Statistical Physics Approaches to High-Dimensional Inference

applications to biological data

Rémi Monasson

Laboratory of Theoretical Physics,
CNRS & Ecole Normale Supérieure, Paris

Winter School on Quantitative Biology, ICTP, December 2012

- Yesterday: Theoretical framework for model inference
  (special case: many interacting & stationary variables)
  Mean-field inference
  Applications covariation in protein families (I)

- Today: Issues & advanced statistical physics methods
  Inverse Hopfield-Potts model & Random Matrix Theory
  Applications to covariation in proteins (II)
  to neural data (I)

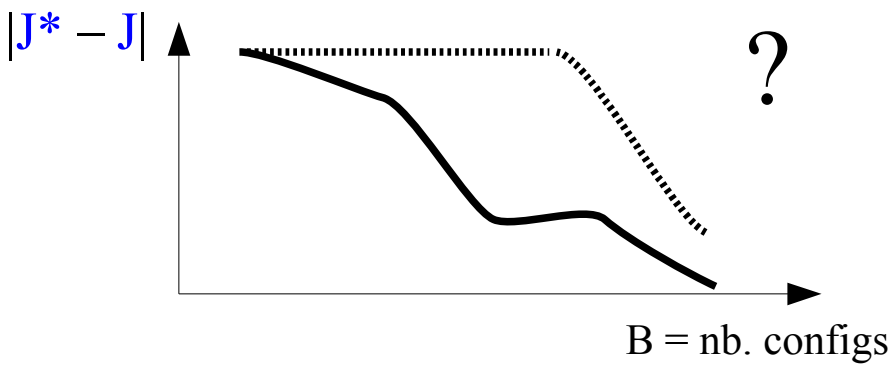
- Tomorrow: Case of interacting & non-stationary variables
  Applications to neural data (II)
  to ecological systems

Questions

1. Practical methods to find interactions J_{ij} from the correlations c_{ij}?
   (fast, accurate algorithms)

2. How much data does one need to get reliable interactions?
   (overfitting ...)

$J^* \rightarrow$ configurations $\{S_i\} \rightarrow c \rightarrow J$



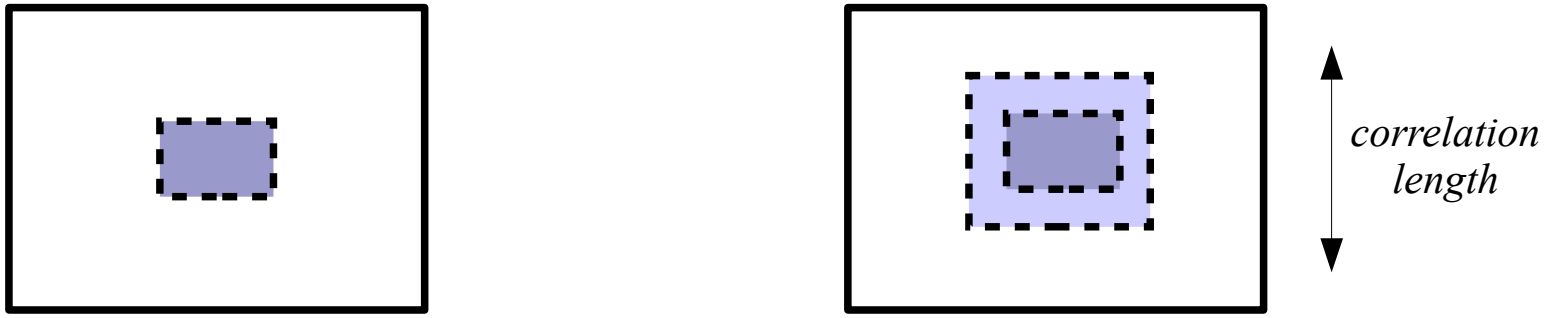
Questions

Asymptotic inference : B $\rightarrow$ infinity, while N is kept fixed

Error on each parameter of the order of $B^{-1/2}$

What happens in practice, i.e. when B and N are of the same order of magnitude ?

3. How large should be the sampled sub-system?



Is the inverse problem well-behaved ?

Analytical approaches

- Mean field inference
- Importance of prior(s)
- Pseudo-likelihood algorithms

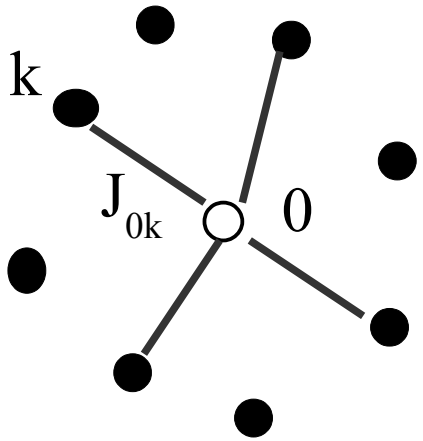
- Advanced statistical physics methods
- Inverse Hopfield model

Pseudo-likelihood methods (1)

Idea: avoid calculation of partition function using Callen identities (1963)

$$< \sigma_0 > \; = \; < \tanh(\sum_{k \neq 0} J_{0k} \, \sigma_k + h_0) >$$

$$\approx \sum_{\substack{\text{sampled} \\ \text{configurations}}} \tanh(\sum_{k \neq 0} J_{0k} \, \sigma_k^b + h_0) \; / \; B$$



Pseudo cross entropy:

