Computational Neuroscience ESPRC Workshop
--Warwick

Information-Geometric Studies on Neuronal Spike Trains

Shun-ichi Amari
RIKEN Brain Science Institute
Mathematical Neuroscience Unit

Neural Firing

$x_1 \quad x_2 \quad x_3 \quad \cdots\cdots \quad x_n \qquad x_i = 1, 0$

$p(\mathbf{x}) = p(x_1, x_2, \ldots, x_n)$: joint probability

$r_i = E[x_i]$ ----firing rate $\qquad S = \{p(x_1, x_2, \ldots, x_n)\}$

$v_{ij} = Cov[x_i, x_j]$ ----covariance: correlation

higher-order correlations

orthogonal decomposition

Multiple spike sequence:

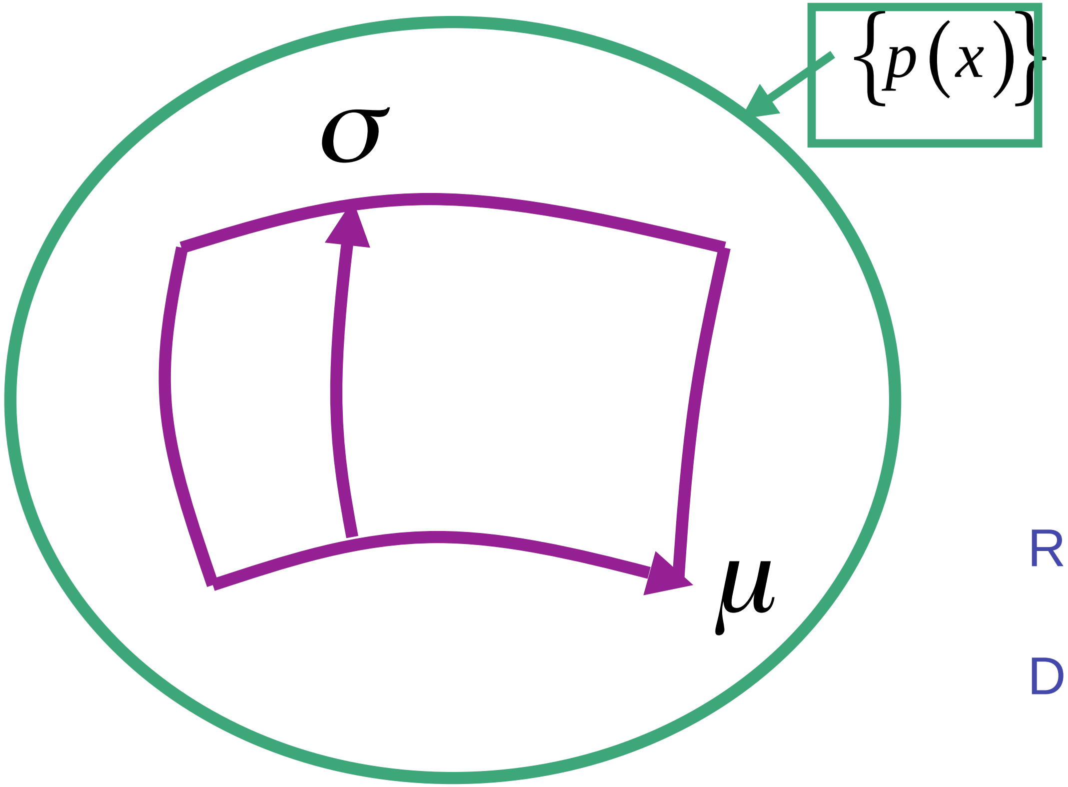
$$\{x_i(t),\quad i=1,\ldots,n;\ t=1,\ldots,T\}$$

$$x_i(t)=0,1$$

neuron i \ time t		1	⋯	T
1	●	$x_1(1)$		$x_1(T)$
⋮	●	$x_2(1)$		$x_2(T)$
	⋮	⋮		⋮
n	●	$x_n(1)$	⋯	$x_n(T)$

Information Geometry ?

$$S = \{p(x;\mu,\sigma)\} \qquad p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$



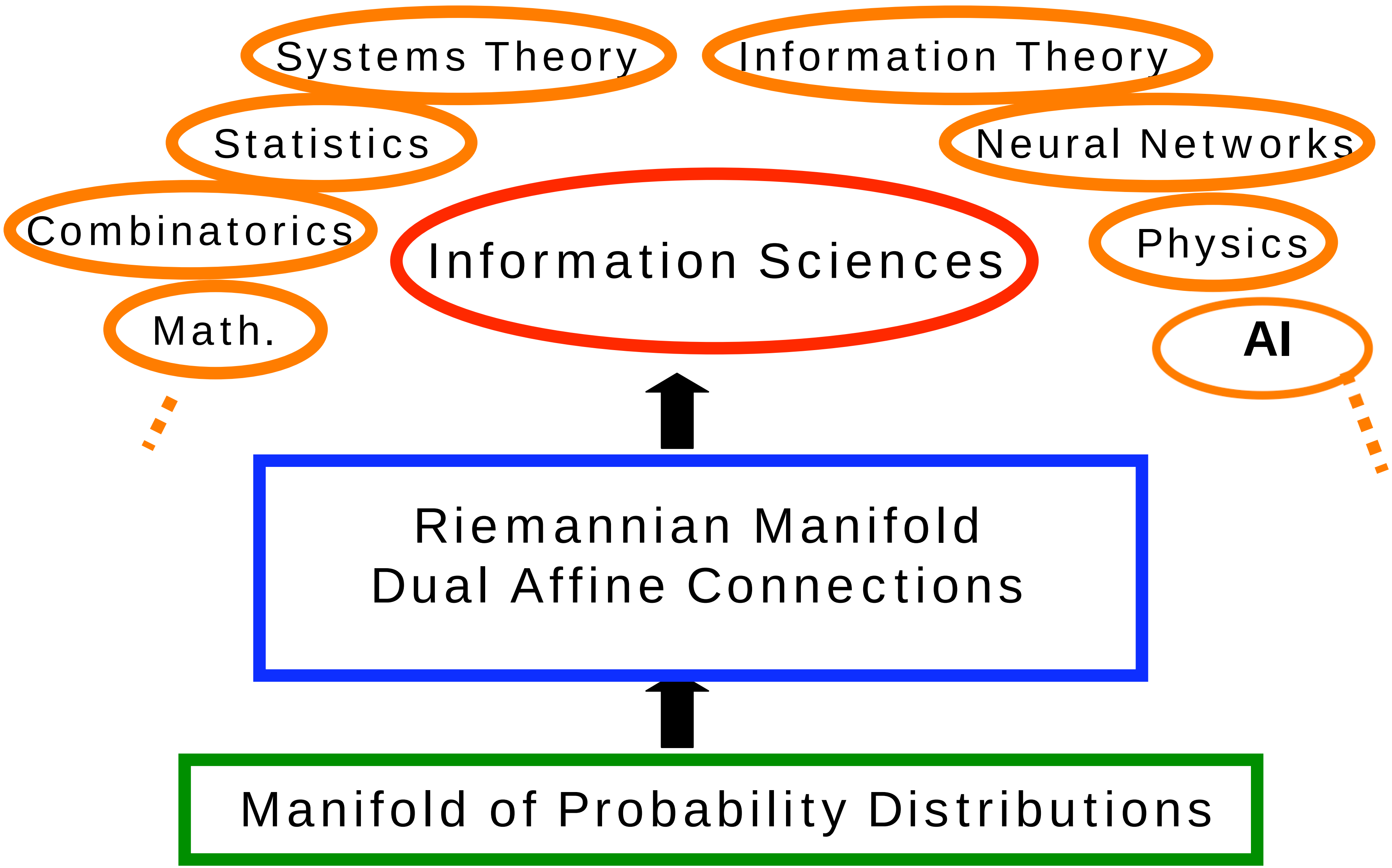
$$S = \{p(x;\text{è})\}$$

$$\text{è} = (\mu,\sigma)$$

Riemannian metric

Dual affine connections

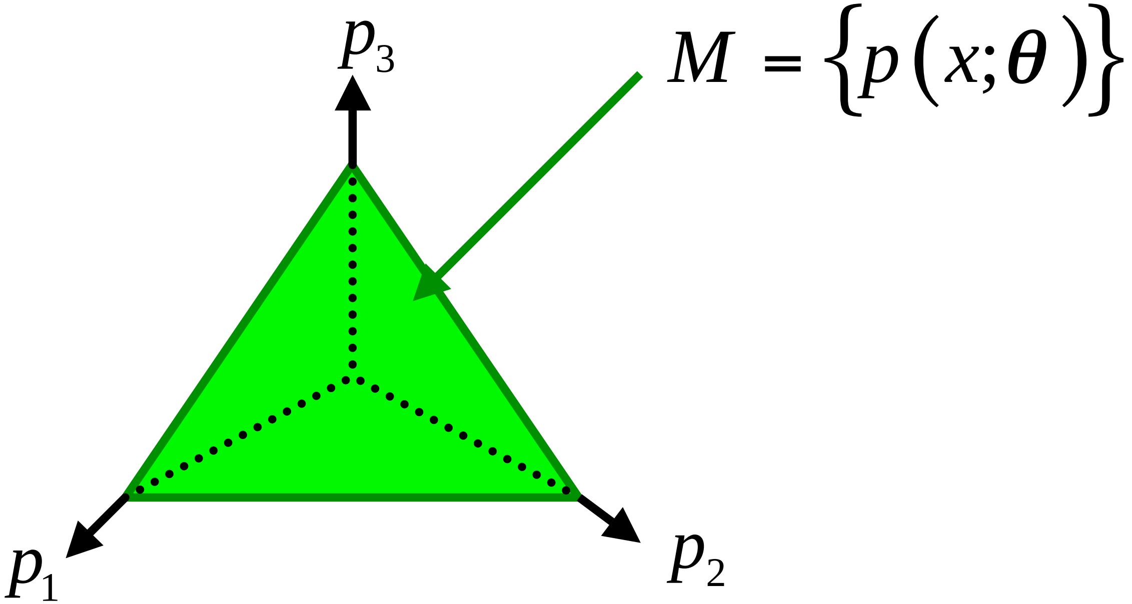
Information Geometry



Manifold of Probability Distributions

$$x = 1, 2, 3 \qquad \{p(x)\}$$

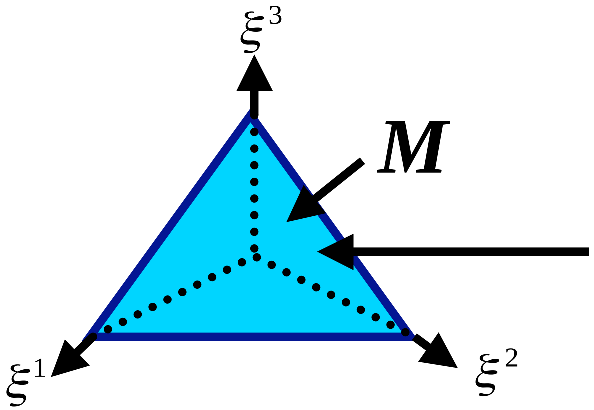
$$\boldsymbol{p} = \left(p_1, p_2, p_3\right) \qquad p_1 + p_2 + p_3 = 1$$



