}\exp[-x^2]$

$$\frac{\mathrm{d}}{\mathrm{d}q} = 0: \qquad \frac{1}{(1-q)^2} - \frac{1}{1-q} - 2\alpha\int \mathrm{D}z\,\Big(\frac{\partial u}{\partial q}\Big)\frac{\mathrm{Erf}'\ u(z,q)}{1-\mathrm{Erf}\ u(z,q)} = 0$$

$$\frac{q}{(1-q)^2} = \frac{4\alpha}{\sqrt{\pi}}\int \mathrm{D}z\,\Big(\frac{\partial u}{\partial q}\Big)\frac{\mathrm{e}^{-u^2(z,q)}}{1-\mathrm{Erf}\ u(z,q)}$$

work out: $\dfrac{\partial u}{\partial q} = \dfrac{1}{\sqrt{2}}\dfrac{\partial}{\partial q}\dfrac{\kappa - z\sqrt{q}}{(1-q)^{1/2}} = \ldots = \dfrac{\kappa\sqrt{q}-z}{2\sqrt{2q}(1-q)^{3/2}}$

insert into eqn for q:

$$q\sqrt{q} = \alpha\sqrt{\frac{2}{\pi}}\sqrt{1-q}\int \mathrm{D}z\ \frac{\mathrm{e}^{-u^2(z,q)}(\kappa\sqrt{q}-z)}{1-\mathrm{Erf}\ u(z,q)}$$

$$2\overline{F} = \log\pi + (1-2\alpha)\log 2 + \frac{1}{1-q} + \log(1-q) + 2\alpha \int \mathrm{D}z \, \log\Big[1-\mathrm{Erf}\, u(z,q)\Big]$$

$$q\sqrt{q} = \alpha\sqrt{\frac{2}{\pi}}\sqrt{1-q}\int \mathrm{D}z \, \frac{\mathrm{e}^{-u^2(z,q)}(\kappa\sqrt{q}-z)}{1-\mathrm{Erf}\, u(z,q)}, \qquad u(z,q) = \frac{\kappa - z\sqrt{q}}{\sqrt{2(1-q)}}$$

remember:

$\overline{F}$ =finite: random data linearly separable with margin κ
$\overline{F} = -\infty$: random data not linearly separable with margin κ

- $\alpha = 0$ (so $1 \ll N \ll p$):

  $q = 0, \quad 2\overline{F} = \log\pi + \log 2 + 1$ *random data linearly separable* (*overfitting*)

- $\alpha > 0$ (so $1 \ll N \sim p$):

  transition point: value of α where $q \to 1$

$$1 = \alpha_c(\kappa)\sqrt{\frac{2}{\pi}}\int \mathrm{D}z \, \lim_{q\to 1}\sqrt{1-q}\, \frac{\mathrm{e}^{-\left[\frac{K-z}{\sqrt{2(1-q)}}\right]^2}(\kappa - z)}{1-\mathrm{Erf}\left[\frac{\kappa - z}{\sqrt{2(1-q)}}\right]}$$

$$\alpha_c(\kappa) = \Big[\frac{1}{\sqrt{\pi}}\int \mathrm{D}z \, (\kappa+z) \lim_{\gamma\to\infty}\frac{1}{\gamma}\frac{\mathrm{e}^{-\gamma^2(\kappa+z)^2}}{1-\mathrm{Erf}[\gamma(\kappa+z)]}\Big]^{-1}$$

- remaining limit:

$$\lim_{\gamma\to\infty}\frac{1}{\gamma}\frac{e^{-\gamma^2Q^2}}{1-\mathrm{Erf}[\gamma Q]} = Q\sqrt{\pi}\,\theta(Q)$$

proof:

$$Q<0: \qquad \mathrm{Erf}[\gamma Q]\to -1 \quad so \quad \lim_{\gamma\to\infty}\frac{1}{\gamma}\frac{e^{-\gamma^2Q^2}}{1-\mathrm{Erf}[\gamma Q]} = 0$$

$$Q>0: \qquad \mathrm{Erf}[\gamma Q] = 1-\frac{1}{\gamma Q\sqrt{\pi}}e^{-\gamma^2Q^2}\left(1+\mathcal{O}(\frac{1}{\gamma^2Q^2})\right)$$

$$\lim_{\gamma\to\infty}\frac{1}{\gamma}\frac{e^{-\gamma^2Q^2}}{1-\mathrm{Erf}[\gamma Q]} = \lim_{\gamma\to\infty}\frac{1}{\gamma}\frac{e^{-\gamma^2Q^2}}{\frac{1}{\gamma Q\sqrt{\pi}}e^{-\gamma^2Q^2}\left(1+\mathcal{O}(\frac{1}{\gamma^2Q^2})\right)} = Q\sqrt{\pi}$$

final result

$$\alpha_c(\kappa) = \Big[\int_{-\kappa}^{\infty} \mathrm{D}z \, (\kappa+z)^2 \Big]^{-1} \qquad \alpha_c(0) = \Big[\int_{0}^{\infty} \mathrm{D}z \, z^2 \Big]^{-1} = [\frac{1}{2}]^{-1} = 2$$

p covariates,
N patients,
binary outcomes,
p and *N* large

random data
(i.e. pure binary noise)
is *perfectly* separable if
$N/p < \alpha_c(\kappa)$

algorithms (SVM etc)
will find pars $\theta_1 \ldots \theta_p$
such that $t_i = sgn[\sum_{\mu=1}^{p} \theta_\mu x_\mu^i]$
for all $i = 1 \ldots N$