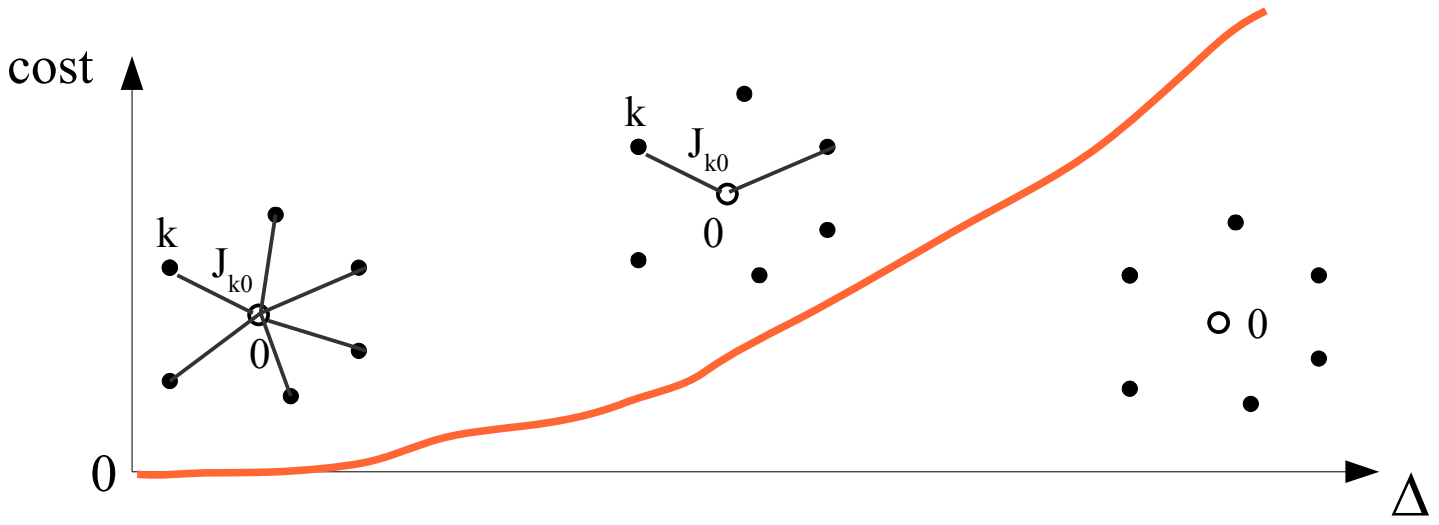
$$\text{« S »} \; = \sum_{\substack{\text{sampled} \\ \text{configurations}}} \log 2 \cosh(\sum_{k \neq 0} J_{0k} \, \sigma_k^b + h_0) \; - B \, (h_0 m_0 + \sum_{k \neq 0} J_{0k} \, c_{0k})$$

Ravikumar, Wainwright, Lafferty (2010)

Prior: increase signal/noise ratio by exploiting the sparsity of J_{ij}

$$\text{cost function } (\{J\}) = \text{pseudo-cross entropy } (h_0, \{J_{0k}\}) \; + \; \Delta \; \sum_k |J_{0k}|$$

Pseudo-likelihood methods (2)



Complexity: if B > a logN, the procedure finds couplings of amplitude $|J| > a' (\log N / B)^{1/2}$ in time poly(B,N)
[a & a' depend on the maximal degree (number of neighbours of i with $J_{ij} \neq 0$)]

Caveat:

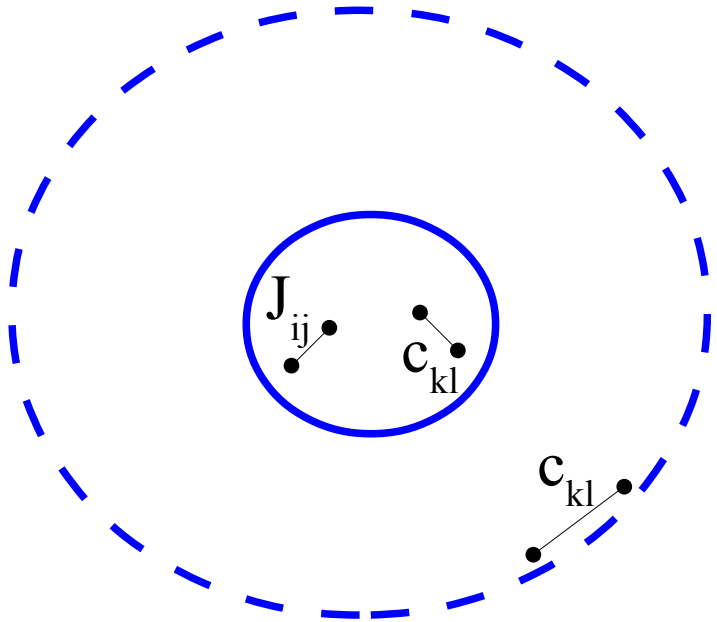
- Ising model should be the true model for data
- Coupling matrix is not symmetric (but is asymptotically consistent)
- Susceptibility χ should be small (fails in the vicinity of the critical point)
- All poly algorithms fail at critical point ?

Bento, Montanari '09

What is the relevant susceptibility for the inverse problem?

Susceptibility: $\chi_{ij,kl} = \left. \frac{\partial \langle s_i s_j \rangle}{\partial J_{kl}} \right|_{J}$ = response of a correlation to a small change in an interaction
may be long range ...

Inverse Susceptibility: $\chi^{-1}_{ij,kl} = \left. \frac{\partial J_{ij}}{\partial \langle s_k s_l \rangle} \right|_{\langle ss \rangle}$ = response of an interaction to a small change in a correlation



Inverse problem is well-behaved if χ^{-1} short range !

