Principal Component Analysis (PCA)

Goals :

- Reduce the dimensionality of a data-set.
- Increase the interpretability of data while preserving the maximum amount of information
- Help the visualization and clusterization of multi-dimensional data:

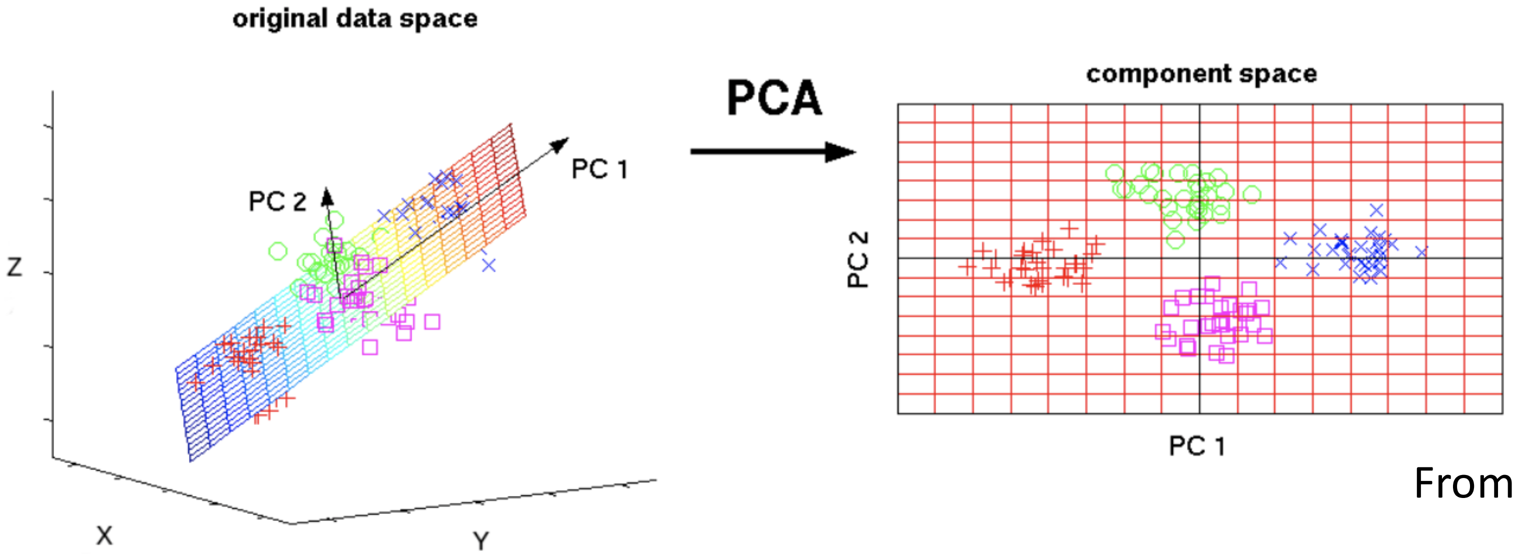
In many applications the first two components of the PCA are used to represent the data, identify clusters of closely related data points, and represent neural trajectories during the recording.

How?

- A linear transformation in a new coordinate system where most of the variability of the data can be explained with fewer dimensions than initial data.
- In practice find the eigenvectors and associated eigenvalues of the correlation matrix.

Principal Component Analysis (PCA)

A change of variables in a new coordinate system: new directions which constitute an **orthonormal** basis.



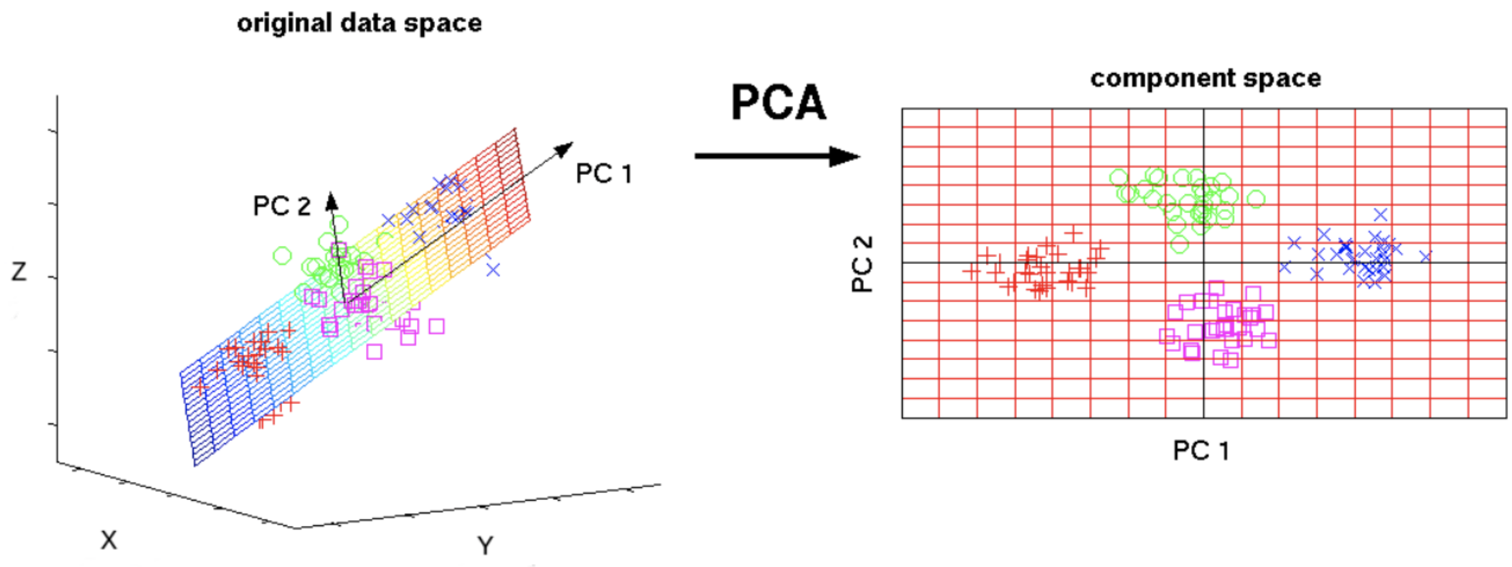
From wikipedia

- The first principal component of a set of variable is the variable formed by a linear combination of the original variables and explaining as much variance in the data as possible

- The second component explains most variance in the data in the space orthogonal to the first direction and so on..

Connection between PCA and K-means

Principal directions connect the center of the cluster.



PC1 is the line connecting the 2 clusters which are the most distant : maximal Variability of the data

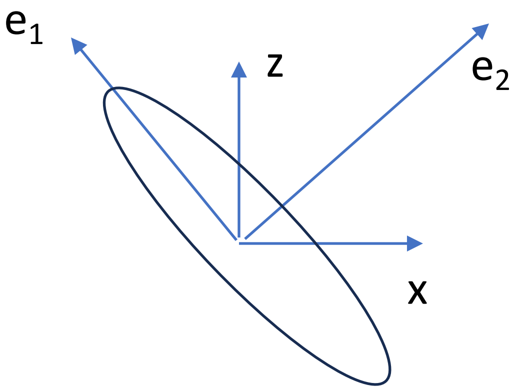
By reducing the dimensionality the distance between cluster is kept while the distance between points in the same cluster is reduced through the projection. This allows to reduce noise.

