•セミパラメトリック統計推論の情報幾何

外生変数に雑音がある場合

甘利俊一
理化学研究所脳科学総合研究センター

線形回帰: Semiparametrics

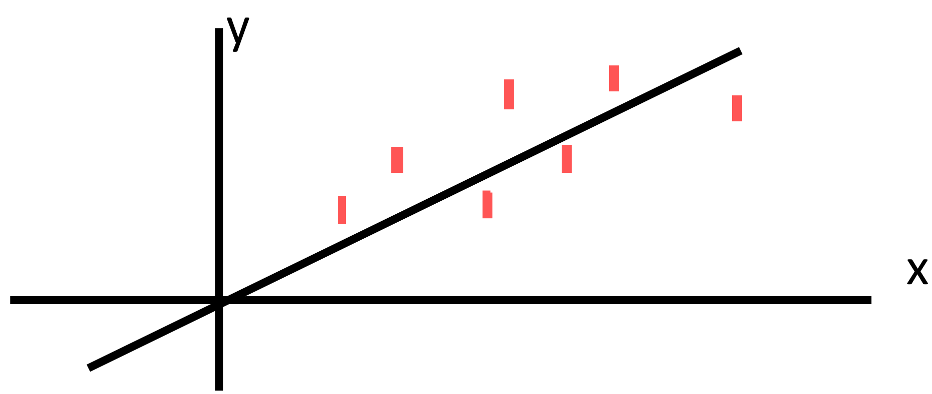
$$(x_1, y_1)$$
$$(x_2, y_2)$$
$$\vdots$$
$$(x_n, y_n)$$

$$y = \theta x$$

$$x_i = \xi_i + \varepsilon_i$$

$$y_i = \theta \xi_i + \varepsilon_i'$$

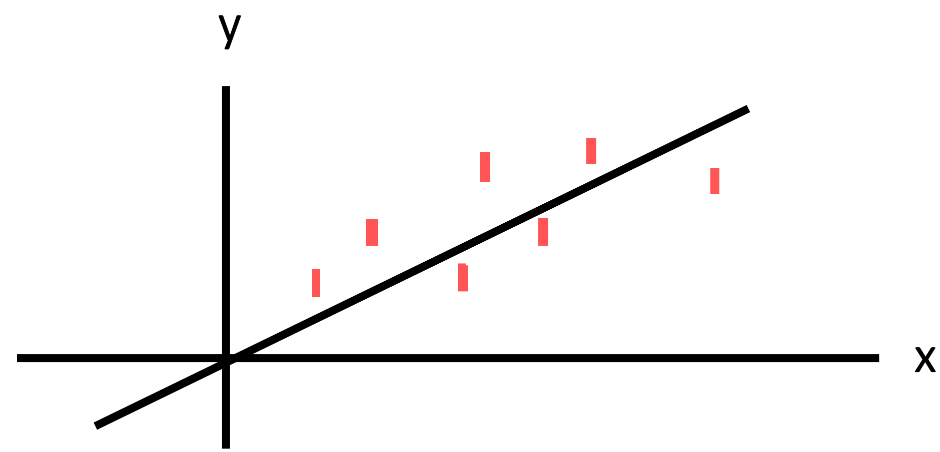
$$\varepsilon_i, \varepsilon_i' \sim N(0, \sigma^2)$$



Semiparametric 統計モデル

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

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



linear relation

$$y = \theta x$$

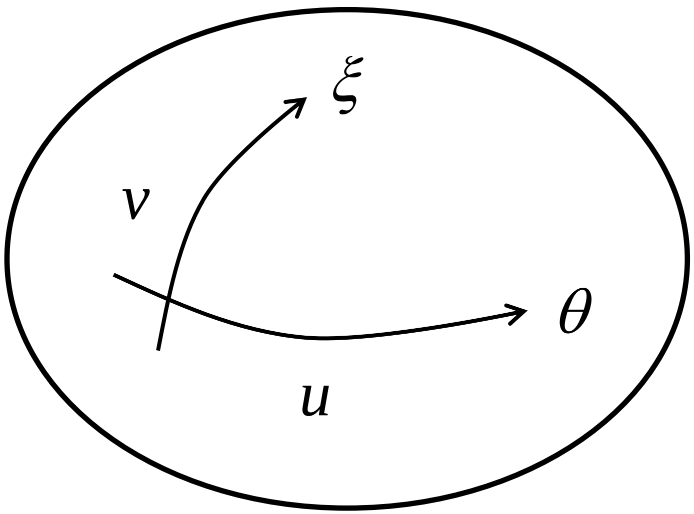
$$\begin{cases} y_i = \theta \xi_i + \varepsilon_i \\ x_i = \xi_i + \varepsilon_i' \end{cases}$$