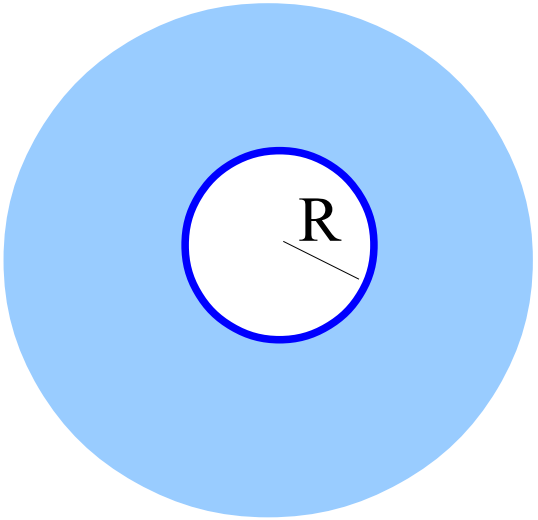
Examples of inverse susceptibility matrices

- Spherical (Gaussian) models : $\chi^{-1}_{ij,kl} = J_{ik}J_{jl}$ has the same sparsity as J

- Liquid theory: χ^{-1} is closely related to the Ornstein-Zernike direct correlation function. The short-rangedness of χ^{-1} is used for closure scheme e.g. Percus-Yevick

- Critical point of ferromagnet: $\chi^{-1}(q) \sim q^{2-\eta}$

$$\int_{r>R} dr\, \chi^{-1}(r) \sim R^{-(3-\eta)}$$



- 1D Ising model : 4-point χ^{-1} is sparse!

- Real data: see tomorrow

Analytical approaches

- Mean field inference
- Importance of prior(s)
- Pseudo-likelihood algorithms
- Advanced statistical physics methods

- Inverse Hopfield-Potts model

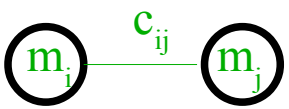
Adaptive cluster expansion (1)

$$S = \min_{J,h} - \sum_{\text{sampled configurations}} \log P(s_1,s_2,\ldots,s_N|\mathbf{J},\mathbf{h})$$

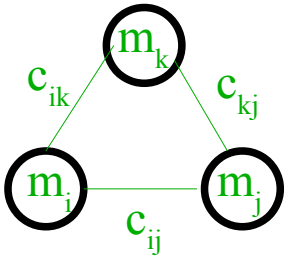
Cross-entropy

$$= \sum_i \Delta S_\iota(m_i) + \sum_{i<j} \Delta S_{\iota\varphi}(m_i,m_j,c_{ij}) + \sum_{i<j<k} \Delta S_{\iota\varphi\kappa}(m_i,m_j,m_k,c_{ij},c_{ik},c_{jk}) + \ldots$$

$\Rightarrow$ h_i



$\Rightarrow$ δh_i , δh_i , J_{ij}

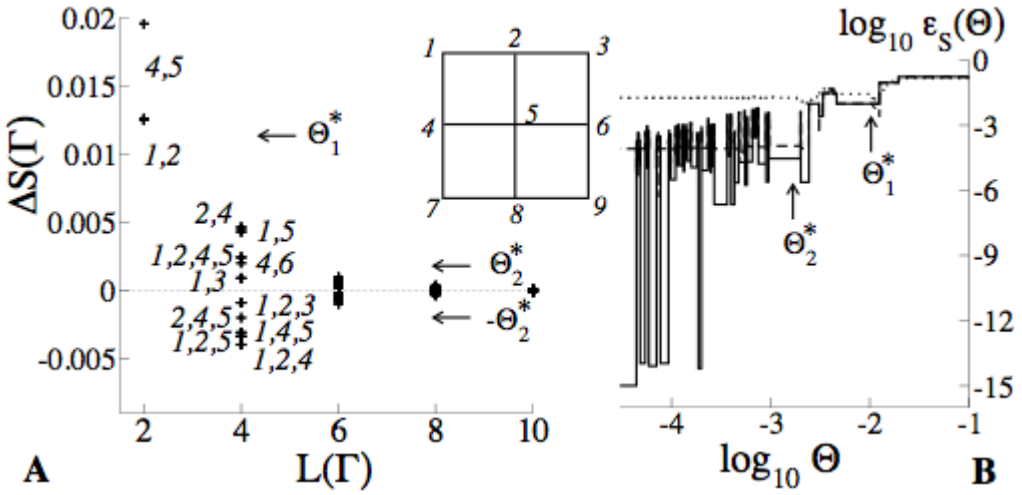


$\Rightarrow$ $\delta\delta h_i$, $\delta\delta h_i$, δJ_{ij} *(and other variables...)*

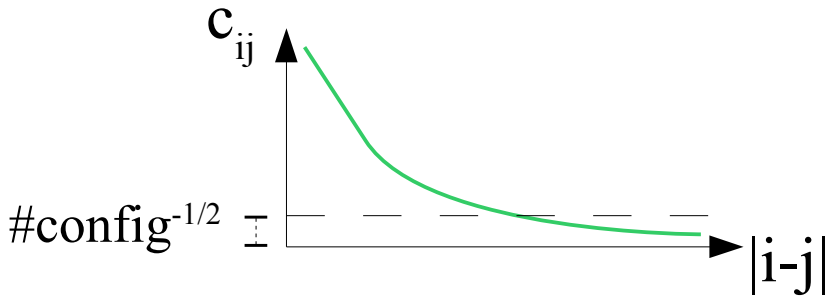
Cocco, M. (2011,2012) ; Barton, Cocco (2012)

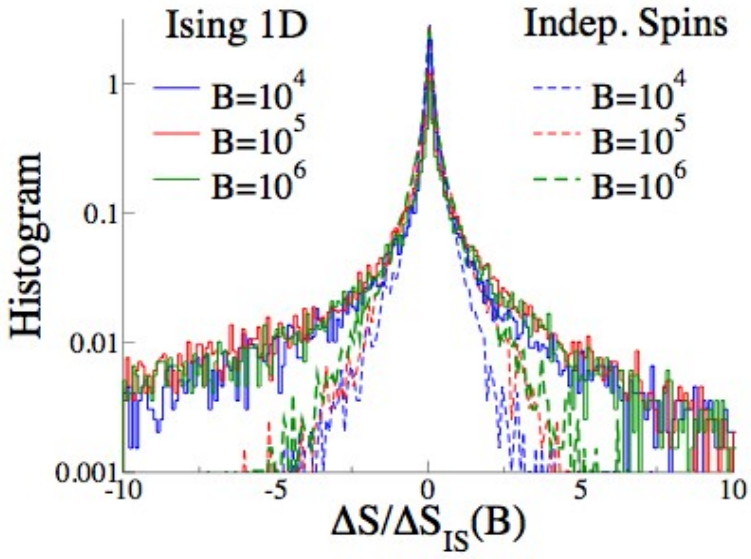
Adaptive cluster expansion (2)