Manifold of Probability Distributions

$$\boldsymbol{M} = \{P(x, \boldsymbol{p})\}$$

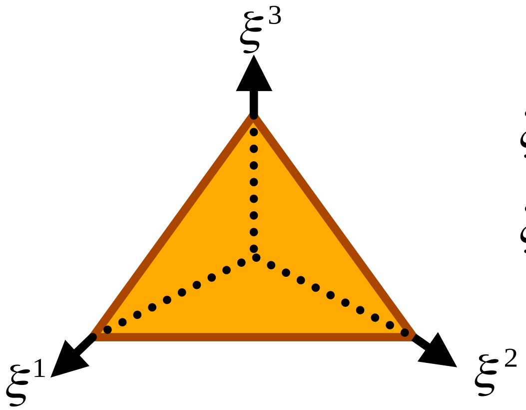
$$x = 1, 2, 3$$



$$P(x, \boldsymbol{p}) = \sum_{i=1}^{3} p_i \delta_i(x) \quad (p_3 = 1 - p_1 - p_2)$$

$$\boldsymbol{p} = (p_1, p_2, p_3)$$

$$p_1 + p_2 + p_3 = 1$$



$$\xi_i = \sqrt{p_i}$$

$$\xi_1^2 + \xi_2^2 + \xi_3^2 = 1$$

$$\begin{cases} \theta_1 = \log \dfrac{p_1}{p_3} \\ \theta_2 = \log \dfrac{p_2}{p_3} \end{cases}$$

$$\boldsymbol{\theta} = (\theta^1, \theta^2)$$

Invariance

$$S = \{p(x, \boldsymbol{\theta})\}$$

1. Invariant under reparameterization

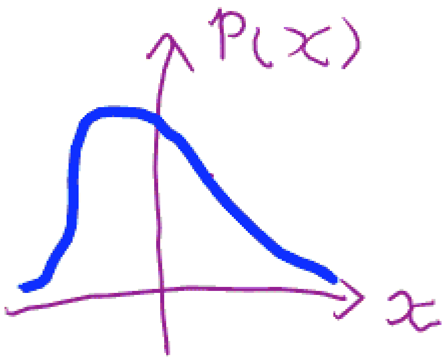
$$p(x,\theta) = \overline{p}(x,\xi) \qquad D = \sum \theta_i^2 \neq \sum \xi_i^2$$

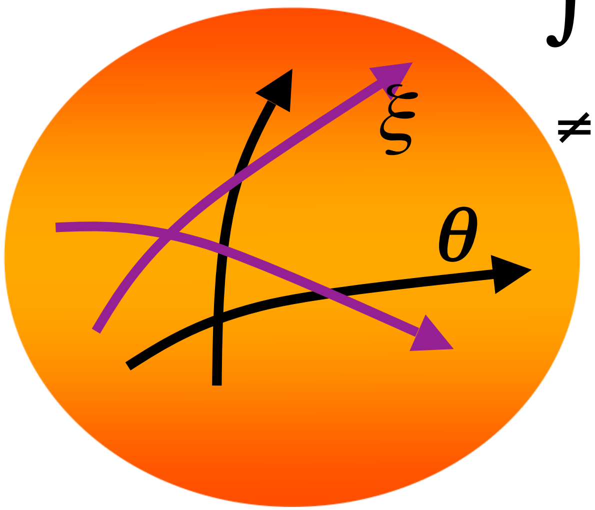
2. Invariant under different representation

$$y = y(x), \quad \{\overline{p}(y,\boldsymbol{\theta})\} = \{p(x,\theta)\}$$

$$\int \left| p(x,\theta_1) - p(x,\theta_2) \right|^2 dx$$

$$\neq \int \left| \overline{p}(y,\theta_1) - \overline{p}(y,\theta_2) \right|^2 dy$$





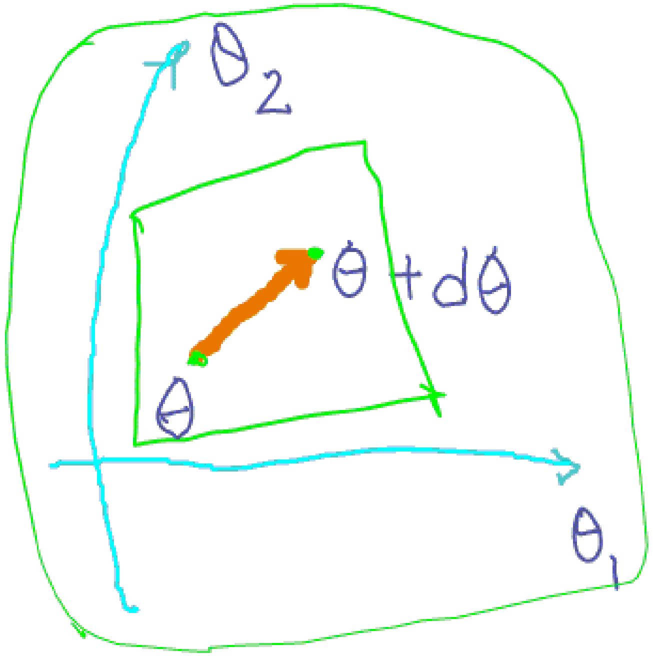
Two Structures

Riemannian metric

affine connection --- geodesic

$$g_{ij} = E\left[\frac{\partial}{\partial\theta_i}\log p\frac{\partial}{\partial\theta_j}\log p\right]$$

Fisher information



$$ds^2 = \sum g_{ij}(\boldsymbol{\theta})d\theta_i d\theta_j = < d\theta, d\theta >$$