PCA in geometry and linear algebra

PCA was invented in 1901 by Karl Pearson as an extension of the principal axis theorem in geometry and linear algebra.

The principal axis are lines generalizing of the major and minor axis of an ellipse. They are perpendicular, and can be found by a matrix diagonalization

Example 2:



$$5\,x^2 + 8\,x\,z + 5\,z^2 = 1$$

$$\widehat{T} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$$

Diagonalizing matrix : solving

$$\sum_j \widehat{T_{ij}}\, e_j = \lambda\, e_i$$

$$\lambda_1 = 1 \qquad \mathbf{e_1} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\lambda_2 = 9 \qquad \mathbf{e_2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

e: eigenvectors, λ:eigenvalues

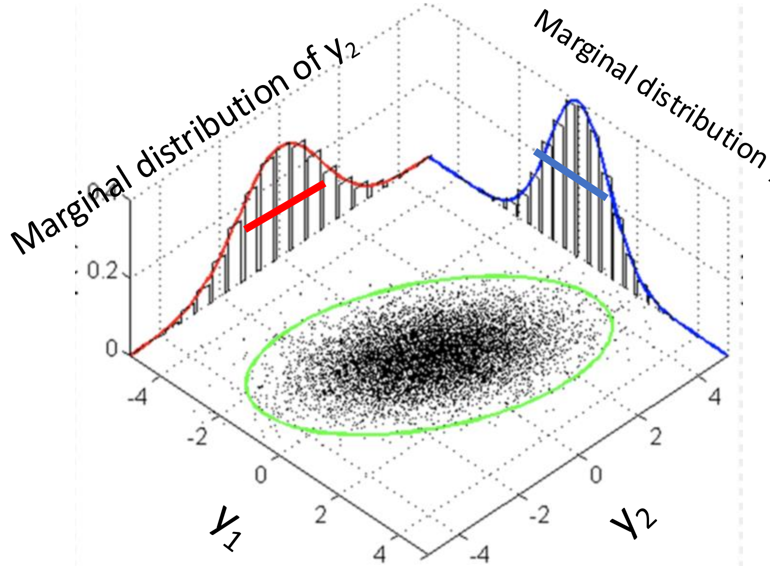
PCA and Multivariate Gaussian distributions

Simple model of stochastic variables which are correlated and in a high dimensional space:

Joint distribution of $\mathbf{y}=(y_1,y_2,\ldots y_N)$:

$$P(\mathbf{y}|\hat{T}) \propto e^{-\frac{1}{2}\sum_{i,j} y_i T_{ij} y_j}$$

Drawing values from the distribution for N=2



PCA and Multivariate Gaussian distributions

Simple model of stochastic variables which are correlated and in a high dimensional space:

Joint distribution of $\mathbf{y}=(y_1,y_2,\ldots y_N)$:

$$P(\mathbf{y}|\hat{T}) \propto e^{-\frac{1}{2}\sum_{i,j} y_i T_{ij} y_j}$$

$\mu_i = \int d\mathbf{y}\, P(y|\hat{T})\, y_i = 0$ Averages

$C_{ij} = \int d\mathbf{y}\, P(y|\hat{T})\, y_i y_j = [\hat{T}^{-1}]_{ij}$ Covariations

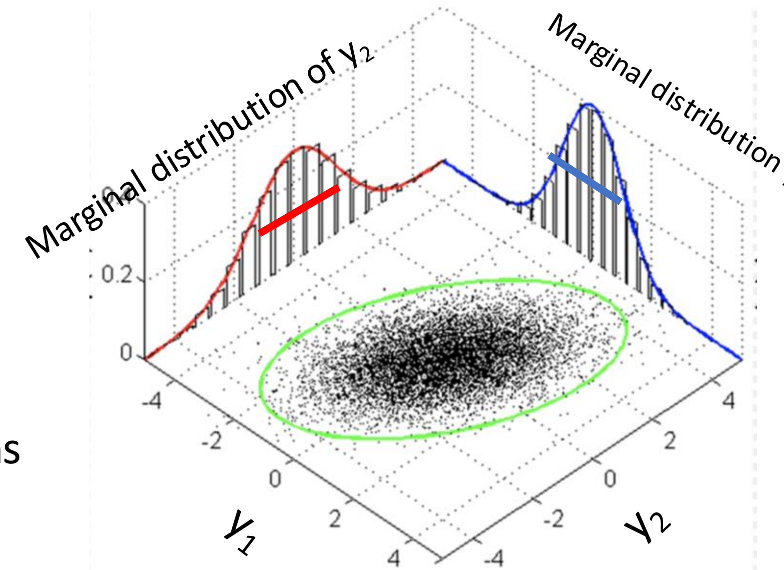
$\hat{T}$ =

3.5	1
1.	1.5

$\hat{C}$ =

0.35	-0.23
-0.23	0.82

Drawing values from the distribution for N=2



PCA and Multivariate Gaussian distributions

Simple model of stochastic variables which are correlated and in a high dimensional space:

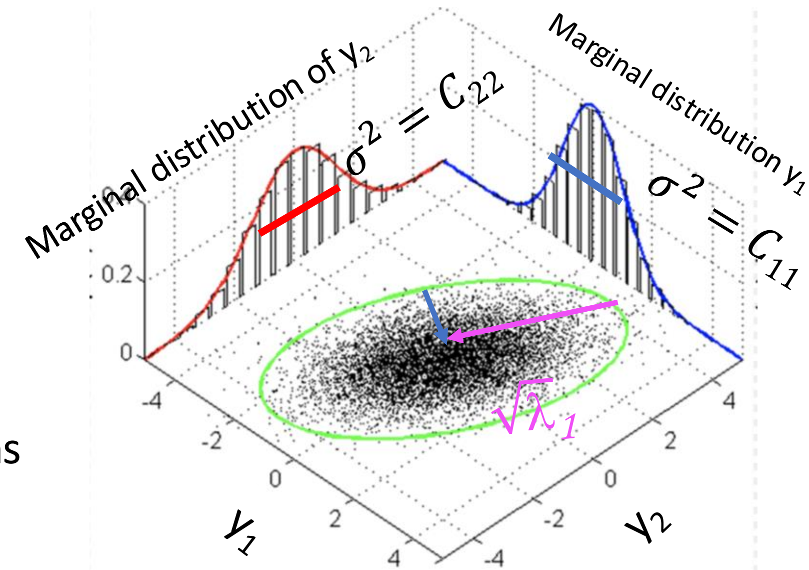
Joint distribution of $\mathbf{y}=(y_1,y_2,\ldots y_N)$:

$$P(\mathbf{y}|\hat{T}) \propto e^{-\frac{1}{2}\sum_{i,j} y_i T_{ij} y_j}$$

$\mu_i = \int d\boldsymbol{y}\, P(y|\hat{T})\, y_i = 0$ Averages

$C_{ij} = \int d\boldsymbol{y}\, P(y|\hat{T})\, y_i y_j = [\hat{T}^{-1}]_{ij}$ Covariations

Drawing values from the distribution for N=2



$$\hat{T} = \begin{array}{c|c} 3.5 & 1 \\ \hline 1. & 1.5 \end{array} \qquad \hat{C} = \begin{array}{c|c} 0.35 & -0.23 \\ \hline -0.23 & 0.82 \end{array}$$

$\lambda^c_1=0.92$, $\lambda^c_2=0.26$,

Principal components are (principal axis of the ellipse, or direction of maximal variability) are the eigenvectors of the covariance matrix $\hat{C}$ (same as the ones of $\hat{T}$), and the eigenvalues of $\hat{C}$ (inverse of the ones of $\hat{T}$) give the variances along them.