Large cluster entropies are network-specific



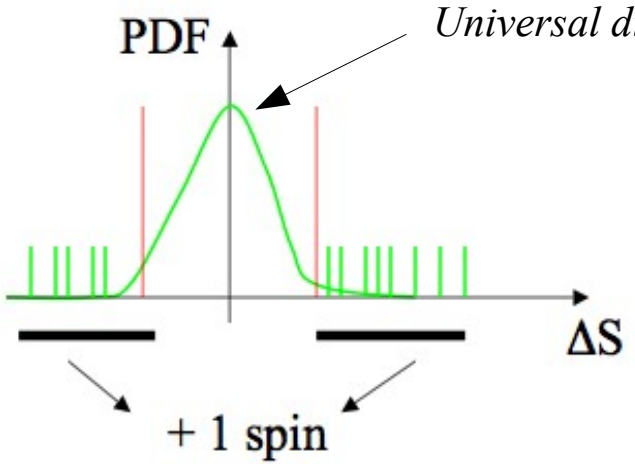
Small cluster entropies are due to sampling noise



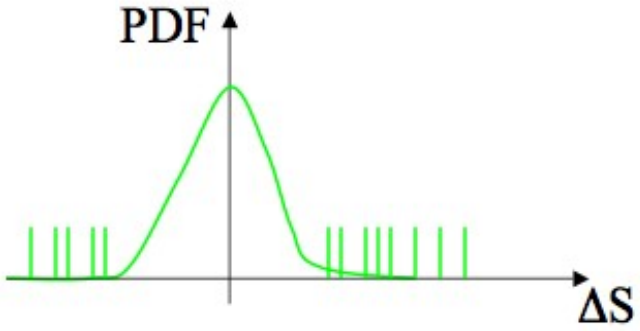


Adaptive cluster expansion (3)

clusters of K spins



clusters of K+1 spins

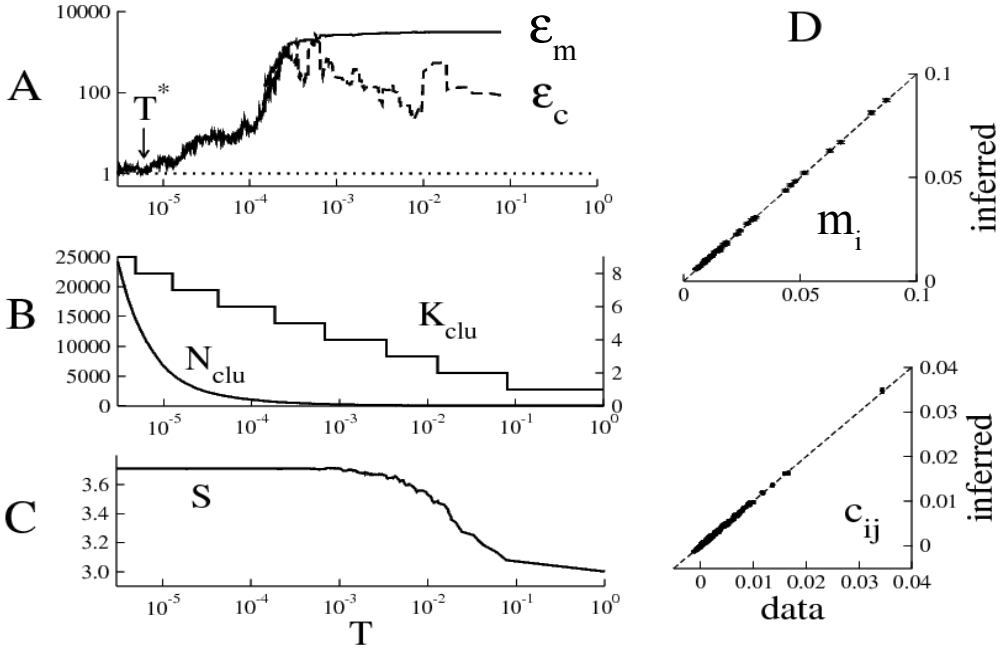


- Number & size of clusters adapt to data structure
- J is (almost surely) unveiled if B >> log N (and not N)
- Successful on critical Ising models in 2D

Application to retinal recordings (1)

N=32-60 ganglion cells recorded for about 2000 sec (spontaneous activity)