Affine Connection

covariant derivative

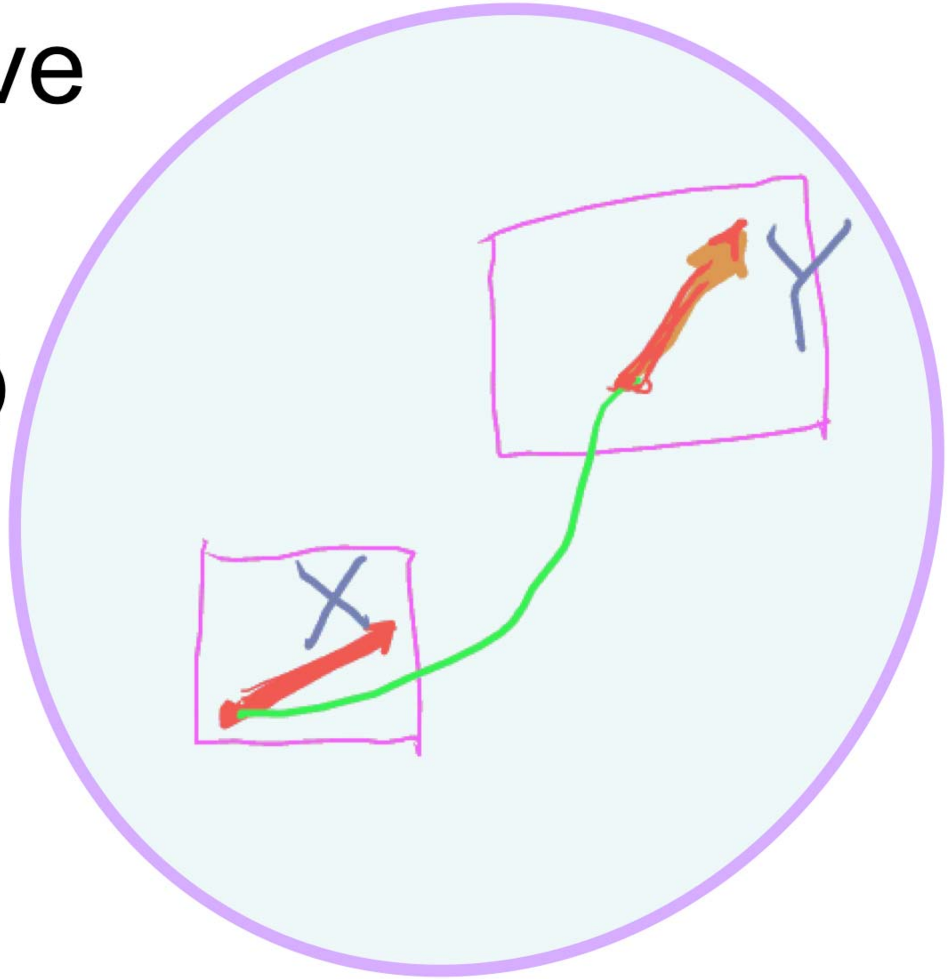
$\Pi_c X = Y$

geodesic $\Pi \dot{X} = \dot{X}$ $X = X(t)$

$$s = \int \sqrt{\sum g_{ij}(\theta) d\theta^i d\theta^j}$$

minimal distance

straight line



Affine Connection

covariant derivative; parallel transport

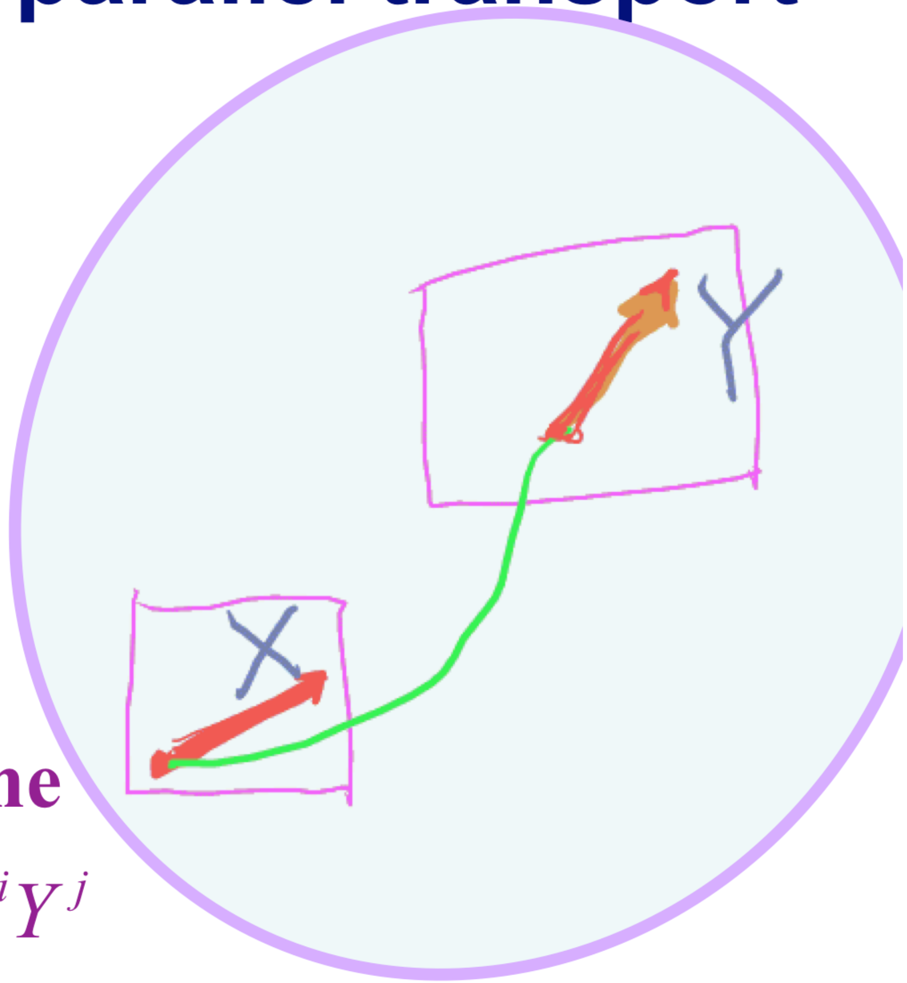
$$\nabla_X Y, \quad \Pi_c X = Y$$

geodesic $\quad \Pi\dot{x} = \dot{x} \quad x = x(t)$

$$s = \int \sqrt{\sum g_{ij}(\theta) d\theta^i d\theta^j}$$

minimal distance : straight line

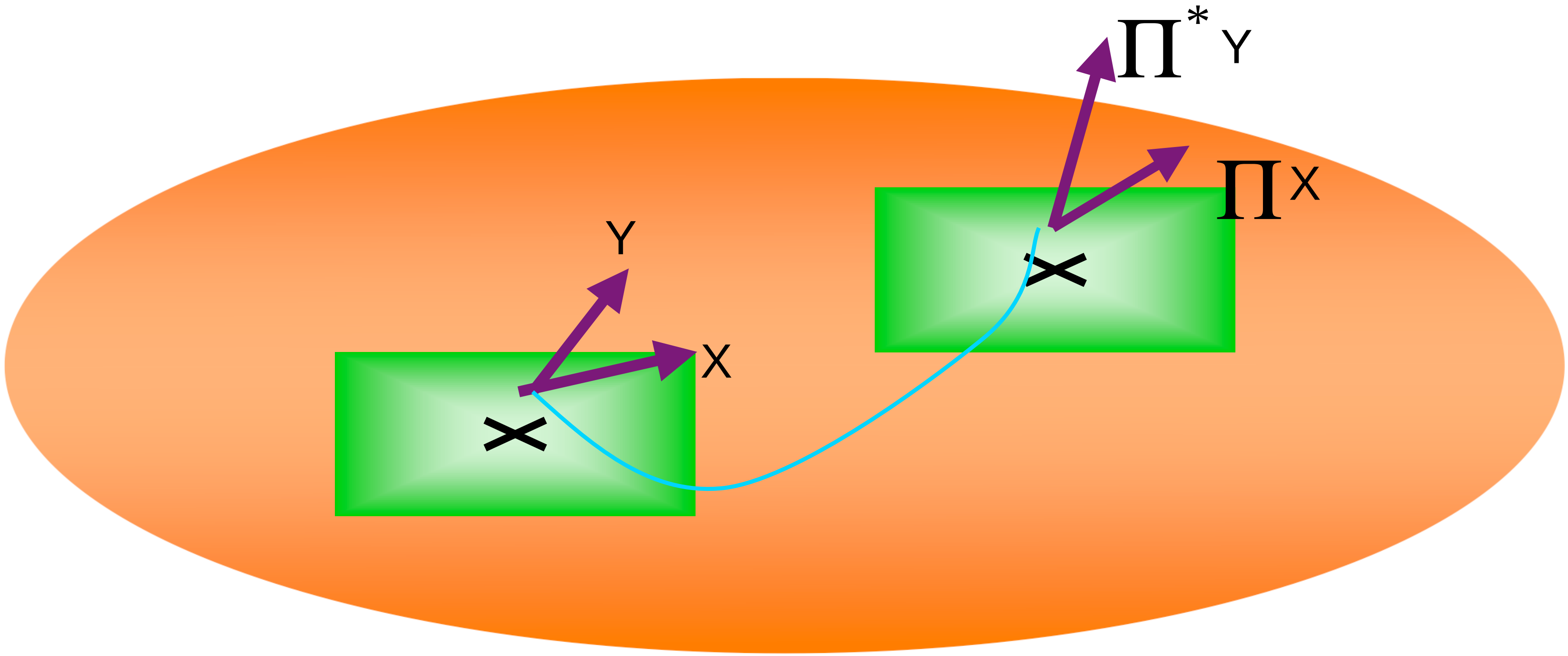
$$< \Pi X, \Pi Y > = < X, Y > \qquad g_{ij} X^i Y^j$$



Duality: two affine connections

$\{S, g, \nabla, \nabla^*\}$

$$\langle X, Y \rangle = \langle \Pi X, \Pi^* Y \rangle \qquad < X, Y > = \sum g_{ij} X^i Y^j$$



Riemannian geometry: $\Pi = \Pi^*$

Dual Affine Connections

$$(\nabla, \nabla^*)$$

$$(\Pi, \Pi^*)$$

e-geodesic

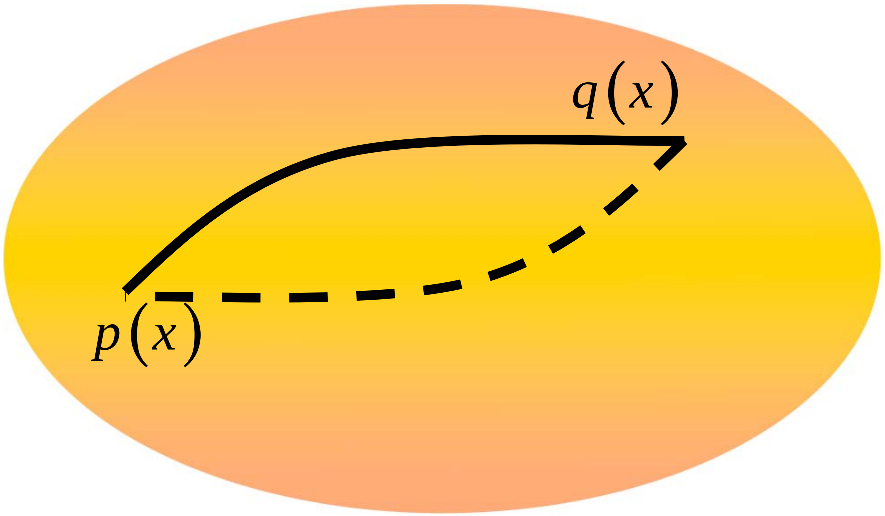
$$\log r(x,t) = t \log p(x) + (1-t) \lg q(x) + c(t)$$

m-geodesic

$$r(x,t) = tp(x) + (1-t)q(x)$$

$$\nabla_{\dot{x}} \dot{x}(t) = 0$$

$$\nabla^*_{\dot{x}} \dot{x}(t) = 0$$



Information Geometry

-- Dually Flat Manifold

Convex Analysis
Legendre transformation
Divergence
Pythagorean theorem
I-projection

Dually Flat Manifold

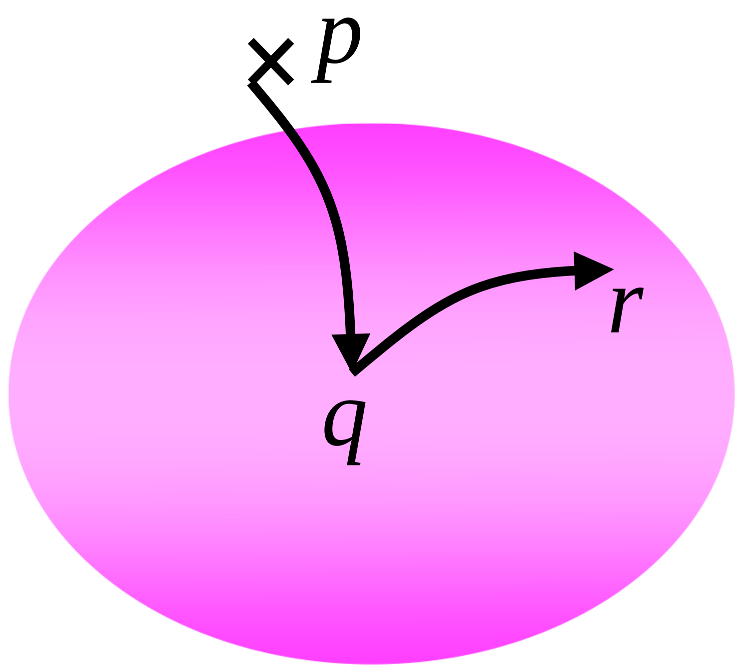
1. Potential Functions

---convex (Bregman, Legendre transformation)

2. Divergence $D[p:q]$

3. Pythagoras Theorem

$$D[p:q]+D[q:r]=D[p:r]$$

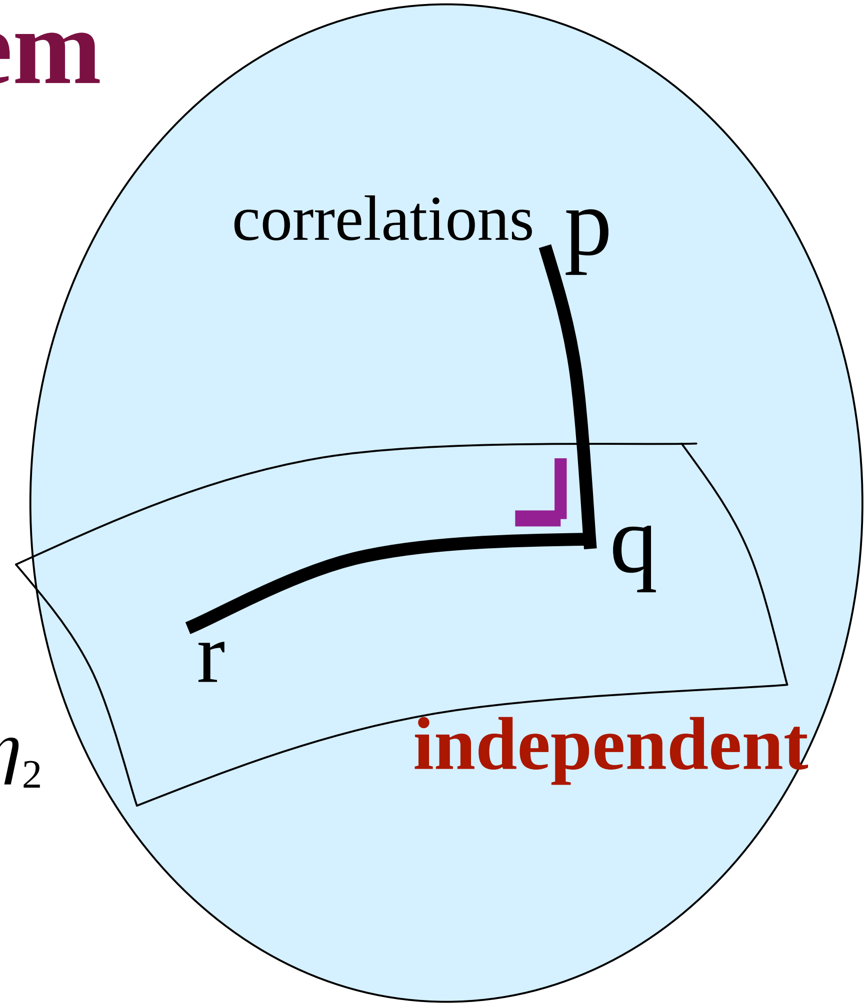


4. Projection Theorem

5.Dual foliation

Pythagoras Theorem

D[p:r] = D[p:q]+D[q:r]