PCA and Multivariate Gaussian distributions

Case of L= 3 variables

$$P(\mathbf{y}|\hat{T}) \propto e^{-\frac{1}{2}\sum_{i,j} y_i T_{ij} y_j}$$

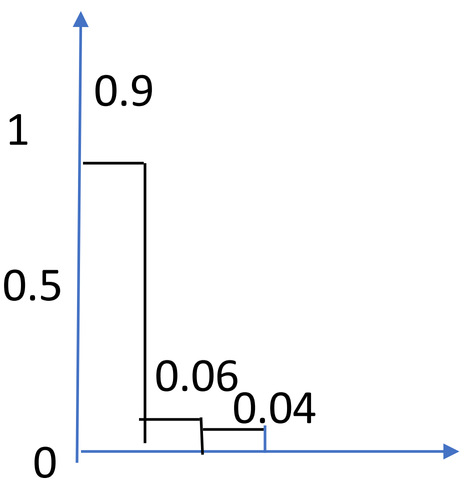
$\hat{T}$ =

3.5	-2	0
-2	1.5	0.5
0	0 .5	2.5

Eigenvalues and eigenvectors of the covariance matrix:

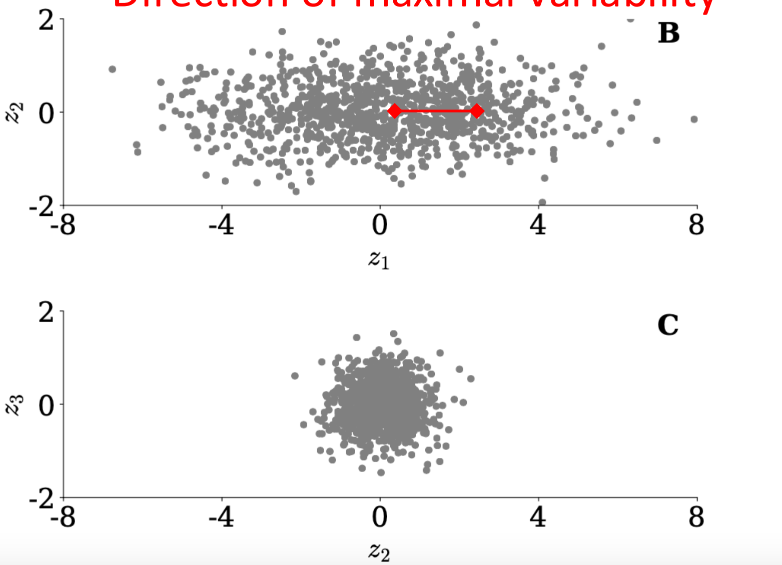
$\lambda^c{}_1 = 5.4 \quad \lambda^c{}_2 = 0.39 \quad \lambda^c{}_3 = 0.21$ $\qquad \mathbf{e}_{1,}\mathbf{e}_{2,}\mathbf{e}_3$

Fraction of explained variability $\dfrac{\lambda_i}{\sum \lambda_i}$



$$z_\mu = \mathbf{e}_\mu^T \cdot \mathbf{y}$$

Direction of maximal variability



$\sqrt{\lambda^c{}_1/\lambda_2{}^c}$ =4.7

$\sqrt{\lambda_2{}^c\ /\lambda_3{}^c}$ =1.4

Exemple 1: PCA and facial recognition: EigenFace

Image is decomposed on pixels (128 X 128: 2^{14} dimensions)
and gray levels (2^8) at each position: y(**x**)
M=115 Pictues

Average face:

[Sirovich & Kirby 1986 J. OptcSoc Am]

Exemple 1: PCA and facial recognition: EigenFace

Image is decomposed on pixels (128 X 128: 2^{14} dimensions)
and gray levels (2^8) at each position: y(**x**)
M=115 Pictues

Principal component analysis of the covaraince of the ensemble of faces : $\hat{\mathbf{e}}_\mu$

Average face:

[Sirovich & Kirby 1986 J. OptcSoc Am]

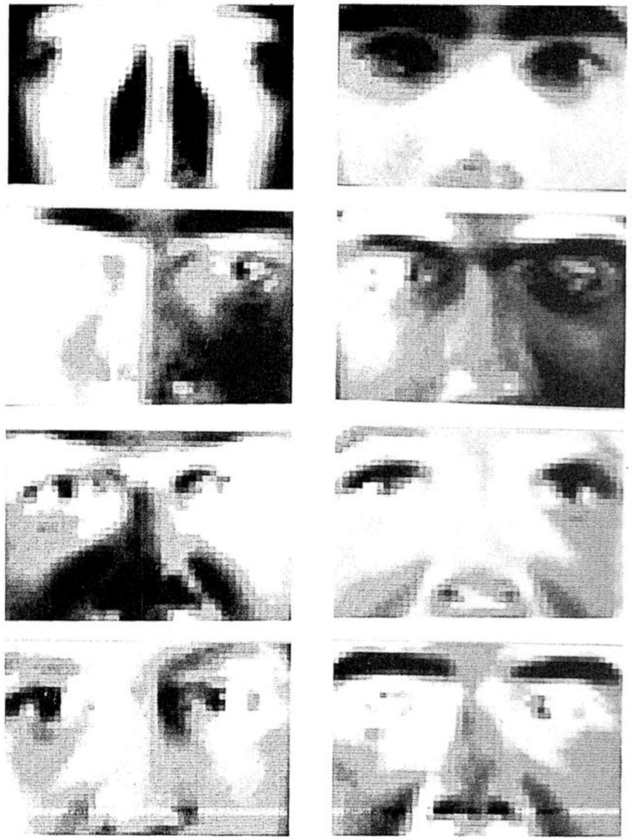


Fig. 4. First eight eigenpictures starting at upper left, moving to the right, and ending at lower right.