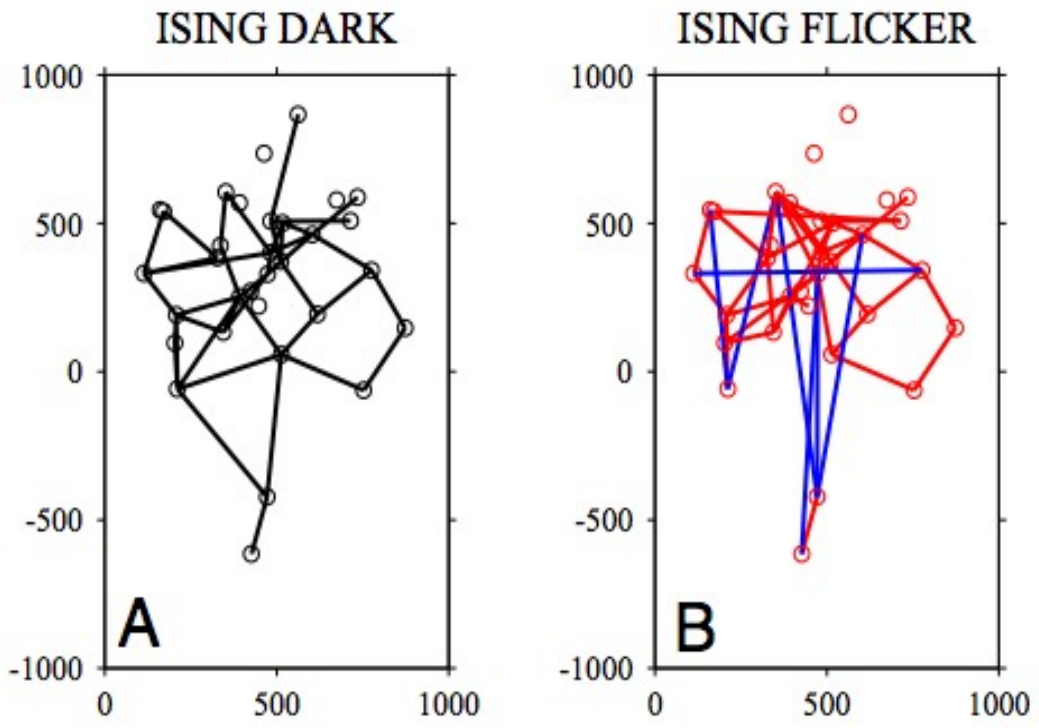
Meister et al. (2003)



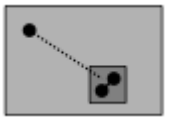
Application to retinal recordings (2)

Maps in retinal plane

Cocco, M., Leibler (2009)



- network of conserved and « nearest-neighbour » interactions
- long-range, stimulus-dependent interactions
- susceptibility vanishes for d > 500 micro-meters



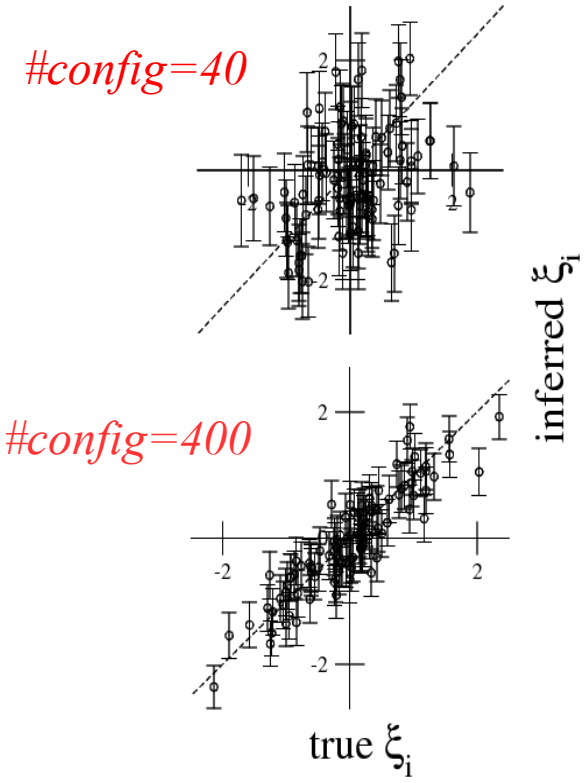
Useful for : population coding (cf talk by Vijay)
Tomorrow ...

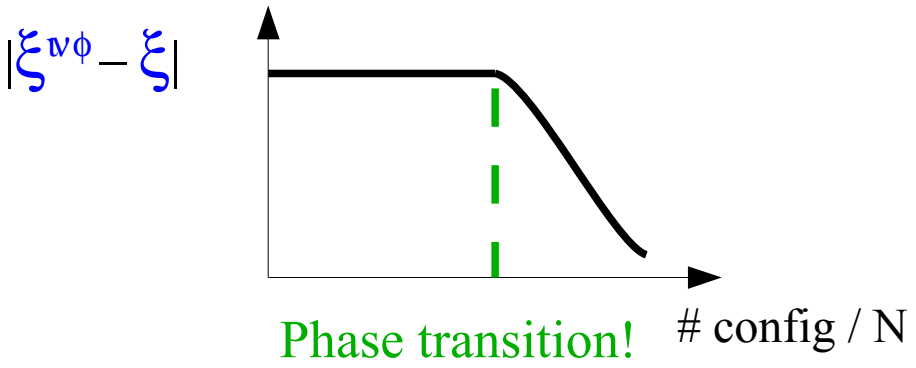
Analytical approaches

- Mean field inference
- Importance of prior(s)
- Pseudo-likelihood algorithms
- Advanced statistical physics methods
- Inverse Hopfield model

Inverse Hopfield model : retarded learning phase transition

Example: N=100, ξ = Gaussian (0,.7)