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

mle, least square, total least square

攪乱変数のある統計モデル

$$p(x, \theta, \xi):$$



スコア関数

$$u(x, \theta, \xi) = \frac{\partial}{\partial \theta} \log p(x, \theta, \xi)$$

$$v(x, \theta, \xi) = \frac{\partial}{\partial \xi} \log p(x, \theta, \xi)$$

同時推定とFisher情報量

$$x_1, \cdots, x_n$$

最尤推定

$$\sum u\left(x_i, \theta, \xi\right) = 0$$
$$\sum v\left(x_i, \theta, \xi\right) = 0$$

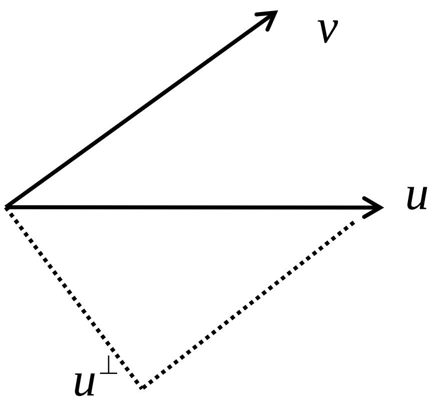
Fisher情報行列

$$g = \begin{bmatrix} g_{\theta\theta} & g_{\theta\xi} \\ g_{\theta\xi} & g_{\xi\xi} \end{bmatrix} = \begin{bmatrix} E\left[u^2\right] & E[uv] \\ E[uv] & E\left[v^2\right] \end{bmatrix}$$

情報量損失と直交性

$$E[uv]=\langle u,v\rangle=0$$ スコア直交の時

$$\left(g^{-1}\right)_{\theta\theta}=1/g_{\theta\theta}$$



非直交:

$$\left(g^{-1}\right)_{\theta\theta}=E\left[\left(u^{\perp}\right)^2\right]$$

$$u=u^{\perp}+cv \quad ; \quad c=\frac{E[uv]}{E\left[v^2\right]}$$

$$E[(u^{\perp})^2]=E[u^2]-c^2E[v^2]$$

無限個の外乱母数–Neyman–Scott問題

$$p\left(x,\theta,\xi\right)$$

$$x_1 \sim p\left(x,\theta,\xi_1\right)$$

$$x_2 \sim p\left(x,\theta,\xi_2\right)$$

$$\vdots$$

$$x_n \sim p\left(x,\theta,\xi_n\right)$$

未知パラメータ:

$$\theta,\xi_1,\cdots,\xi_n$$

Semiparametric Statistical Model

$$p\left(x,\theta,\xi\right): x_i \sim \xi_i$$

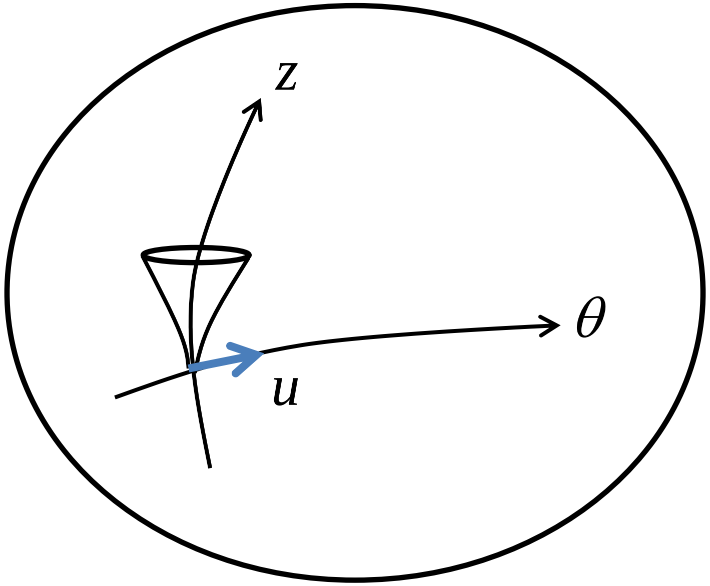
$$\xi_1,\cdots,\xi_n \sim z\left(\xi\right)$$

$$x_i \sim p\left(x,\theta,z\right)=\int p\left(x,\theta,\xi\right)z\left(\xi\right)d\xi$$