p,q: same marginals η_1, η_2

r,q: same correlations θ

$$D[p:r] = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

estimation correlation testing

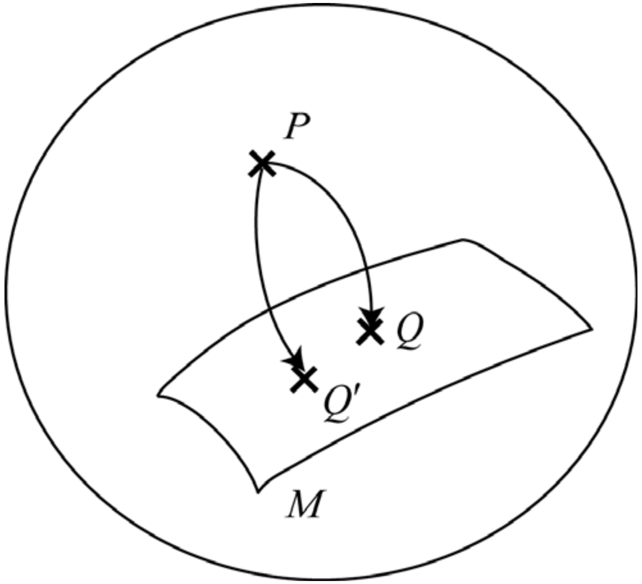
invariant under firing rates

Projection Theorem

$$\min_{Q \in M} D[P : Q]$$

$Q =$ m-geodesic

projection of P to M



$$\min_{Q \in M} D[Q : P]$$

$Q =$ e-geodesic projection of P to M

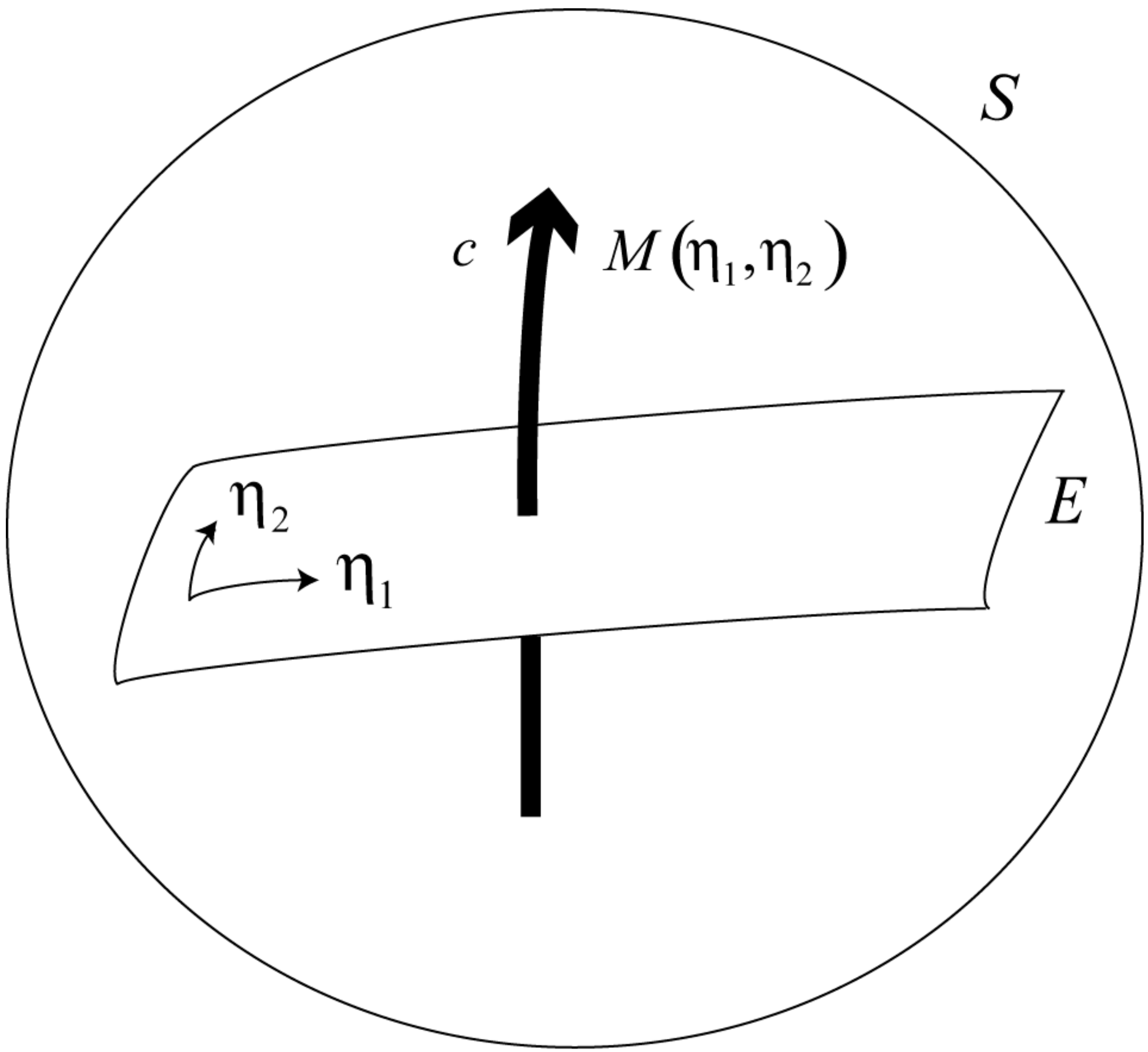


Fig. 1a

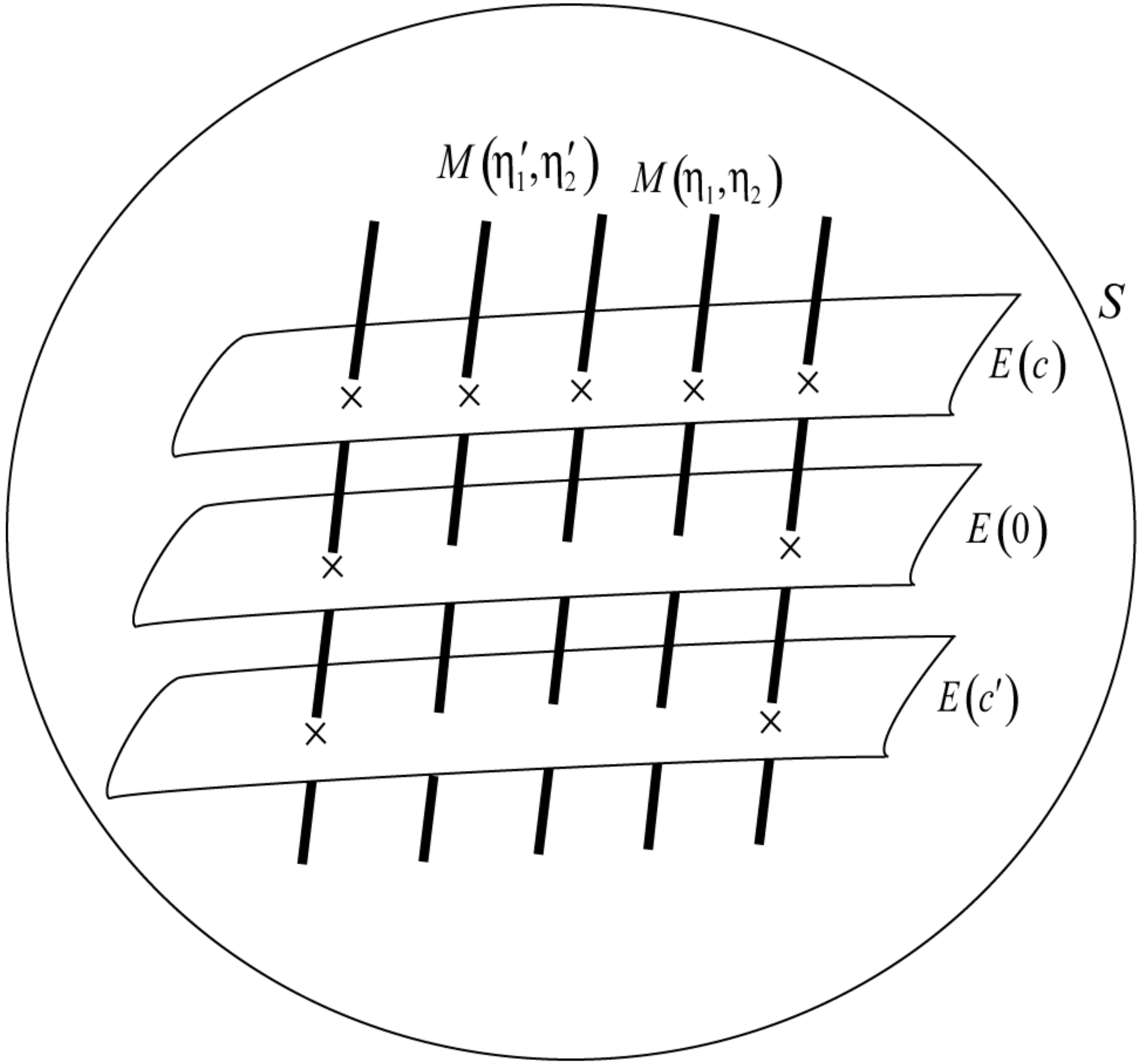


Fig. 1b

Multiple spike sequence:

$$\{x_i(t),\quad i=1,\ldots,n;\ t=1,\ldots,T\}$$

$$x_i(t)=0,1$$

neuron i \ time t	1	………………………	T
1	$x_1(1)$		$x_1(T)$
⋮	$x_2(1)$		$x_2(T)$
	⋮		⋮
n	$x_n(1)$	……….	$x_n(T)$

spatial correlations

$$\boldsymbol{x} = \left(x_1, \cdots, x_n\right)$$

$$p\left(\boldsymbol{x}\right) = \exp\left\{\sum \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j + \cdots\cdots + \theta_{1\cdots n} x_1 \cdots x_n - \psi\left(\boldsymbol{\theta}\right)\right\}$$

$$r_i = E\left[x_i\right] = \text{Prob}\left\{x_i = 1\right\}$$

$$r_{ij} = E\left[x_i x_j\right] = \text{Prob}\left\{x_i = x_j = 1\right\}$$

$$\vdots$$

$$\boldsymbol{\theta} = \left(\theta_i, \theta_{ij}, \cdots, \theta_{1\cdots n}\right)$$

$$\boldsymbol{r} = \left(r_i, r_{ij}, \cdots, r_{1\cdots n}\right)$$

$S = \left\{p\left(\boldsymbol{x}\right)\right\}$: coordinates

orthogonal structure

Spatio-temporal correlations

correlated Poisson

$p(\boldsymbol{x})$: temporally independent

correlated renewal process

$p(\boldsymbol{x}_t)$: firing rate $r_i(t)$ modified

spatial correlations fixed

Two neurons: $\{p_{00}, p_{01}, p_{10}, p_{11}\}$

x_1 0011000101101

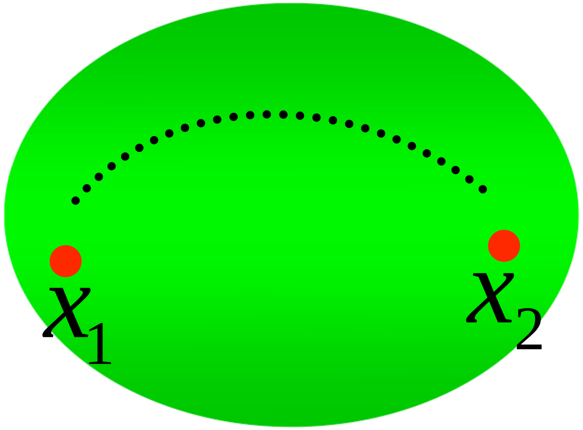
x_2 0100100110100

x_3 0101101001010

firing rates: $r_1, r_2; r_{12}$

correlation—covariance?