Exemple 1: PCA and facial recognition: EigenFace

Image is decomposed on pixels (128 X128: 2^{14} dimensions)
and gray levels (2^8) at each position: y(**x**)
M=115 Pictues

$$\mathbf{y} = \sum_{\mu} z_{\mu} \mathbf{e}_{\mu}$$

Reconstruct a face through its projections on the principal components

Reconstructed images

Real image

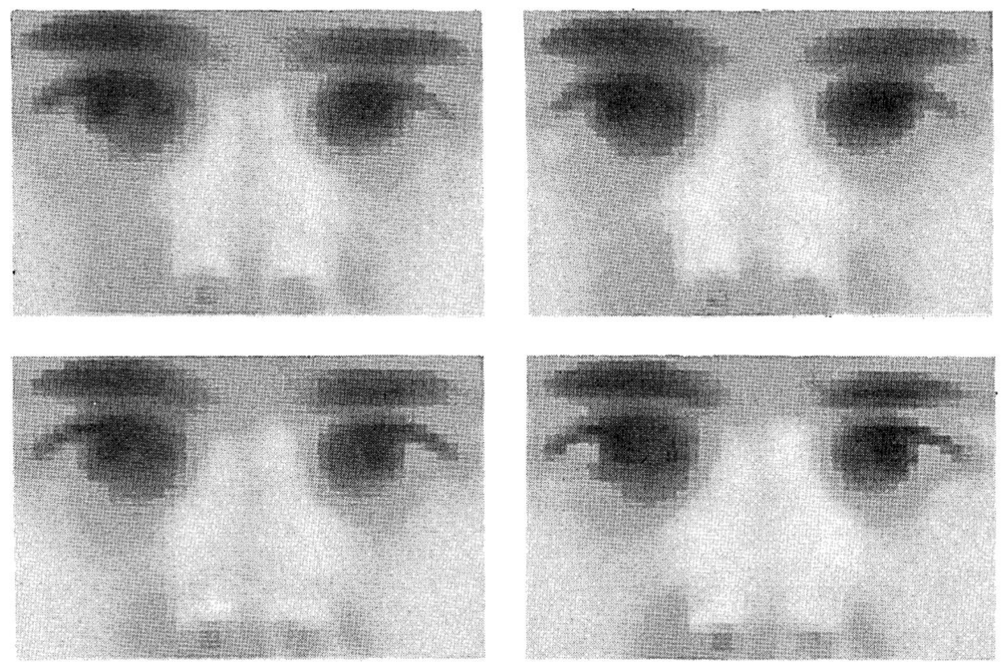
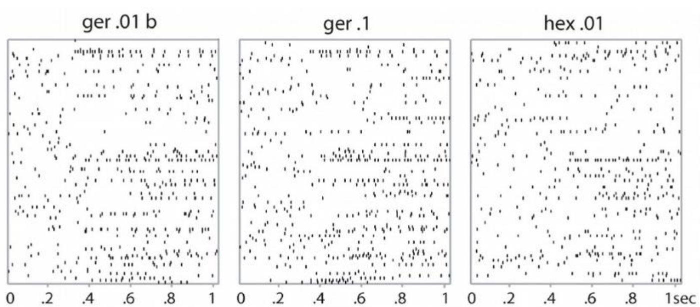


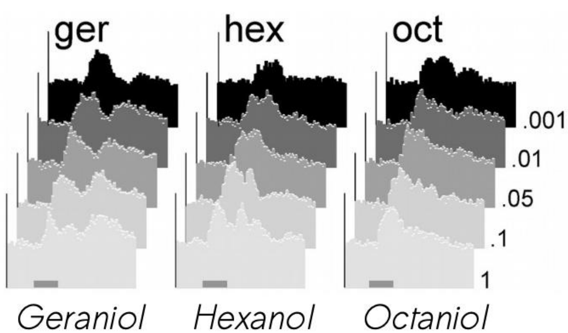
Fig. 5. Approximation to the exact picture (middle panel of Fig. 3) using 10, 20, 30, and 40 eigenpictures.

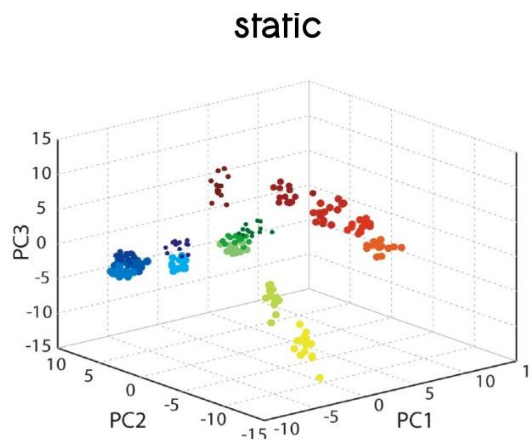
[Sirovich & Kirby 1986 J. OptcSoc Am]
[Turk & Pentland 1991 J. Cognitive Neuroscience]

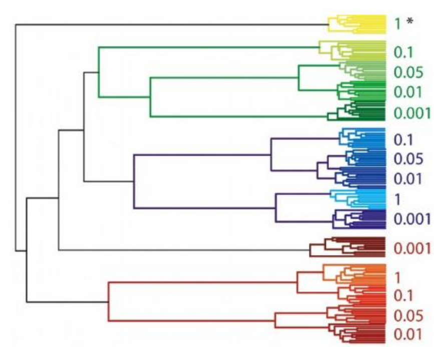
From 16000 to 10-40 dimensions

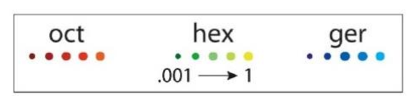
Exemple 2: Odor Identity and Concentration Encoding in the locust olfactory system