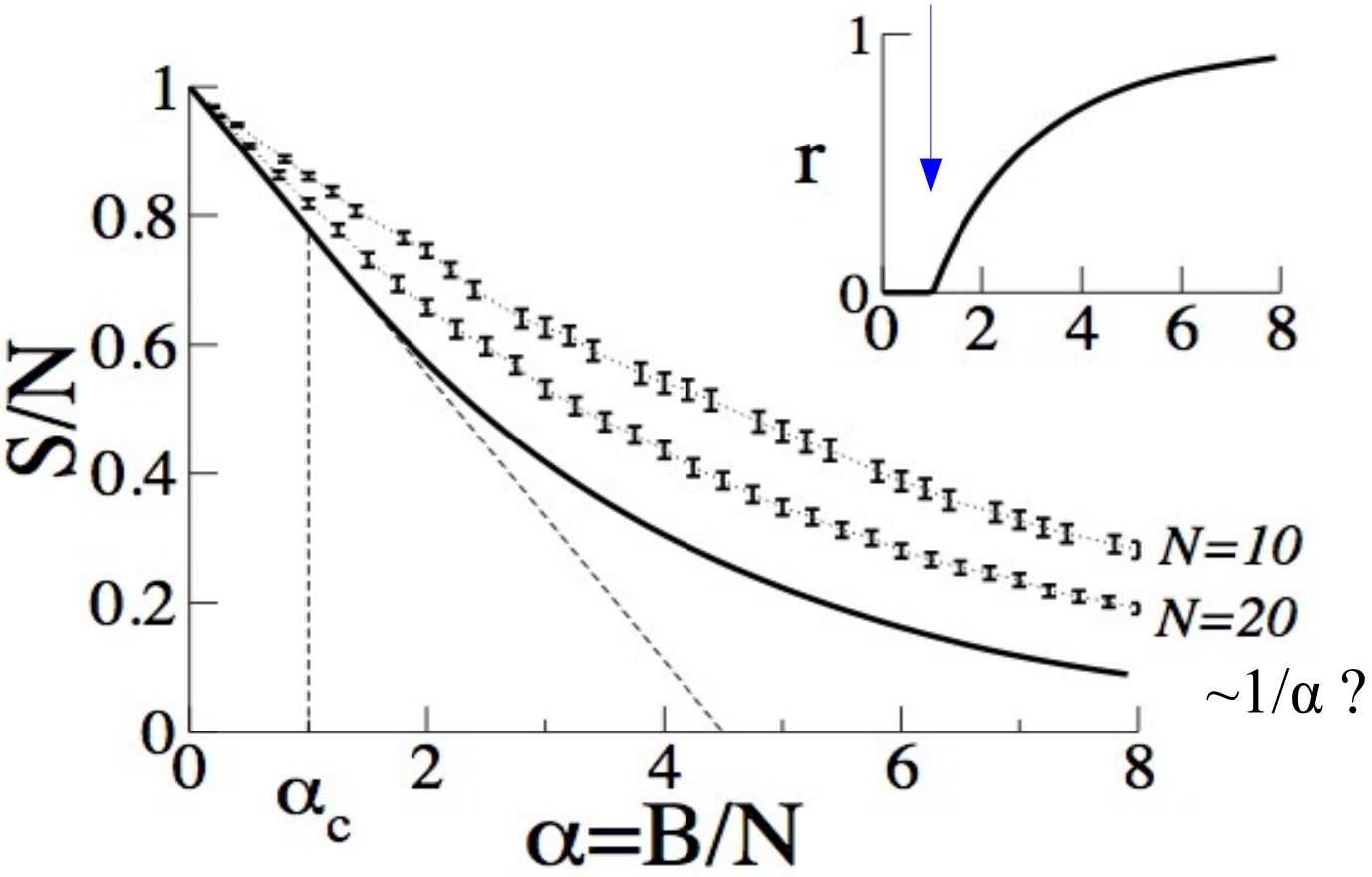


Watkin, Nadal (1994)
Baik, Ben Arous, Peche (2005)
Cocco, R.M. (2011)

Inverse Hopfield Model : posterior entropy of patterns

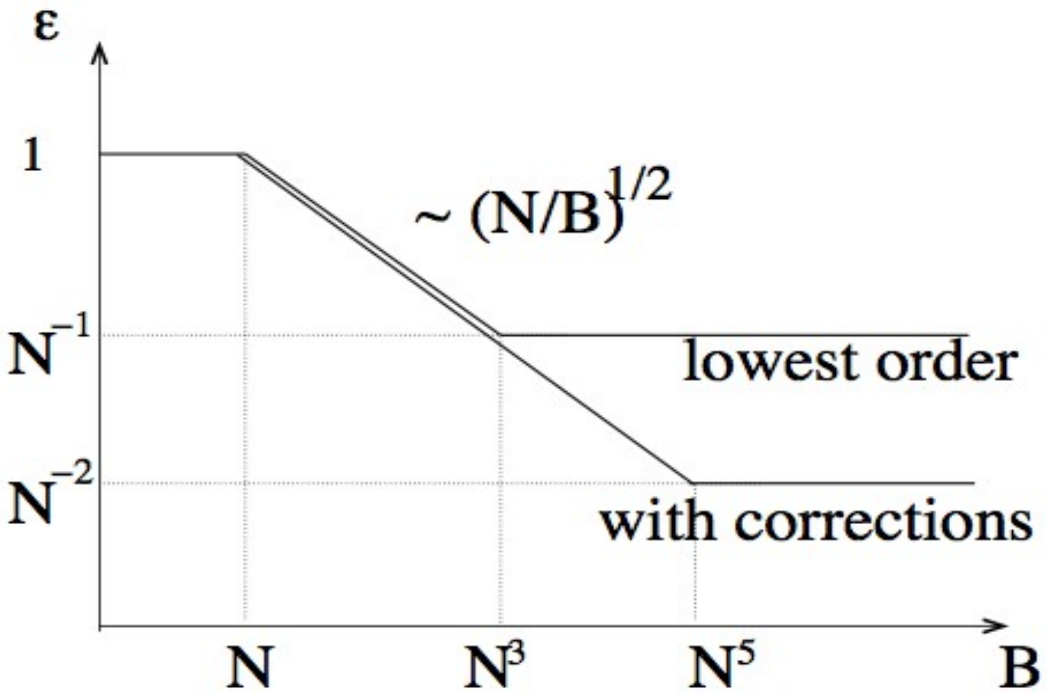
Example : P=1 pattern, L=2, $\alpha_L = 1$

Retarded learning transition



Cocco, M., Sessak (2011)

Inverse Hopfield Model : error on the inferred patterns



For pattern components ~1, corrections to S_0 are useless (at best) unless many configurations are available ...