Correlations of Neural Firing



$$\{p(x_1, x_2)\}$$

$$\{p_{00}, p_{10}, p_{01}, p_{11}\}$$

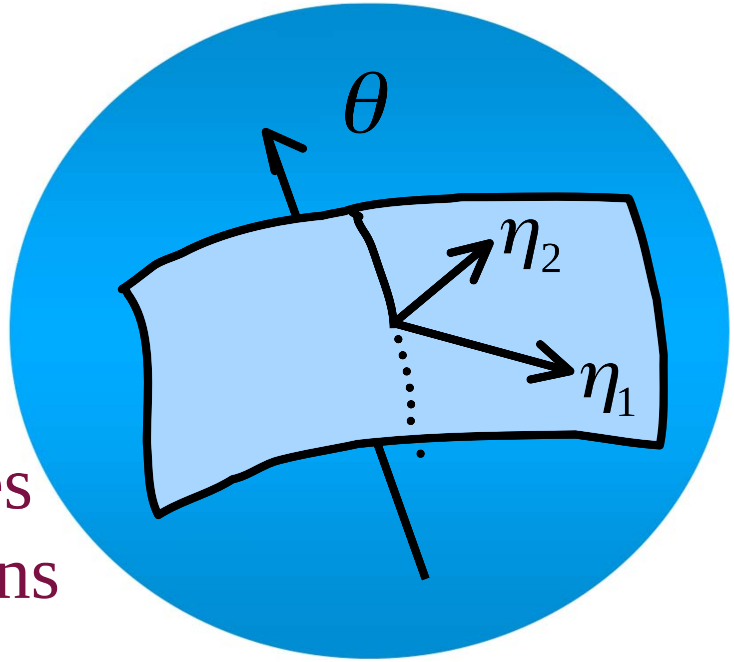
$$r_1 = p_{1\cdot} = p_{10} + p_{11}$$

$$r_2 = p_{\cdot 1} = p_{01} + p_{11}$$

$$\theta = \log \frac{p_{11} p_{00}}{p_{10} p_{01}}$$

firing rates

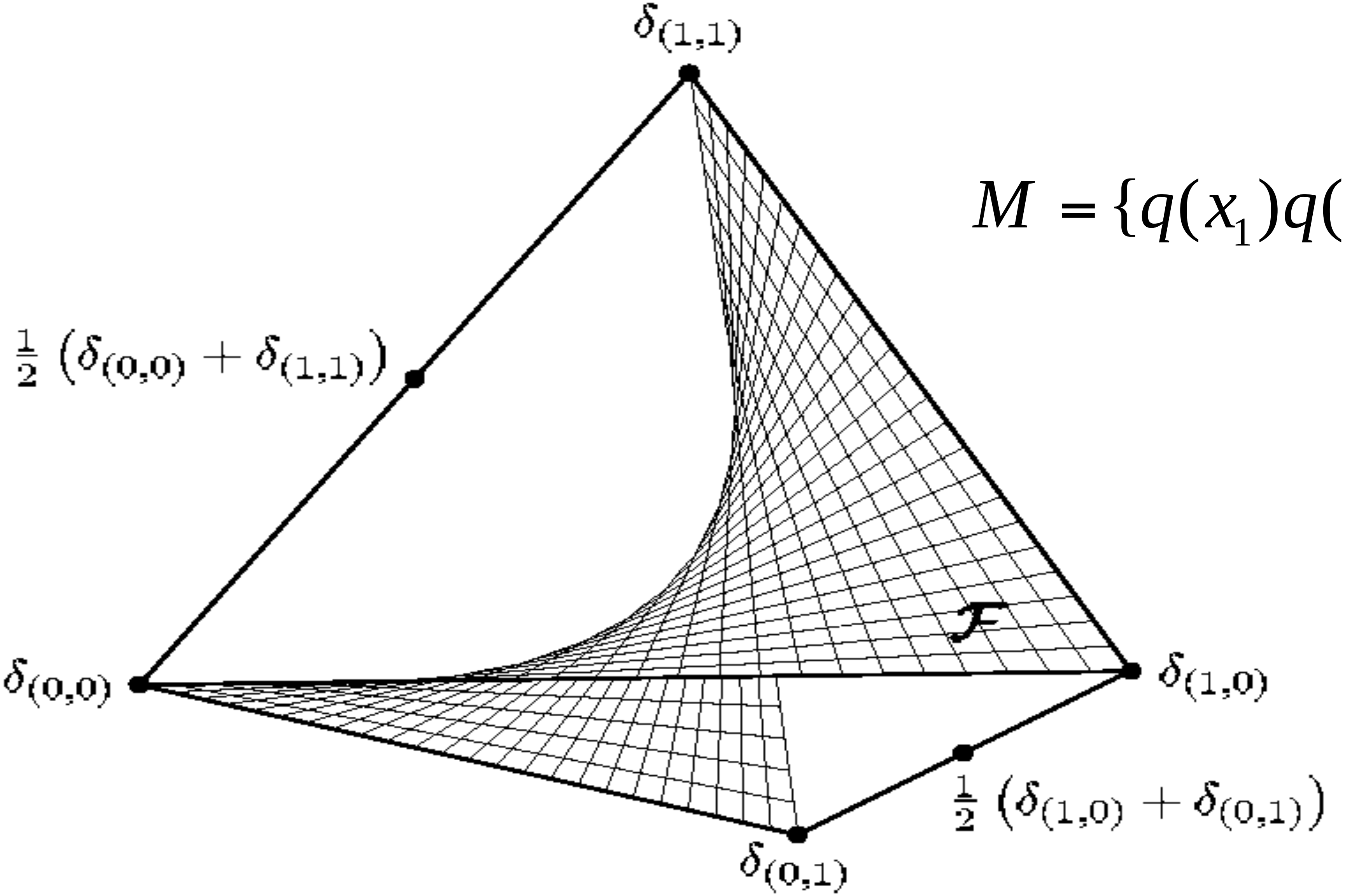
correlations



$$\{(r_1, r_2), \theta\}$$

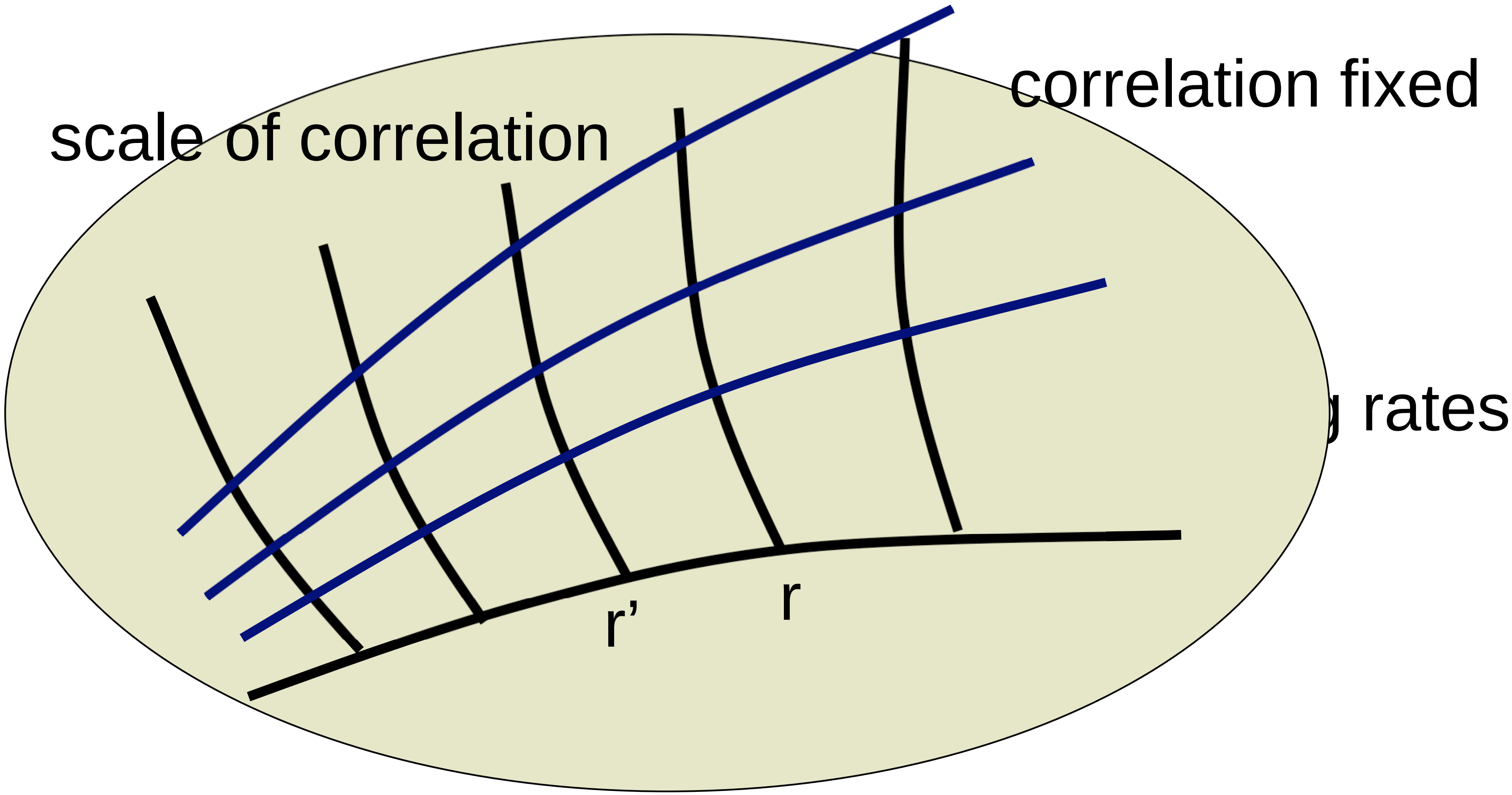
orthogonal coordinates

Independent Distributions



$$M = \{q(x_1)q(x_2)\}$$

Orthogonal Coordinates:



two neuron case

$$r_1, r_2, r_{12}; \quad \theta_1, \theta_2, \theta_{12}$$

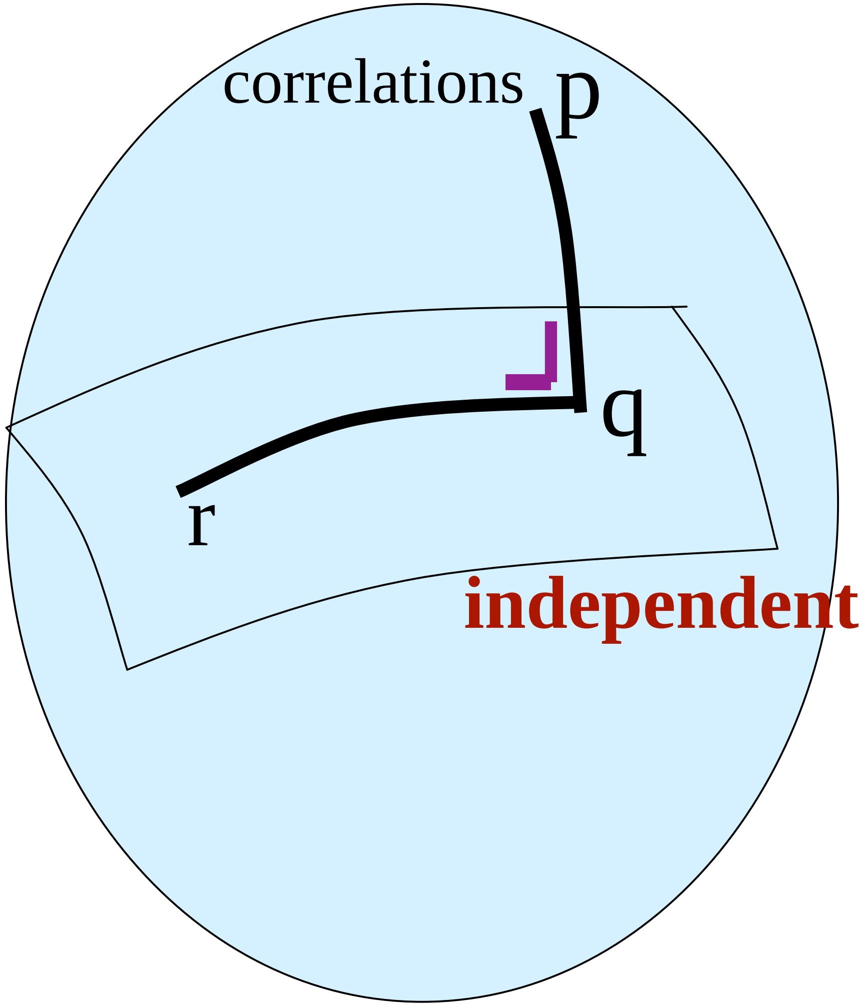
$$\theta_{12} = \log \frac{p_{00} p_{11}}{p_{01} p_{10}} = \log \frac{r_{12}\left(1 + r_{12} - r_1 - r_2\right)}{\left(r_1 - r_{12}\right)\left(r_2 - r_{12}\right)}$$

$$r_{12} = f\left(r_1, r_2, \theta\right)$$

$$r_{12}\left(t\right) = f\left(r_1\left(t\right), r_2\left(t\right), \theta\right)$$

Decomposition of KL-divergence

D[p:r] = D[p:q]+D[q:r]



p,q: same marginals η_1, η_2

r,q: same correlations θ

$$D[p:r] = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

pairwise correlations

covariance: $c_{ij} = r_{ij} - r_i r_j$ **not orthogonal**

independent distributions

$$r_{ij} = r_i r_j, \qquad r_{ijk} = r_i r_j r_k, \ldots$$

How to generate correlated spikes?
(Niebur, Neural Computation [2007])

higher-order correlations

Orthogonal higher-order correlations

$$\boldsymbol{\theta} = \left(\theta_i, \theta_{ij}; \ldots, \theta_{1\ldots n}\right)$$

$$\boldsymbol{r} = \left(r_i, r_{ij}; \ldots, r_{1\ldots n}\right)$$

tractable models of distributions $M = \{p(\mathbf{x})\}$

Full model: $2^n - 1$ parameters

Homogeneous model: n parameters

Boltzmann machine: only pairwise
correlations

Mixture model

Models of Correlated Spike Distributions (1)

Reference	$y(t)$	0110001011·····
	$\tilde{x}_1(t)$	1001011001·····
	$\tilde{x}_2(t)$	0101101101·····
	$\tilde{x}_3(t)$	·····

Models of Correlated Spike Distributions (2)

$$x_i(t) = \tilde{x}_i(t) \circ y(t)$$

additive interaction

eliminating interaction

Niebur replacement

$y(t)$	0110001011 …..
$\tilde{x}_1(t)$	1001001101 …..
$\tilde{x}_2(t)$	1100010011