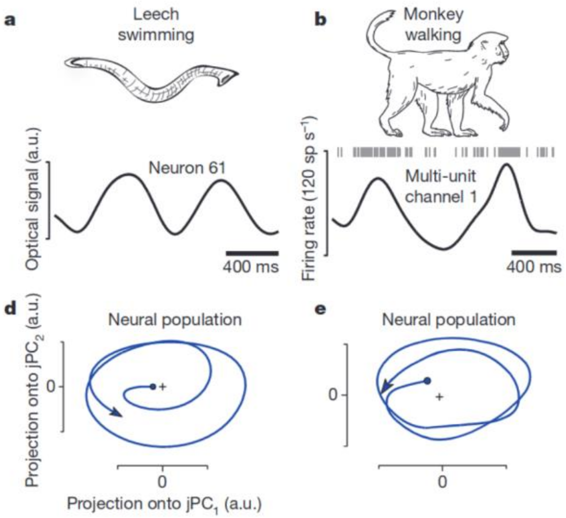


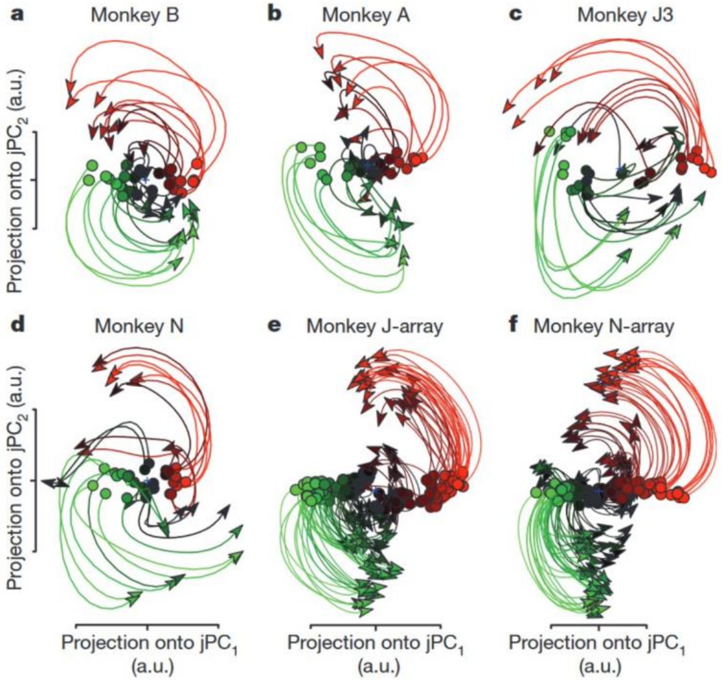




Stopfer 2023

Exemple 3: Neural Trajectories during reach in macaque

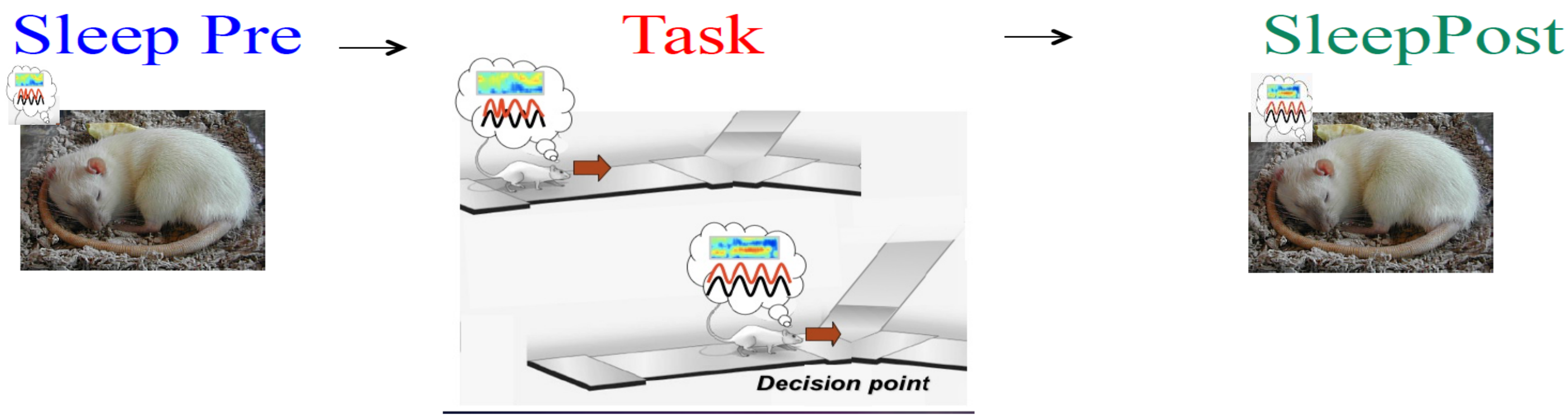




Churchland 2013

Stopfer 2023

Exemple 4 Application of PCA in sleep studies



Replay of rule-learning related neural patterns in the prefrontal cortex during sleep

nature neuroscience (2009)

Adrien Peyrache[1], Mehdi Khamassi[1,2], Karim Benchenane[1], Sidney I Wiener[1] & Francesco P Battaglia[1,3]

Peyrache et al JCN 2009

How many Principal Components to take?

Method 2: use Marcenko Pastur Spectrum for Random Covariance Matrices

Consider a multivariate distribution on uncorrelated random variables $\langle y_i \rangle = 0$, $\langle y_i y_j \rangle = 0$, $\langle y_i y_i \rangle = 1$

$$\widehat{C} = \begin{array}{c|c} 1 & 0 \\ \hline 0 & 1 \end{array}$$

Take M= measures

$$C_{ij} = \frac{1}{M} \sum_{b=1}^{M} y_{bi} y_{bj} \ .$$

$$\rho_{MP}(\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\, r\, \lambda} \ ,$$

$$r = \frac{N}{M}$$

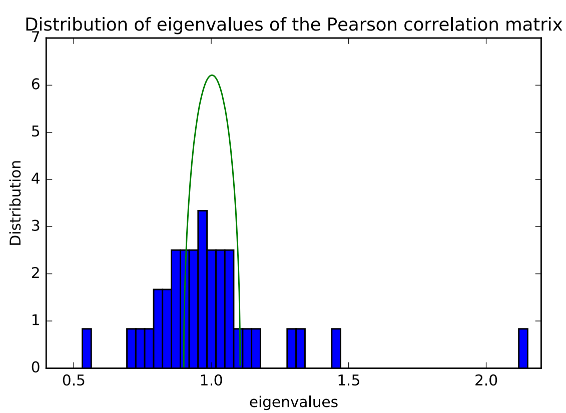
In our case

$M = 1.4\ 10^4$

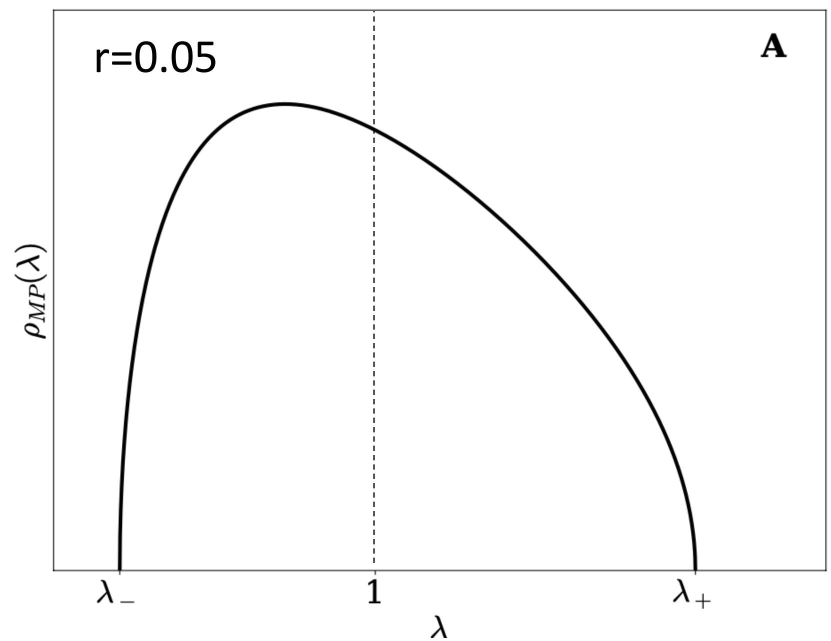
$N = 37$

$r = N/M = 2.6\ 10^{-3}$

$\lambda_- = 0.9$, $\lambda_+ = 1.1$,



True case



$$\lambda_- = (1 - \sqrt{r})^2 \qquad \lambda_+ = (1 + \sqrt{r})^2$$

Limitations of Principal Components Analysis

1. It is not the optimal way to find cell assembly: an arbitrary threshold to select Neurons participating in it.
2. One neuron may participate to several cell assembly.