θ: parameter of interest

z: nuisance parameter 関数次元

幾何学



$$u(x, \theta, z)$$

$$v(x, \theta, z)$$

$$v_t = \frac{\partial}{\partial t} \log p\left(x, \theta, z(\xi, t)\right)$$

$$v(x, \theta, \xi) = \frac{\partial}{\partial \xi} \log p$$

直交スコア

$$V^N = \left\{ v_t\left(x, \theta, z\right) \right\} = \left\{ v\left(x, \theta, \xi\right) \right\} : \text{mixture model}$$

$$u^{\perp} = u - \int c_t v_t dt$$

Projected score　直交スコア

推定関数法　estimating function

$$f(x,\theta):$$

$$E_{\theta,Z}\left[f(x,\theta)\right]=0: \text{ unbiased}$$

$$\sum_{i=1}^{n} f(x_i,\theta)=0: \text{ 推定方程式}$$

$$E\left[\left(\hat{\theta}-\theta\right)^2\right]=\frac{1}{n}\frac{E\left[f^2\right]}{E\left[\left(\partial_\theta f\right)^2\right]}$$

平行移動

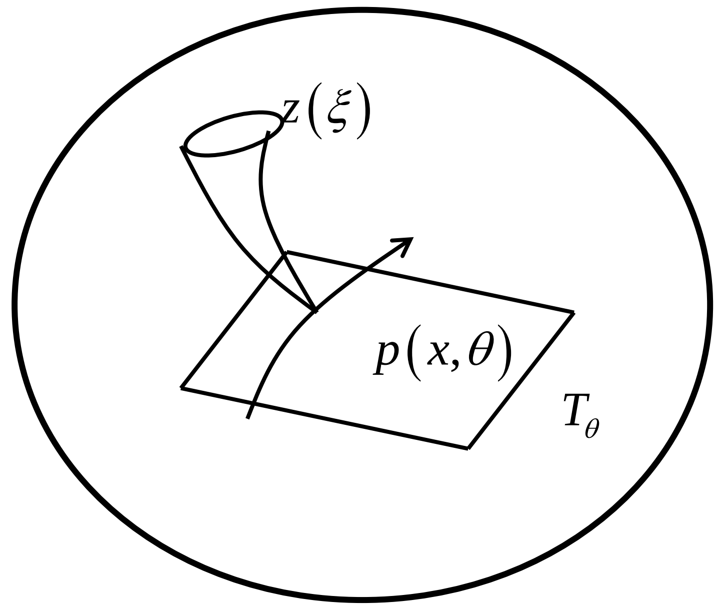
$$T_\theta = \left\{ r(x) \middle| E_\theta\left[r(x)\right] = 0 \right\}$$

$$\text{内積} \langle r_1, r_2 \rangle = E_\theta\left[r_1(x) r_2(x)\right]$$

$$\overset{e}{\prod}{}_z^{z'} r(x) = r(x) E_{\theta,z'}\left[r(x)\right]$$

$$\overset{m}{\prod}{}_z^{z'} r(x) = \frac{p(x,\theta,z)}{x,\theta,z'} r(x)$$

$$\langle r_1, r_2 \rangle_{z'} = \left\langle \overset{e}{\prod}{}_z^{z'} r_1, \overset{m}{\prod}{}_z^{z'} r_2 \right\rangle$$