Mixture Model

$y(t)$: 0110001011 mixing sequence

$$p_1(\boldsymbol{x};\boldsymbol{u}) = \prod p(x_i, u_i) \quad \text{when } y = 1$$

$$p_2(\boldsymbol{x};\boldsymbol{u}) = \prod p(x_i, v_i) \quad \text{when } y = 0$$

$$p(x, u) : x = 1 \ \text{ with probability } u$$

$$p(\mathbf{x}) = \sum p(y)\, p(\mathbf{x} \mid y)$$

Mixture model

$$p(\boldsymbol{x};\boldsymbol{u}) = \prod p(x_i, u_i)$$: independent distribution

$$p(\boldsymbol{x};\boldsymbol{u},\boldsymbol{v},m) = mp(\boldsymbol{x};\boldsymbol{u}) + \bar{m}p(\boldsymbol{x};\boldsymbol{v})$$

$$\boldsymbol{u} = (u_1, \cdots, u_n) \qquad \bar{m} = 1 - m$$

$$\boldsymbol{v} = (v_1, \cdots, v_n)$$

$$M_n = \{p(\boldsymbol{x};\boldsymbol{u},\boldsymbol{v},m)\}, \; S_n = \{p(\boldsymbol{x})\}$$

firing rates

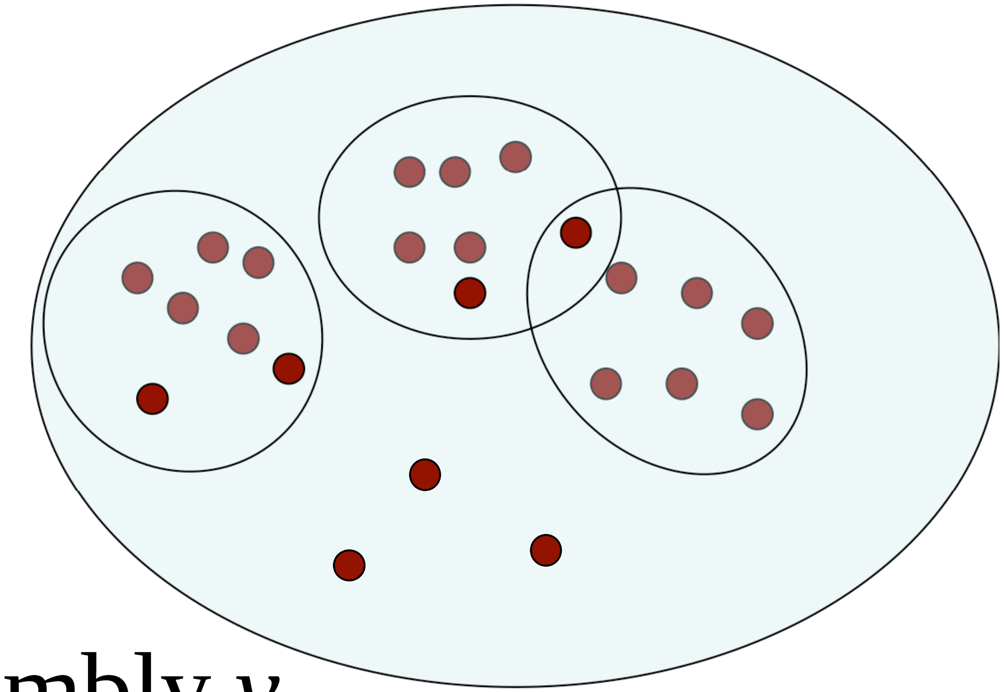
$$r_i = mu_i + \bar{m}v_i$$

$$r_{ij} = mu_iu_j + \bar{m}v_iv_j$$

$$r_{ijk} = mu_iu_ju_k + \bar{m}v_iv_jv_k$$

L

Conditional distributions and mixture



$$y = 1, 2, 3, \ldots$$

$$p(\mathbf{x} \mid y): \quad \text{Hebb assembly } y$$

$$p(\mathbf{x}) = \sum p(y)\, p(\mathbf{x} \mid y)$$

State transition of Hebb assemblies

mixture and covariances

$$p(\boldsymbol{x},t)=(1-t)p_u(\boldsymbol{x})+tp_v(\boldsymbol{x})$$

$$c_{ij}(t)=(1-t)c_{ij}(u)+tc_{ij}(v)+t(1-t)(u_i-v_i)(u_j-v_j)$$

Extra increase (decrease) of covariances

within Hebb assemblis---- increase
between Hebb assemblies ---decrease

Temporal correlations

$\{x(t)\}$, t =1, 2, …T

$$\mathrm{Prob}\{x(1),\ldots,x(T)\}$$

Poisson sequence:

Markov chain

Renewal process

Orthogonal parameter of correlation ---- Markov case

$$p(x_{t+1} \mid x_t)$$

$$a_{ij} = \Pr\{x_{t+1} = i \mid x_t = j\}$$

$$\Pr\{x_t\} = \Pr(x_1)\Pi p(x_{t+1} \mid x_t)$$

$$r_1 = \Pr\{x_t = 1\}$$

$$c = r_{11} - r_1^2$$

$$\theta = \log \frac{a_{11}a_{00}}{a_{10}a_{01}}$$

Spatio-temporal correlation

correlated Poisson

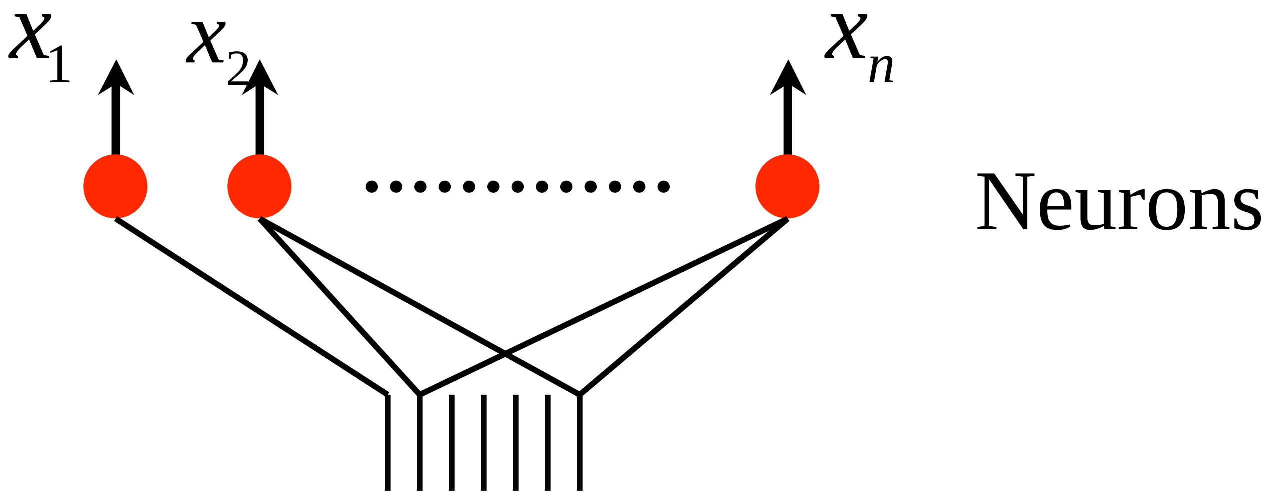
$$p(\{\mathbf{x}_t\}) = \prod_{t=1}^{T} p(\mathbf{x}_t)$$

Correlated renewal process

$$\Pr(x_t) = k(t - t') : \quad t' \text{ last spike time}$$

$$\Pr(x_{i,t}) = k(t - t') p(x_i \mid \mathbf{x})$$

Population and Synfire



$$x_i = 1(u_i)$$

$$u_i = \text{Gaussian} \qquad E[u_i u_j] = \alpha$$

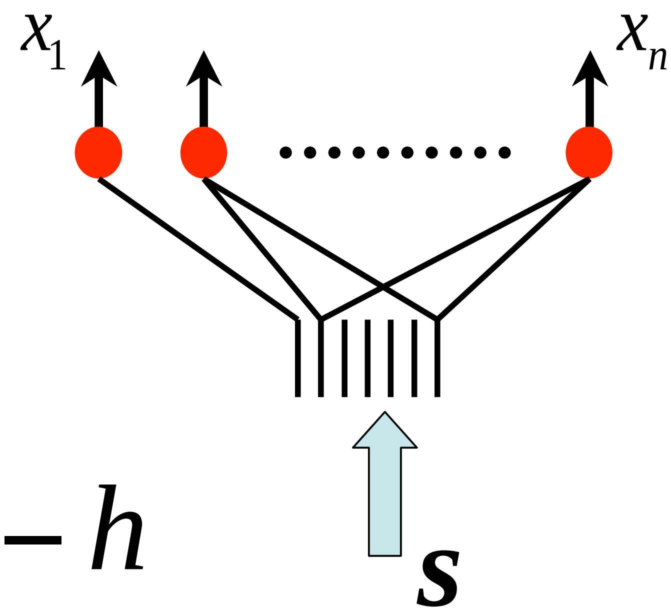
Population and Synfire

$$u_i = \sum w_{ij} s_j - h$$

$$x_i = 1[u_i]$$

$$u_i = \sqrt{(1-\alpha)}\delta_i + \sqrt{\alpha}\varepsilon - h$$

$$\delta_i,\ \varepsilon \approx N(0,\ 1)$$



$$E[u_i u_j] = \alpha$$

$$E[u_i^2] = 1$$

$$p_i = \text{Prob}\,\{i \text{ neurons fire at the same time}\}$$

$$r = \frac{i}{n} \qquad\qquad P_r = \Pr\{nr \text{ neurons fire}\}$$

$$q(r,\alpha) = e^{nH(r)} \int e^{-nz(\varepsilon)} d\varepsilon$$

$$z(\varepsilon) = \frac{\varepsilon^2}{2n} - r \log F - (1-r)\log(1-F)$$

$$F = F(a\varepsilon - h) = \frac{1}{\sqrt{2\pi}} \int_0^{a\varepsilon - h} e^{-\frac{t^2}{2}}\,\mathrm{dt}$$

$$q(r,\alpha) = c\exp[\frac{2\alpha - 1}{2(1-\alpha)}\{F^{-1}(\alpha) - \frac{\sqrt{\alpha}}{2\alpha - 1}h\}^2]$$

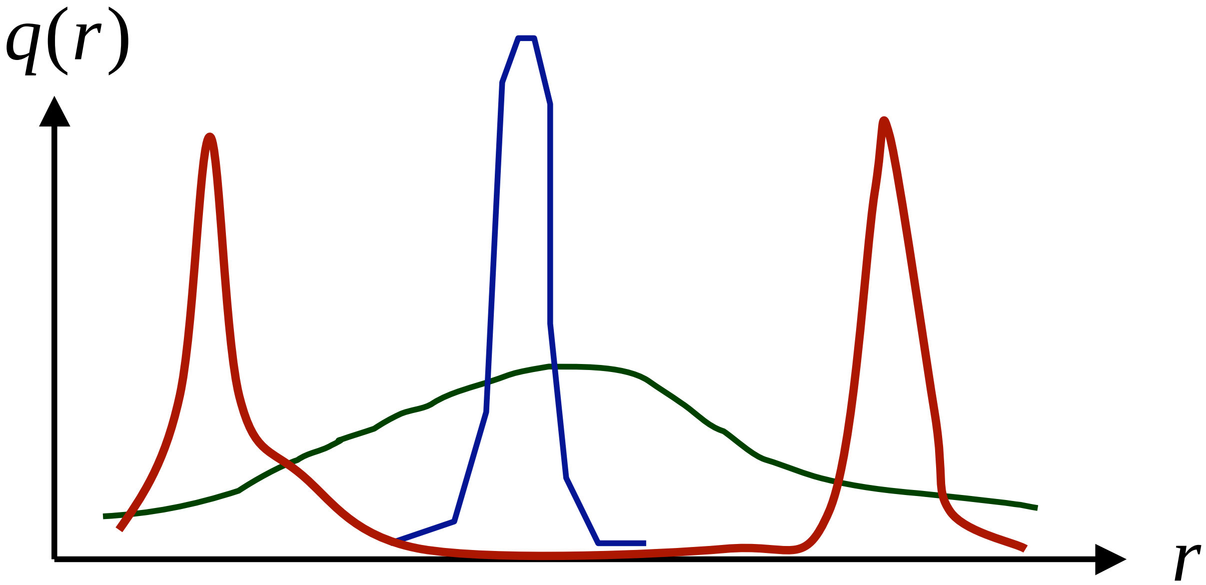
$$p(\mathbf{x},\theta) = \exp\{\sum \theta_i x_i + \sum \theta_{ij} x_i x_j + \sum \theta_{ijk} x_i x_j x_k + \ldots\}$$