1. It can only perform linear transformations

Beyond PCA

1. Sparse PCA : approximation of the principal vector, with many component to zero

1. Independent Component Analysis, enforce separations of neurons in different Components

3. Kernel methods -> non linear transformation

Take Home Message

- It is difficult to interpret data in an high dimensional space

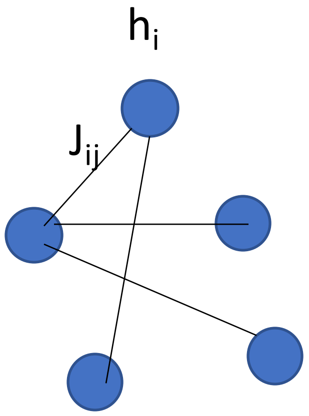
- PCA Reduce the dimensionality of the data, by projecting them in the directions of their maximal variability

- PCA is a powerfull method in many fields of cognitive science

- Models from theory of random matrices to decide how many component to take .

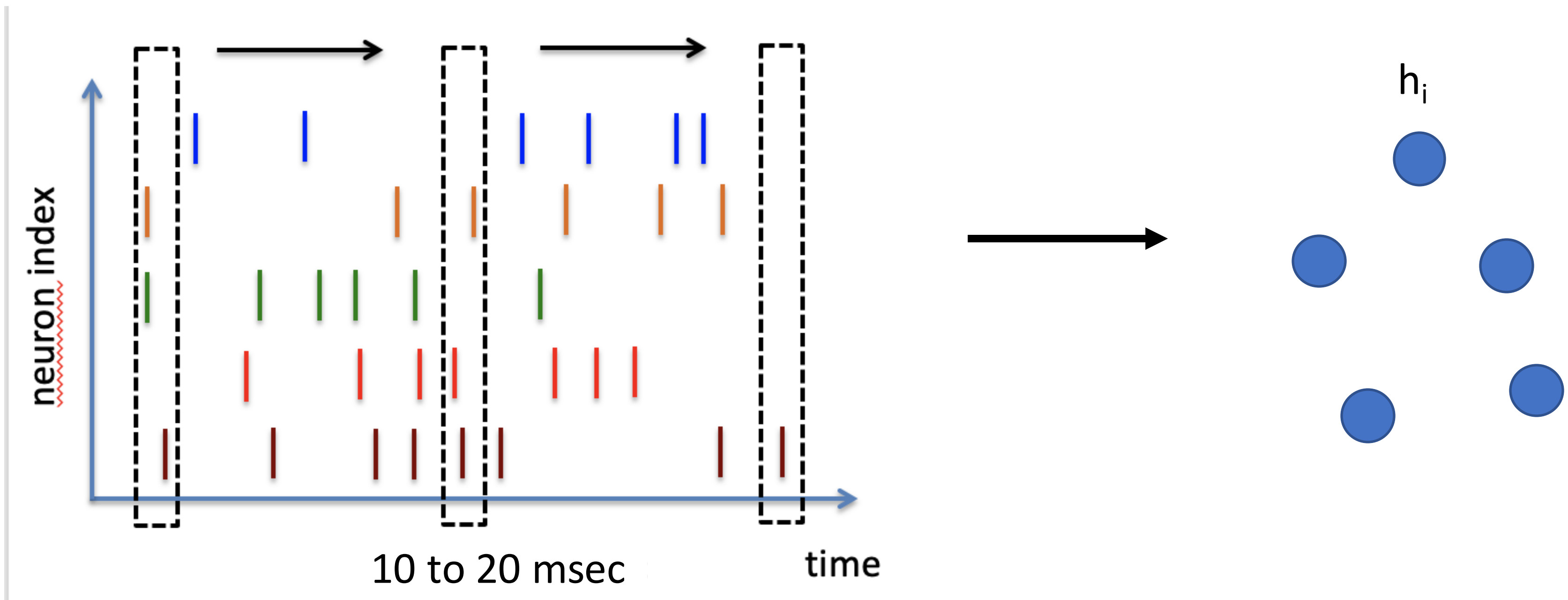
- **Network Inference: infer effective connectivity from neural recordings, at the basis of population coding.**

- Network inference for Gaussian variabels: from the correlation matrix of a Multi-Variate Gaussian distribution to the Interaction matrix .



Constrain the Max Ent model to reproduce firing frequencies

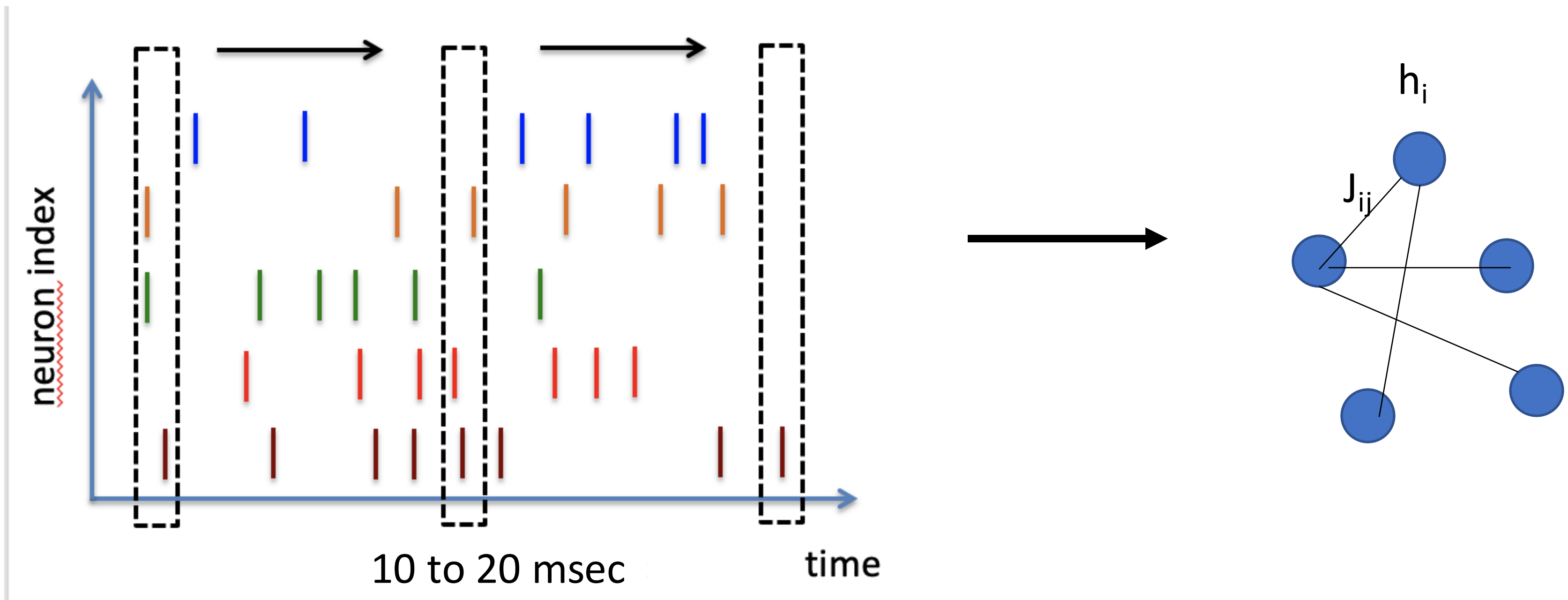
- Independent model



From recording to inference of local field parameters.