推定関数 $f(x,\theta)$

$$e-\text{不変}: E\left[\, f(x,\theta)\right] = 0$$

$$\prod_z^{z'} f(x,\theta) = f$$

$$m-\text{直交性}: \langle v, f \rangle = 0$$

$$\left\langle \overset{m}{\prod_z^{z'}} v, f \right\rangle = 0$$

$$T_\theta = T_\theta^I \oplus T_\theta^N \oplus T_\theta^A$$

$$u^I(x,\theta,z): \text{最適推定関数}$$

悪い推定量

$$z = z(\xi, \eta)$$

$$p(x, \theta, z(\xi, \eta)) \to \hat{\theta}, \hat{\eta}$$

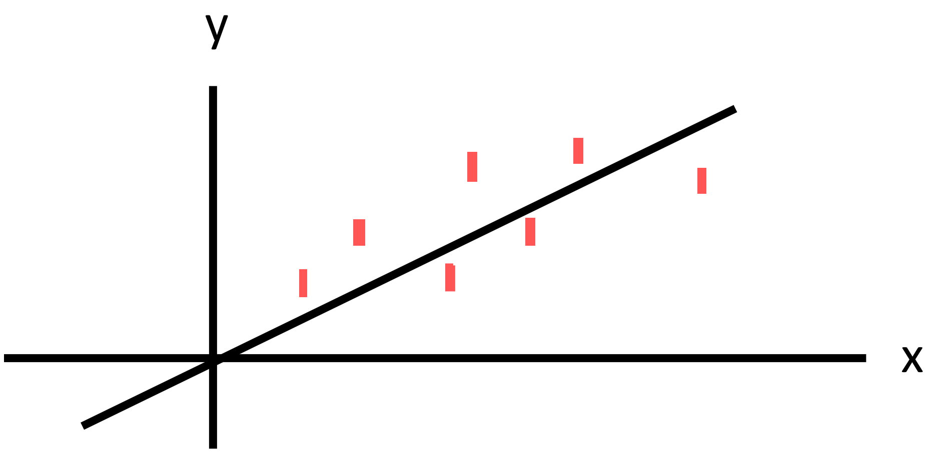
$$u^2(x, \theta, \hat{\eta}) \qquad \times$$

$$u^I(x, \theta, \hat{\eta}) \qquad \text{OK}$$

Semiparametric Statistics

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

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



linear relation

$$y = \theta x$$

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

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

mle, least square, total least square

Linear Regression: Semiparametrics

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

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

$$\vdots$$

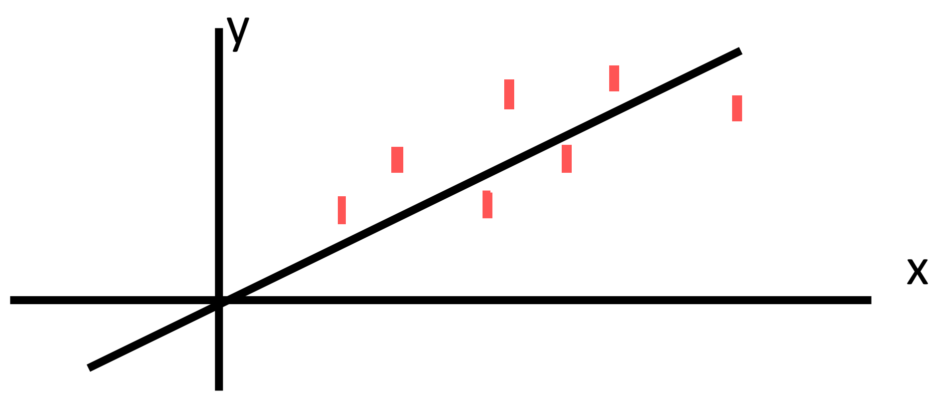
$$\left(x_n, y_n\right)$$

$$y = \theta x$$

$$x_i = \xi_i + \varepsilon_i$$

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

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



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

Semiparametric statistical model

$$x_1, x_2, \cdots \sim p\left(x, \theta, Z\right)$$

Estimating function

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

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

Estimating equation

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

Example of estimating functions

$$f\left(x, y; \theta\right) = k\left(x + \theta y\right)\left(y - \theta x\right)$$

$$\int c \exp\left\{-\frac{1}{2}\left(x-\xi\right)^2 - \frac{1}{2}\left(y-\theta'\xi\right)^2 \right.$$

$$\left. \times\left(y-\theta x\right)k\left(x+\theta y\right)Z\left(\xi\right)dxdyd\right\}$$

$$= 0, \qquad \theta = \theta'$$

$$\neq 0, \qquad \theta \neq \theta'$$

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

$$\frac{1}{\sqrt{n}}\sum y\left(x_i,\theta\right)$$

$$=\frac{1}{\sqrt{n}}\sum y\left(x_i,\theta\right)+\frac{\sqrt{n}}{n}\sum y'\left(x_i,\theta\right)\left(\theta-\theta\right)$$

$$\theta-\theta=\frac{1}{\sqrt{n}}\frac{\varepsilon}{E\left[y'\left(x,\theta\right)\right]}$$