→ Mean-Field is better (even if wrong) when few data are available

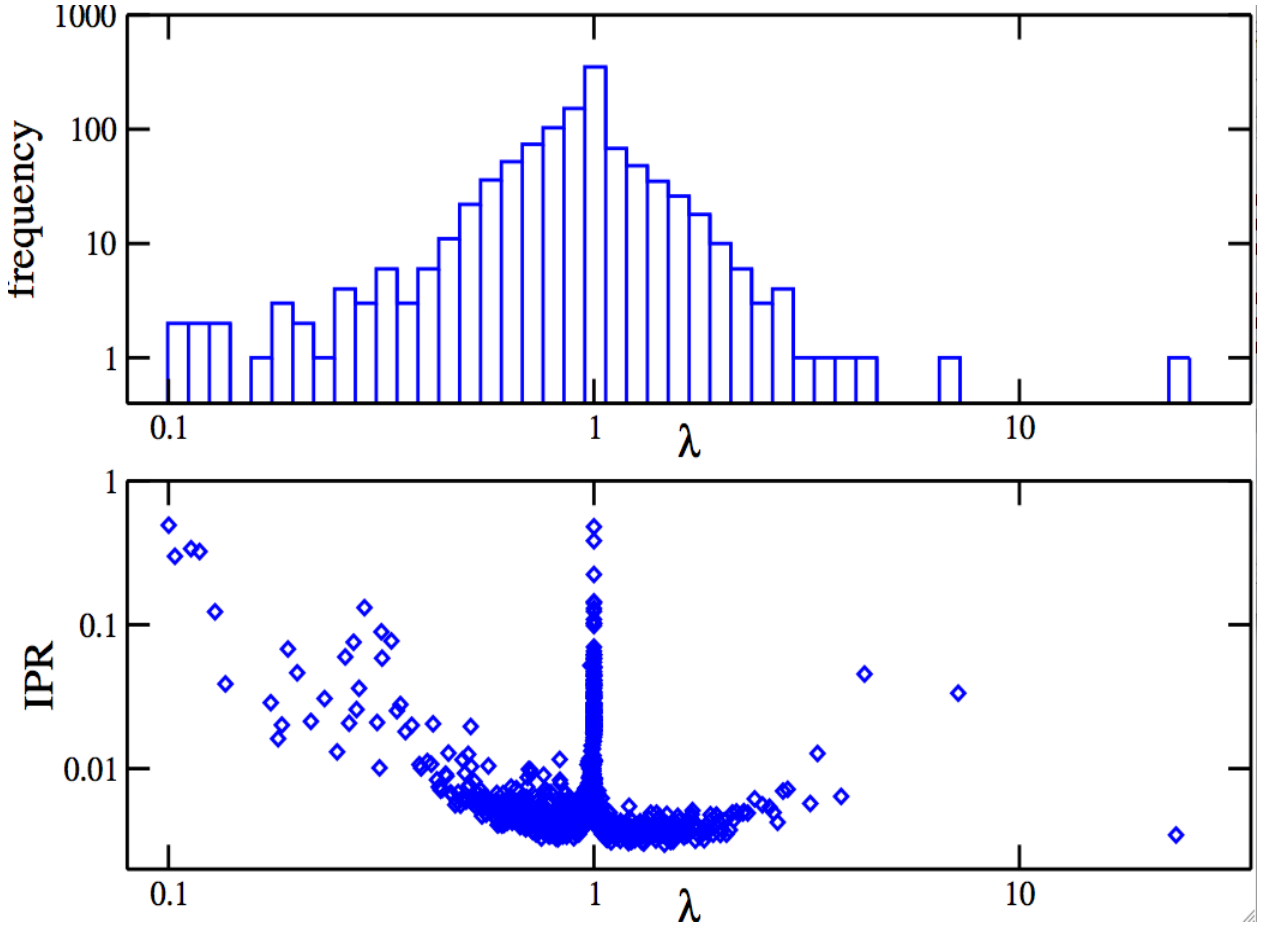
Applications to covariation in protein families (1)

- Hopfield-Potts model : 20 + 1 symbols
- Multi-sequence alignment : compute 1-, 2-residue frequencies
- Preprocessing of data :
  clustering (discount groups of similar sequences)
  pseudo-count for amino-acids a never present on site i
- Computation of patterns

Applications to covariation in protein families (2)