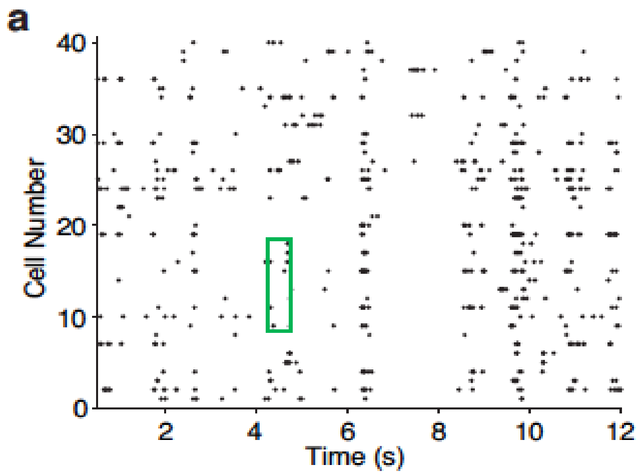
Constrain the Max Ent model to reproduce pairwise correlatations in the neural activity.

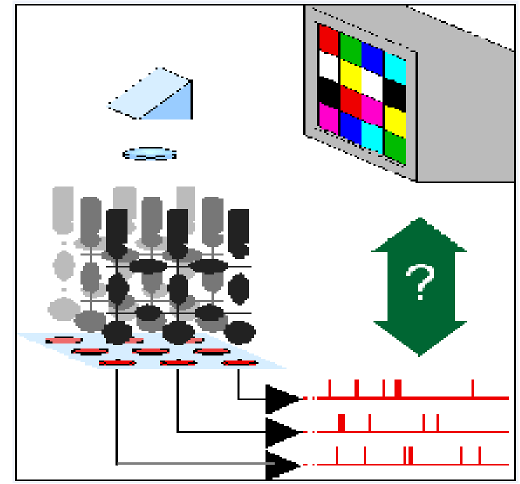
- Hypothesis neurons do not spike independently but there is a population coding



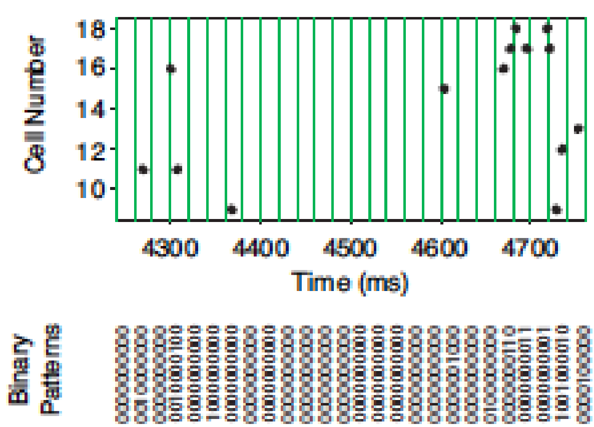
From recording to inference of local field parameters and effective connectivity.

MaxEnt model to reproduce correlated activity in a neural population (Schneidman et al. 2006, Shlens et al. 2006)





Time is discretized in time windows of size Δt = 20ms



$$s_i(k) = \begin{cases} 1 \text{ if at least one spike in time window k} \\ 0 \text{ if no spike in time windows k} \end{cases}$$

Spiking probabilities and **pairwise correlations from DATA**

$$p_i = \frac{1}{M}\sum_{b=1}^{M} s_i(b)\ , \qquad p_{ij} = \frac{1}{b}\sum_{b=1}^{M} s_i(b)s_j(b)$$

$$P(s_1\ , s_2\ , \ldots\ , s_N)?$$

Inference of Coupling Matrix in Multivariate Gaussian distributions

Multi-Variate Gaussian Distribution: $$p(\boldsymbol{y}|\widehat{T}) = \frac{\sqrt{\det\,\widehat{T}}}{(2\pi)^{L/2}}\, e^{-\frac{1}{2}\,\boldsymbol{y}^T\cdot\widehat{T}\cdot\boldsymbol{y}}$$

Log-likelihood from M sampled configurations:

$$\frac{1}{M}\log P(Y|T) = \frac{1}{M}\sum_{m=1}^{M}\log P(\boldsymbol{y}^m|T) \quad = \frac{1}{2}\log\det T - \frac{1}{2M}\sum_{m=1}^{M}(\boldsymbol{y}^m)^T\cdot T\cdot\boldsymbol{y}^m$$

Derivative with respect to T_{ij} to infer the maximal likelihood values of **T:**

Inference of Coupling Matrix in Multivariate Gaussian distributions

Multi-Variate Gaussian Distribution: $$p(\boldsymbol{y}|\widehat{T}) = \frac{\sqrt{\det\ \widehat{T}}}{(2\pi)^{L/2}}\, e^{-\frac{1}{2}\,\boldsymbol{y}^T\cdot\widehat{T}\cdot\boldsymbol{y}}$$

Log-likelihood from M sampled configurations:

$$\frac{1}{M}\log P(Y|T) = \frac{1}{M}\sum_{m=1}^{M}\log P(\boldsymbol{y}^m|T) \quad = \frac{1}{2}\log\det T - \frac{1}{2M}\sum_{m=1}^{M}(\boldsymbol{y}^m)^T\cdot T\cdot\boldsymbol{y}^m$$

Derivative with respect to T_{ij} to infer the maximal likelihood values of **T:**

Using the general identity: $$\frac{\partial}{\partial T_{ij}}\log\det T = (T^{-1})_{ji}$$

Inference of Coupling Matrix in Multivariate Gaussian distributions

Multi-Variate Gaussian Distribution: $$p(\boldsymbol{y}|\widehat{T}) = \frac{\sqrt{\det\,\widehat{T}}}{(2\pi)^{L/2}}\, e^{-\frac{1}{2}\,\boldsymbol{y}^T\cdot\widehat{T}\cdot\boldsymbol{y}}$$

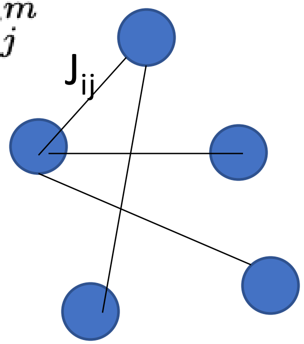
Log-likelihood from M sampled configurations:

$$\frac{1}{M}\log P(Y|T) = \frac{1}{M}\sum_{m=1}^{M}\log P(\boldsymbol{y}^m|T) \quad = \frac{1}{2}\log\det T - \frac{1}{2M}\sum_{m=1}^{M}(\boldsymbol{y}^m)^T\cdot T\cdot\boldsymbol{y}^m$$

Derivative with respect to T_{ij} to infer the maximal likelihood values of **T:**