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

$$=\exp\left\{\frac{\xi}{\sigma^{2}}\left(x+\theta y\right)+c\left(x,y\right)-\psi\left(\theta,\xi\right)\right\}$$

$$s\left(x,y;\theta\right)=x+\theta y$$

$$p\left(x,\theta,Z\right)=\int\exp\left\{\xi s\left(x,y;\theta,\right)+c-\psi\right\}Z\left(\xi\right)d\xi$$

$$\sum f\left(x_i, y_i; \theta\right) = 0$$

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

$$c = \frac{\bar{\xi}\sigma^2}{\overline{\xi^2} - \left(\bar{\xi}\right)^2} \qquad \left(\begin{array}{l} \bar{\xi} = 1 \\ \overline{\xi^2} = 2 \end{array}\right.$$

$$c = 0 : V = \frac{1}{n}\frac{\left(2 + \sigma^2\right)\sigma^2}{4} \qquad : \frac{3}{4}$$

$$c = 1 : V = \frac{1}{n}\left(1 - \frac{1}{\sigma^2 + 2}\right)\sigma^2 \qquad : \frac{2}{3}$$

$$c = \infty : v = \frac{1}{n}\sigma^2 \qquad : 1$$

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

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

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

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

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

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

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

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

$$p\left(x,\theta,Z\right)$$

$$\mathrm{Z}\left(\xi\right)=\mathrm{Z}\left(\xi,\eta\right)$$

$$p\left(x,\theta,Z\left(\cdot,\eta\right)\right)$$ does not work

$$u^{I}\left(x,\theta,Z\right)$$ estimating function

Generalization

$$y = \boldsymbol{\theta} \cdot \boldsymbol{x}$$

$$\boldsymbol{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i$$

$$y_i = \boldsymbol{\theta} \cdot \boldsymbol{\xi}_i + \varepsilon_i^{'}$$

$$\boldsymbol{f}\left(\boldsymbol{x}, y; \boldsymbol{\theta}\right) = \left(y - \boldsymbol{\theta} \cdot \boldsymbol{x}\right) k\left(\boldsymbol{x} + y\boldsymbol{\theta}\right)$$

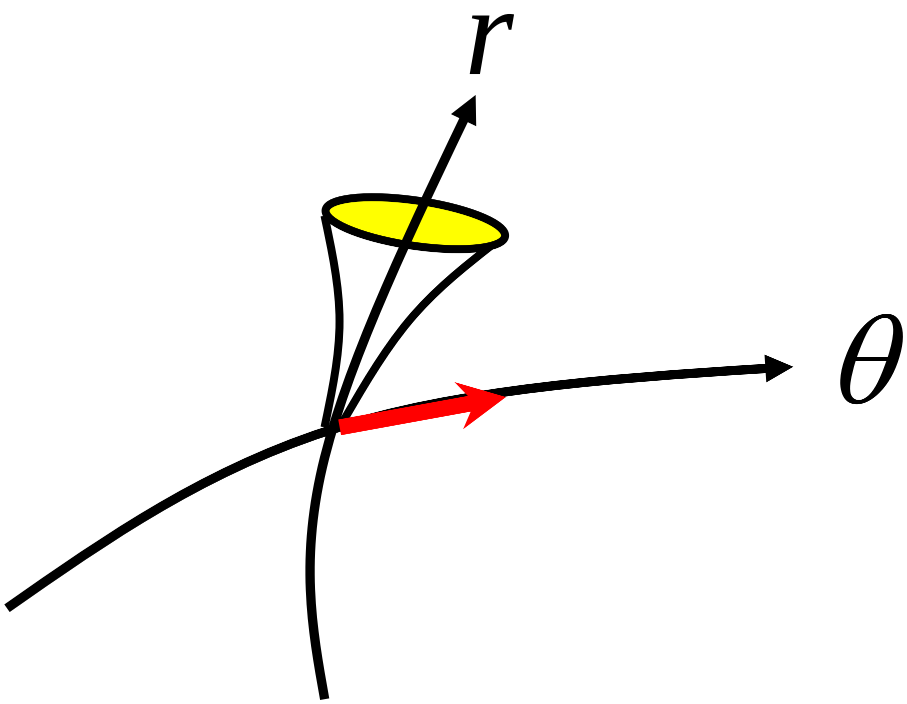
$$\boldsymbol{f}_c\left(\boldsymbol{x}, y; \boldsymbol{\theta}\right) = \left(y - \boldsymbol{\theta} \cdot \boldsymbol{x}\right)\left(\boldsymbol{x} + y\boldsymbol{\theta} + \boldsymbol{c}\right)$$