$$\theta_{i_1 i_2 \ldots i_k} = O(1/n^{k-1})$$

Synfiring

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_n)$$

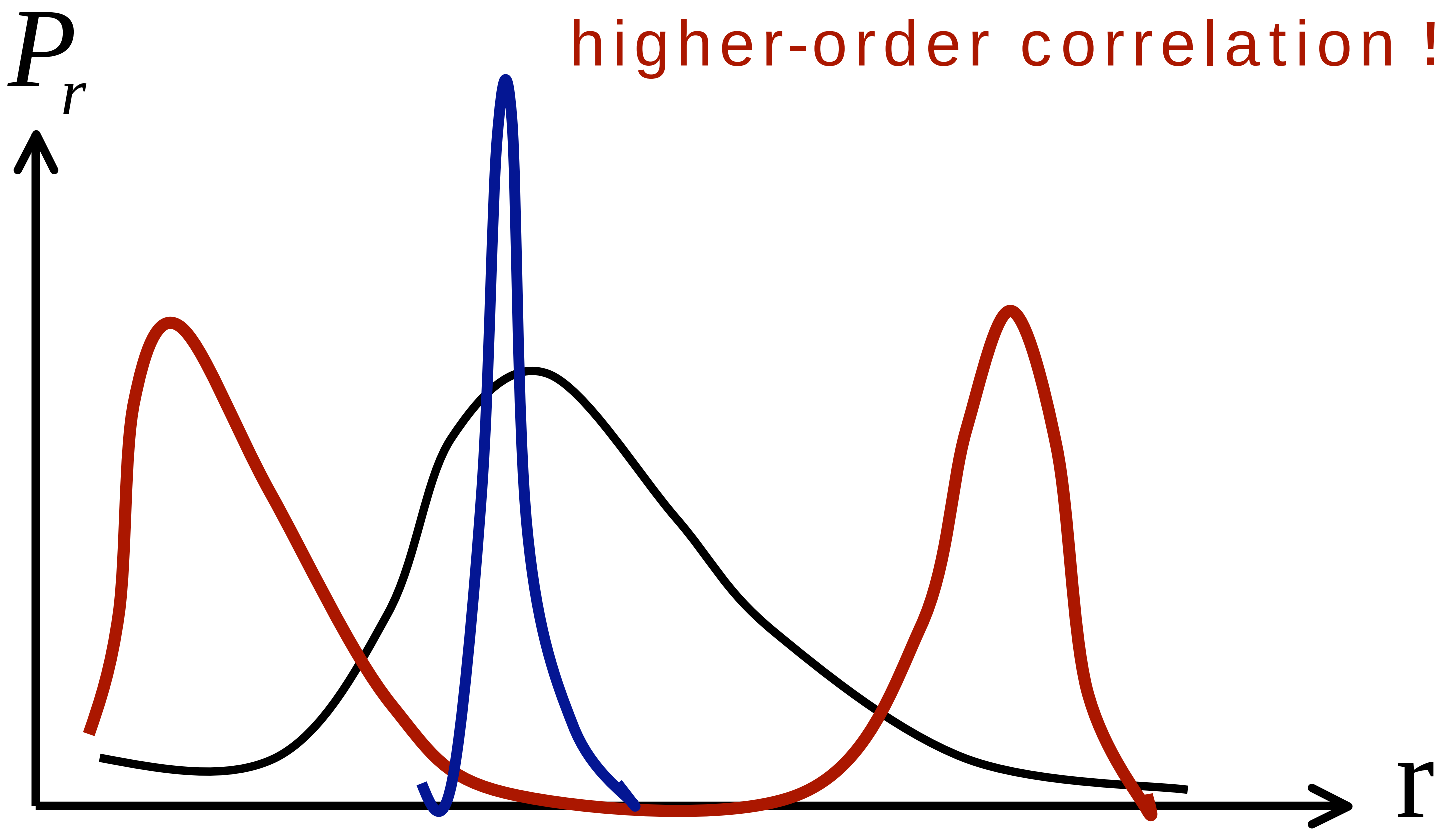
$$r = \frac{1}{n}\sum x_i \qquad\qquad q(r)$$



Bifurcation

x_i : independent---single delta peak
pairwise correlated

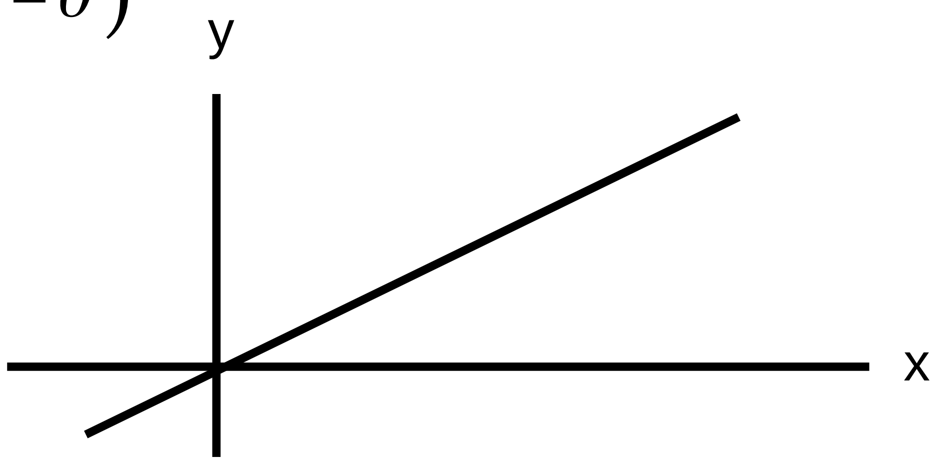
higher-order correlation !



Semiparametric Statistics

$$M = \{p(x, \theta, r)\}$$

$$p(x, \theta) = r(x - \theta)$$



$$y = \theta x$$

$$\begin{cases} y_i = \theta \xi_i + \varepsilon_i \\ x_i = \xi_i + \varepsilon_i^{'} \end{cases}$$

$$p(x, y; \theta) = \int p(x, y; \xi, \theta) r(\xi) d\xi$$

mle, least squre, total least square

Linear Regression: Semiparametrics

$$\left(x_1, y_1\right)$$

$$\left(x_2, y_2\right)$$

$$\vdots$$

$$\left(x_n, y_n\right)$$

$$y = \theta x$$

$$x_i = \xi_i + \varepsilon_i$$

$$y_i = \theta \xi_i + \varepsilon_i^{'}$$

$$\varepsilon_i, \varepsilon_i^{'} \quad N\left(0, \sigma^2\right)$$

Statistical Model

$$p\left(x, y \middle| \theta, \xi\right) = c \exp\left\{-\frac{1}{2}\left(x-\xi\right)^2 - \frac{1}{2}\left(y-\theta\xi\right)^2\right\}$$

$$p\left(x_i, y_i \middle| \theta, \xi_i\right) \; : \; \theta, \xi_1, \cdots, \xi_n$$

$$p\left(x, y \middle| \theta\right) = \int p\left(x, y \middle| \theta, \xi\right) Z\left(\xi\right) d\xi$$

——— semiparametric

Least squares?

$$L(\theta) = \sum (y_i - \theta x_i)^2 \to \min \qquad : \hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\frac{1}{n}\sum \frac{y_i}{x_i}, \qquad \frac{\sum y_i}{\sum x_i}$$

mle, TLS

$$\sum (y_i - \theta x_i)(\theta y_i + x_i) = 0$$

Neyman-Scott

Semiparametric statistical model

$$x_1, x_2, \cdots \qquad p(x, \theta, Z)$$

Estimating function

$$y(x, \theta) \qquad\qquad E_{\theta, Z}\left[y(x, \theta)\right] = 0$$

$$E_{\theta', Z}\left[y(x, \theta)\right] \neq 0$$

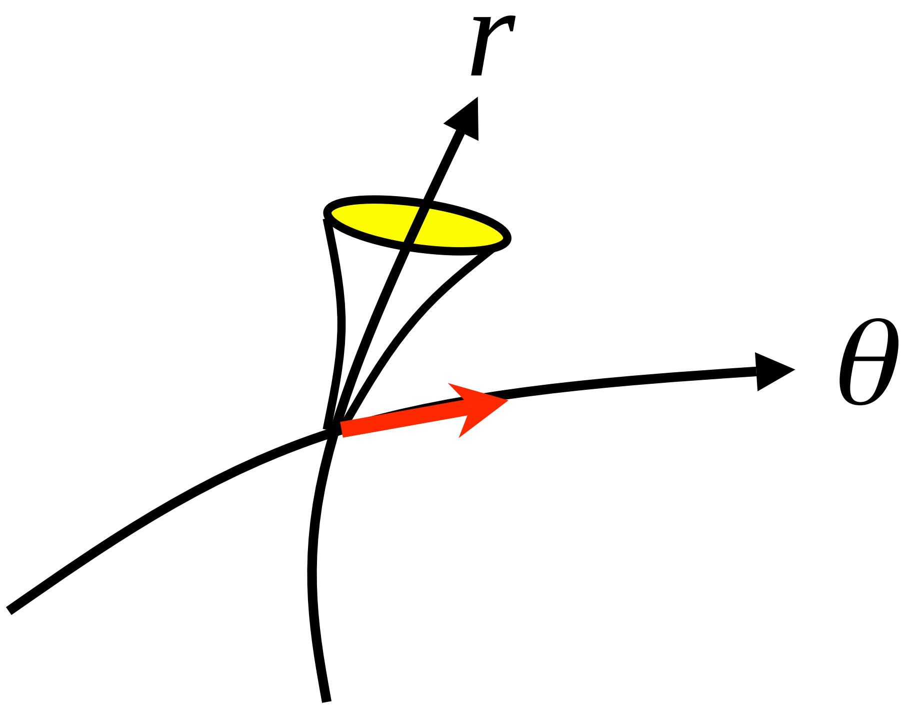
Estimating equation

$$\sum y(x_i, \theta) = 0 \qquad \Rightarrow \hat{\theta}$$

$$
\begin{aligned}
u\left(x, y; \theta, Z\right) &= \frac{\partial}{\partial\theta} \log p \\
&= \partial_{\theta} s E\left[\xi \middle| s\right] \cdots \\
v\left(x, y; \theta, Z\right) &= E\left[f\left(\xi\right) \middle| s\right] \\
&= k\left\{s\left(x, y; \theta\right)\right\} \\
u^{I}\left(x, y, \theta\right) &= u - E\left[u \middle| s\right] \\
&= k\left(x + \theta y\right)\left(y - \theta x\right)
\end{aligned}
$$

Fibre bundre

function space



$$z(\xi) \quad N\left(\mu_\xi, \sigma_\xi^2\right)$$

$$u^I(x, \theta, Z) = (x + \theta y + c)(y - \theta x)$$

$$c = \mu_\xi / \sigma_\xi^2$$

$$f(x, \theta; c) = (x + \theta y + c)(y - \theta x)$$

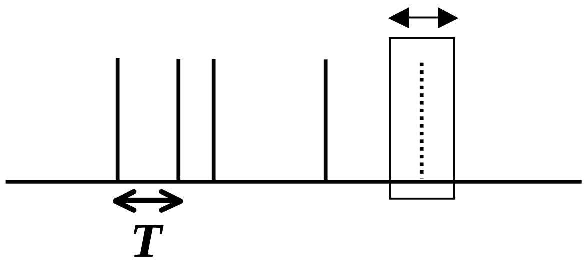
$$c = \mu_\xi / \sigma_\xi^2$$

$$\mu_\xi = \frac{1}{n} \sum x_i$$

$$\sigma_\xi^2 = \frac{1}{n} \sum x_i^2 - \left(\mu_\xi\right)^2 - \sigma^2$$

Poisson process

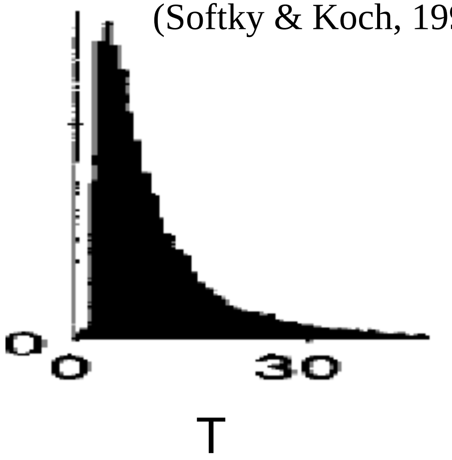
Poisson Process: Instantaneous firing rate is constant over time.



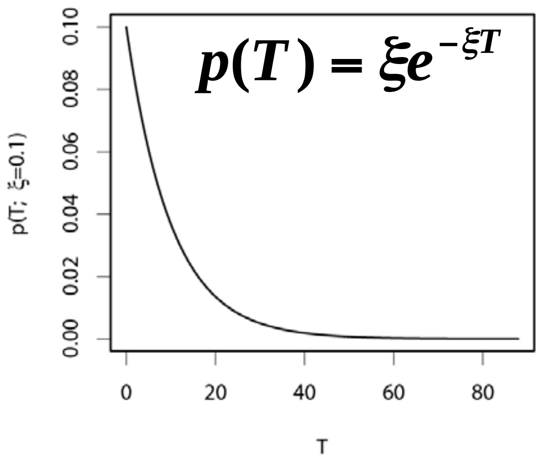
For every small time window dt, generate a spike with probability ξdt.