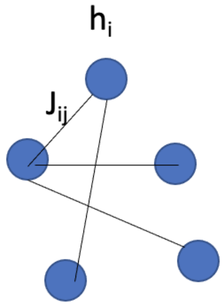
Using the general identity: $$\frac{\partial}{\partial T_{ij}}\log\det T = (T^{-1})_{ji}$$

We obtain: $T^{-1}_{ij}=C_{ij}$ with: $$C_{ij} = \frac{1}{M}\sum_{m=1}^{M} y_i^m y_j^m$$



The Coupling matrix is defined as minus the precision matrix: $J_{ij} \equiv -T_{ij}$

MaxEnt pairwise model



$$P(s_1, s_2, \dots, s_N) = \exp\Big(\sum_{i<j} J_{ij}\, s_i\, s_j + \sum_i h_i\, s_i\Big) / Z$$

Energy based models: $E(s_1, s_2, .., s_N) = -\sum_{i<j} J_{ij}\, s_i\, s_j - \sum_i h_i\, s_i$

Inverse
Ising problem

(From data)

$$p_i = \sum_{s_1,..s_N} P(s_1, s_2, \dots, s_N)\; s_i \quad \text{and} \quad p_{ij} = \sum_{s_1,..s_N} P(s_1, s_2, \dots, s_N)\, s_i\, s_j$$

Inhomogeneous system :40*41/2=820 parametres to infer

Find fields and couplings that minimize the cross-entropy

Remember that: Minimization of S_c is equivqlent to the Maximisation of the Log-Lokelihood of observing the M configurations of activity.

$$S_C = \min_{\{J_{ij},h_i\}} \left(\log Z[\{J_{ij},h_i\}] - \sum_{i<j} J_{ij}\, p_{ij} - \sum_i h_i\, p_i \right)$$

Find fields and couplings that minimize the cross-entropy

Remember that: Minimization of S_c is equivqlent to the Maximisation of the Log-Lokelihood

$$S_C = \min_{\{J_{ij},h_i\}} \left(\log Z[\{J_{ij},h_i\}] - \sum_{i<j} J_{ij}\, p_{ij} - \sum_i h_i\, p_i \right)$$

As in thermodynamics

$F = U - TS \quad T=1;$

$S = -F + U$

Find fields and couplings that minimize the cross-entropy

Remember that: Minimization of S_c is equivqlent to the Maximisation of the Log-Lokelihood

$$S_C = \min_{\{J_{ij},h_i\}} \left(\log Z[\{J_{ij},h_i\}] - \sum_{i<j} J_{ij}\, p_{ij} - \sum_i h_i\, p_i \right)$$

$$\frac{\partial S_C}{\partial h_i} = 0 \Rightarrow \frac{\ln Z[J,h]}{h_i} = p_i \equiv \sum_{s_1,..s_N} P(s_1,..s_N)\, s_i = p_i \,;$$

$$\frac{\partial S_C}{\partial J_{ij}} = 0 \Rightarrow \frac{\ln Z[J,h]}{J_{ij}} = p_{ij} \equiv \sum_{s_1,..s_N} P(S)\, s_j\, s_i = p_{ij} \,;$$

Exact Minimization of S_c by gradient descent is possible for small systems but

Computational problem: calculation of Z, 2^N configurations $\left\} \begin{array}{ll} 2^{10} = 1024 & \text{ok!} \\ 2^{20} = 10^6 & \text{difficult} \end{array} \right.$

Find fields and couplings that minimize the cross-entropy

$$S_C = \log Z[\{J_{ij}, h_i\}] - \sum_{i<j} J_{ij}\, p_{ij} - \sum_i h_i\, p_i \quad + \Gamma \left(\sum_{i<j} J_{ij}^2\right)$$

Is it convex? Yes : The Hessian ≥ 0

Is the solution unique? No: The Hessian may have some zeros

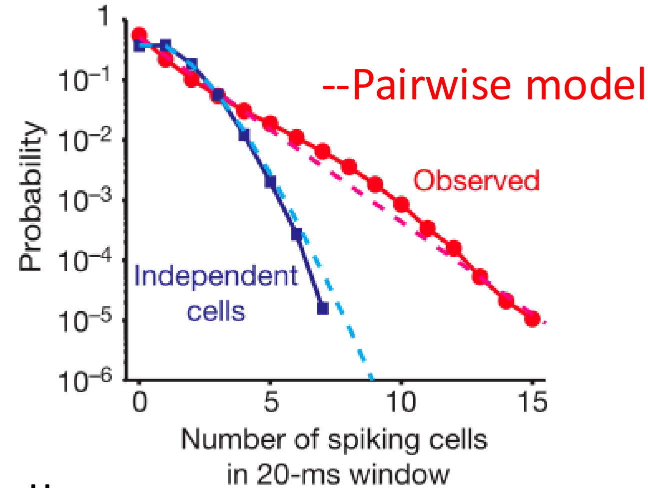
Solved by a regularization parameters $\Gamma \approx 1/M$

Eg: When two neurons never spike in the same time window $J_{ij} \rightarrow -\infty$

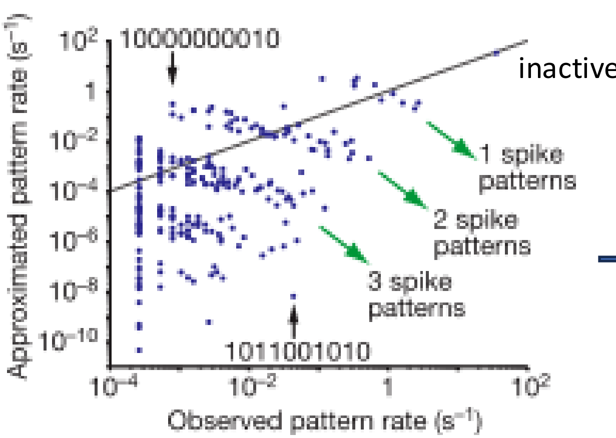
Is the pairwise model good?

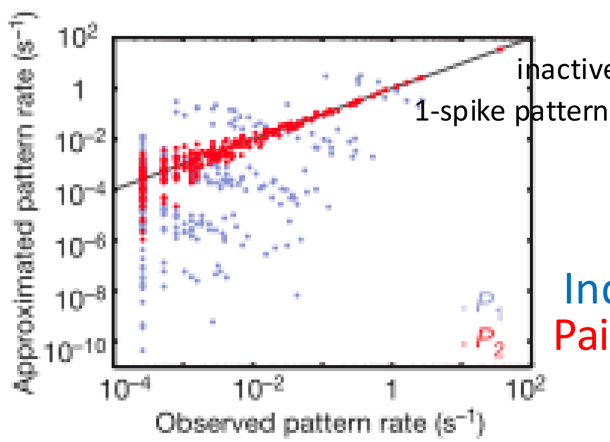
The inferred model correctly predict higher order moments and patter of multi-cell firing.

Schneidmann et al., Nature (2006)



And frequencies of 10 cells firing patterns





Independent

Pairwise model

Qualitatively good also for very rare configurations = unobserved in recording …!

Total Entropy ≈1.25 bits : 0.125 per neurons: Number of dominant/recurrent configurations $2^{1.25}=4$ out of $2^{10}=1024$ configurations