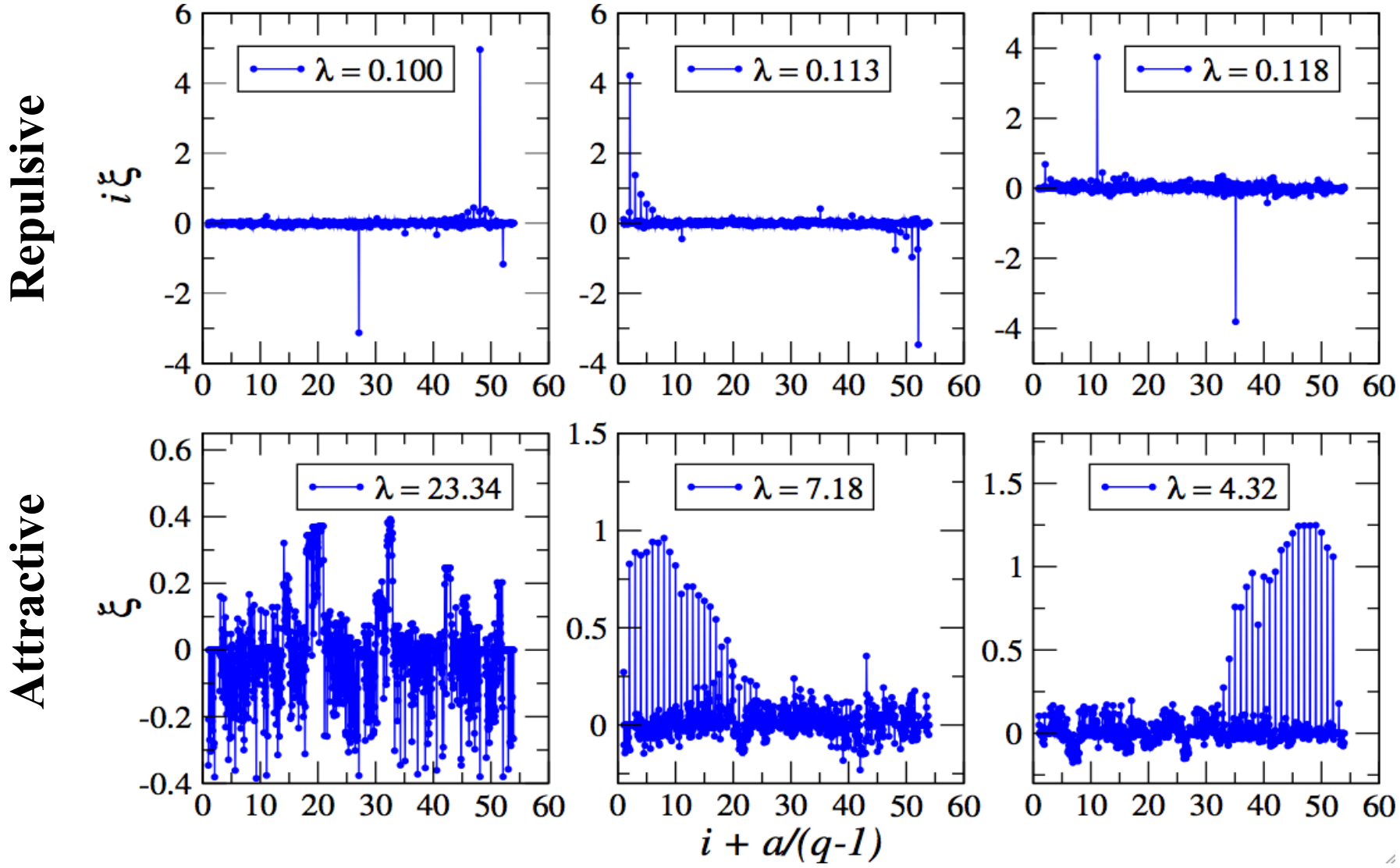
spectrum

inverse participation ratio



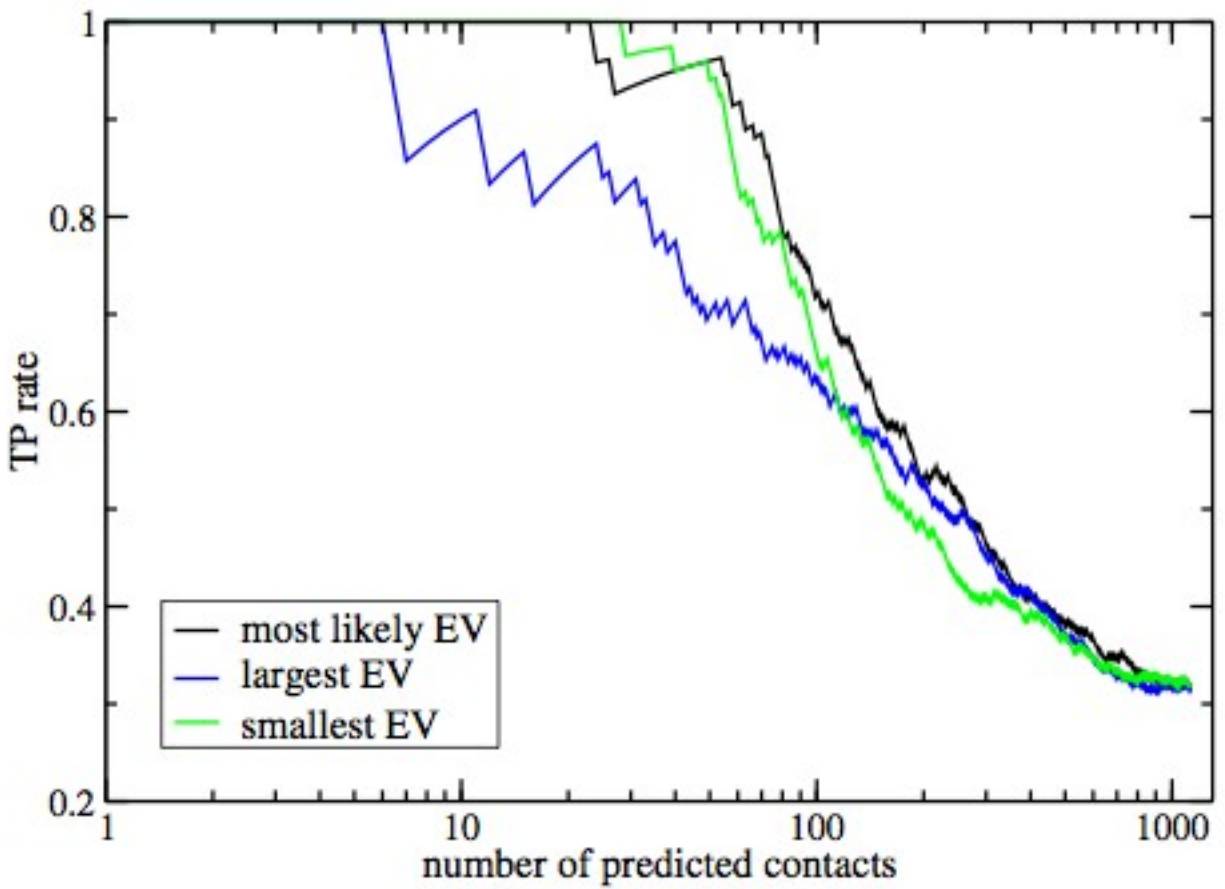
Cocco, M., Weigt (2012)

Applications to covariation in protein families (3)



Cocco, M., Weigt (2012)

Applications to covariation in protein families (5)



Contacts can be predicted from repulsive patterns only

Cocco, M., Weigt (2012)