Cortical Neuron

(Softky & Koch, 1993)



Poisson Process



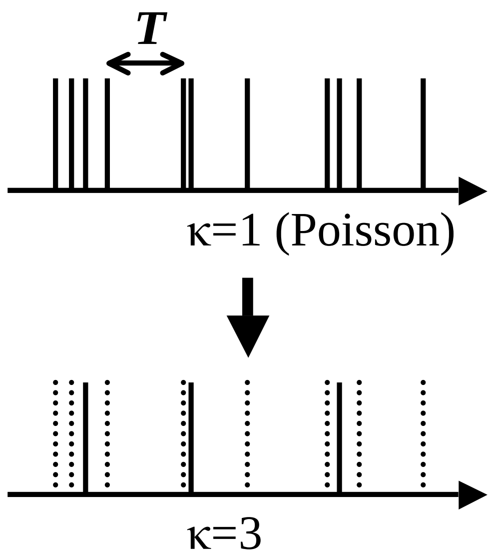
<u>Poisson process cannot explain inter-spike interval distributions.</u>

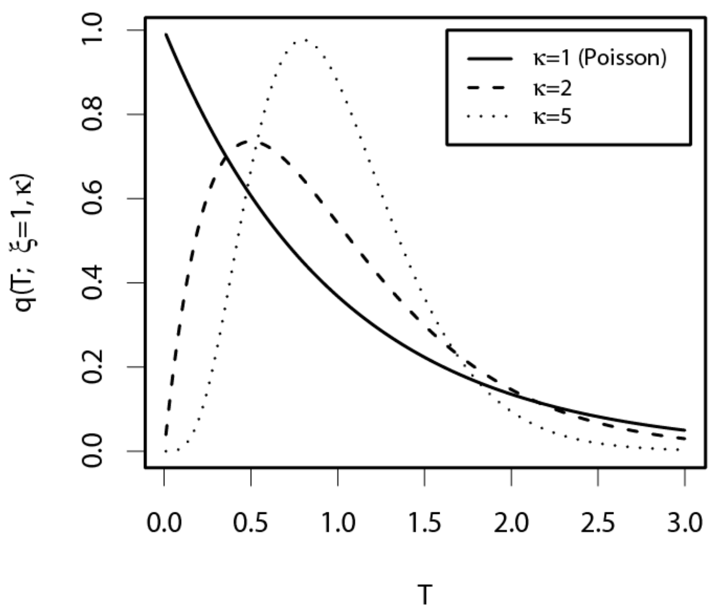
Gamma distribution

Gamma Distribution: Every κ-th spike of the Poisson process is left.

$$\mathbf{q(T;\xi,\kappa) = \frac{(\xi\kappa)^{\kappa}}{\Gamma(\kappa)} T^{\kappa-1} e^{-\xi\kappa T}.}$$

Two parameters
- ξ: **Firing rate**
- κ: **Irregularity**





Gamma distribution

$$f(T) = \frac{(r\kappa)^{\kappa}}{\Gamma(\kappa)} T^{\kappa-1} \exp\{-r\kappa T\}$$

$\kappa = 1$: Poisson

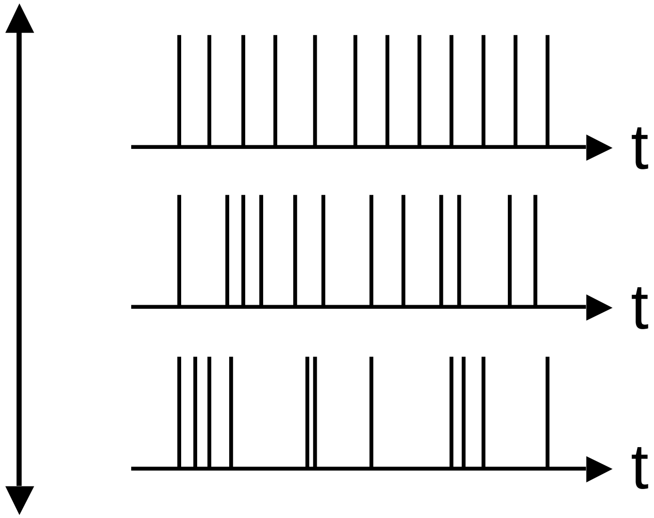
$\kappa \to \infty$: regular

Integrate-and fire

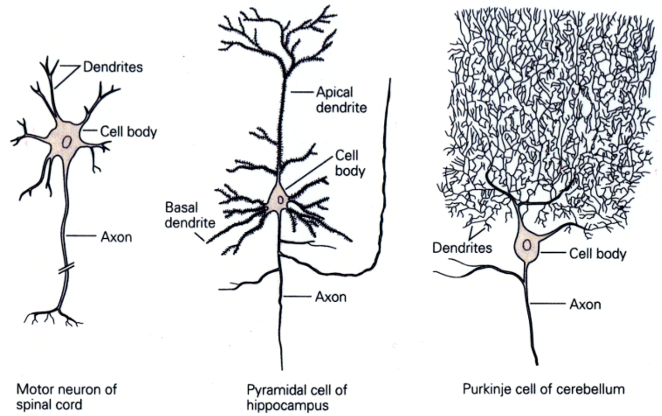
Markov model

Irregularity κ is unique to individual neurons.

Regular（largeκ）



Irregular（smallκ）



Irregularity varies among neurons.

→ **<u>We assume that κ is independent of time.</u>**

Information geometry of estimating function

•**Estimating function $y(T,\kappa)$:**

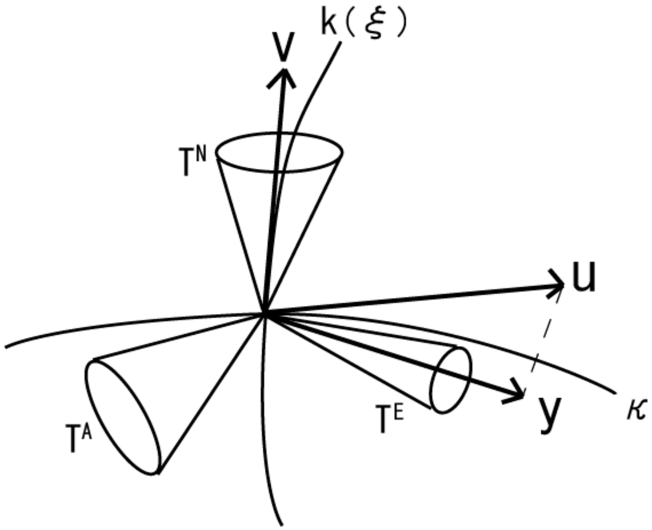
$$\longrightarrow \sum_{l=1}^{N} y(T_l;\kappa) = 0$$

E[$y(T,\kappa)$]=0

$\longleftrightarrow$

•**Maximum likelihood Method:**

$$\frac{d}{d\kappa}\log p(T_1)\cdots p(T_N) = \sum_{l=1}^{N} u(T_l;\kappa) = 0$$



(See poster for details)

<u>How to obtain an estimating function y:</u>

Score functions

$$\begin{cases} u(T;\hat{e},k) \equiv \dfrac{d \log p(T;\hat{e},k)}{d\hat{e}} \\ v(T;\hat{e},k) \equiv \dfrac{\delta \log p(T;\hat{e},k)}{\delta k(\hat{\imath}\ )} \end{cases}$$

$$\longrightarrow \quad y = u - \frac{< u \cdot v >}{< v \cdot v >} v$$

temporal correlations

○: $x(1)x(2)\cdots x(N)$

independent with firing probability $r(t)$

→ spike counts: Poisson
ISI: exponential

renewal:

$$r(t) = k(t - t_i)$$: last spike

ISI distribution $f(t)$

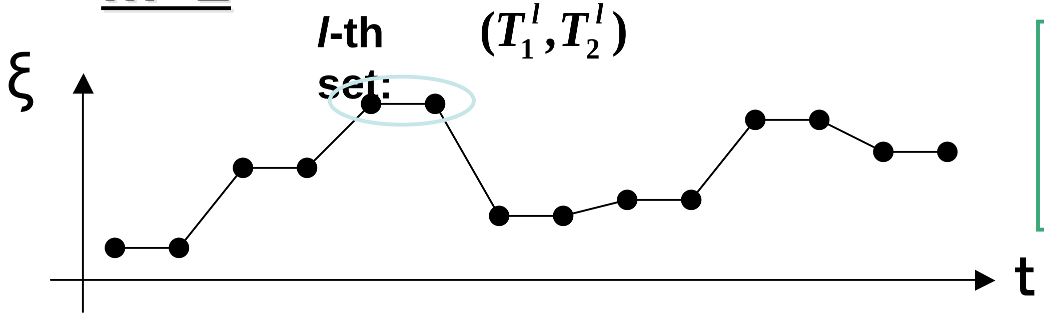
$$f(T) = ck(T)\exp\left\{\int_0^T k(t)dt\right\}$$

Estimation of κ by estimating functions

- No estimating function exists if the neighboring firing rates are different.
- m(≧2) consecutive observations must have the same firing rate.

Model: $p(T_1, T_2; \kappa, k) = \int_0^\infty q(T_1; \xi, \kappa) q(T_2; \xi, \kappa) k(\xi) d\xi$

Example:
m=2

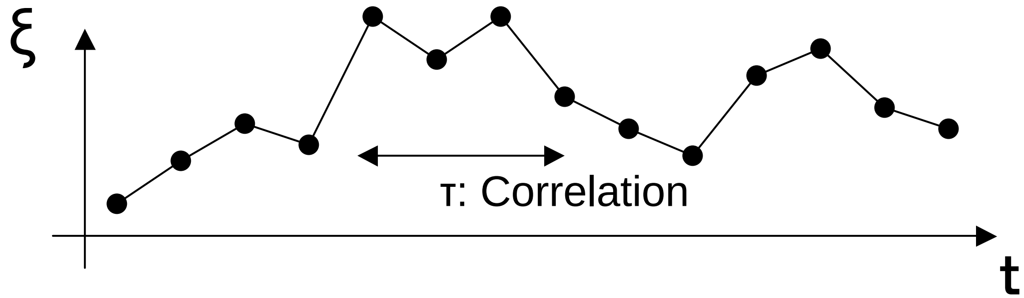


Estimating function:
(E[y]=0)

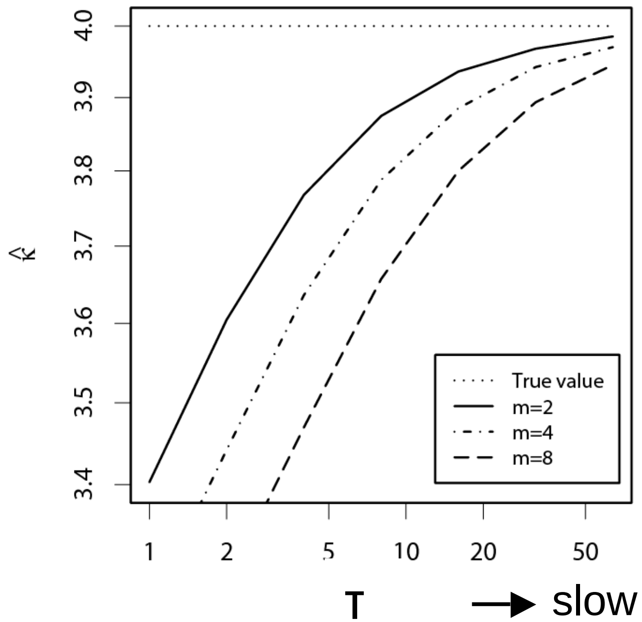
$$y = \log \frac{T_1 T_2}{(T_1 + T_2)^2} + 2\ddot{o}(2\hat{e}) - 2\ddot{o}(\hat{e})$$

$$\bar{y} = \frac{1}{N} \sum_{l=1}^{,m} \log \frac{T_1^l T_2^l}{(T_1^l + T_2^l)^2} + 2\ddot{o}(2\hat{e}) - 2\ddot{o}(\hat{e}) = 0$$

Case of m=1 (spontaneous discharge)



Firing rate continuously changes and is driven by Gaussian noise.



<u>κ can be approximated if the firing rate changes slowly.</u>

Estimating function (m=2):

$$\bar{y} = \frac{1}{N}\sum_{l=1}^{m} \log \frac{T_1^{\,l} T_2^{\,l}}{\left(T_1^{\,l} + T_2^{\,l}\right)^2} + 2\ddot{o}(2\hat{e}) - 2\ddot{o}(\hat{e}) = 0$$