$$\sum\left(y_i - \boldsymbol{\theta} \cdot \boldsymbol{x}_i\right)\left(\boldsymbol{x}_i + y_i\boldsymbol{\theta} + \boldsymbol{c}\right) = 0$$

$$\boldsymbol{c}^* = \sigma^2 V_\xi^{-1} \overline{\boldsymbol{\xi}}$$

Fibre bundre of probability distributions

function space



Least squares?

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

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

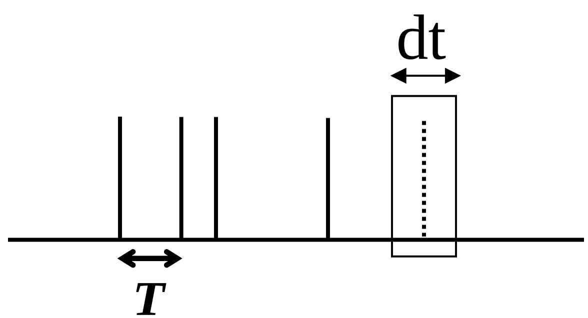
mle, TLS

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

Neyman-Scott

Poisson process

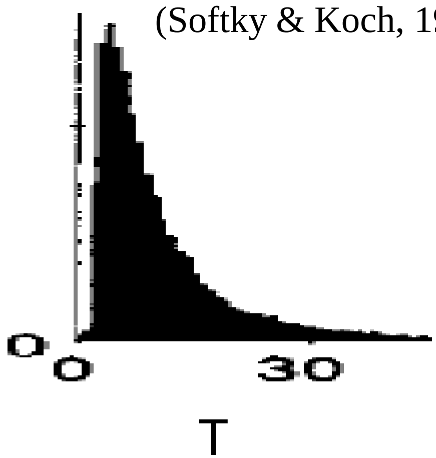
Poisson Process: Instantaneous firing rate is constant over time.



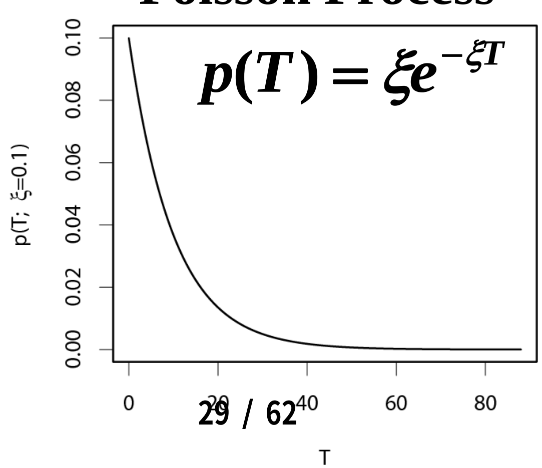
For every small time window dt, generate a spike with probability ξdt.

Cortical Neuron

(Softky & Koch, 1993)



Poisson Process



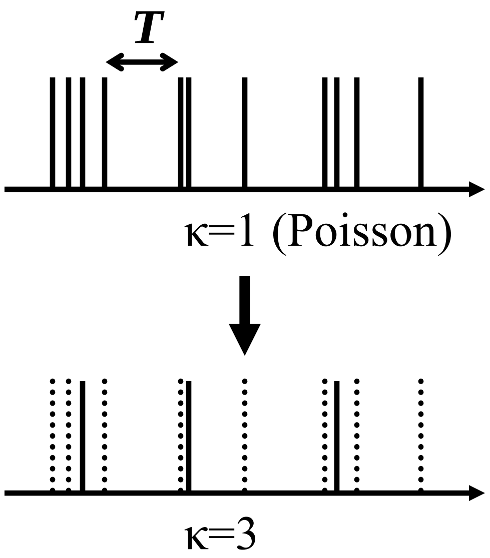
<u>**Poisson process cannot explain inter-spike interval distributions.**</u>

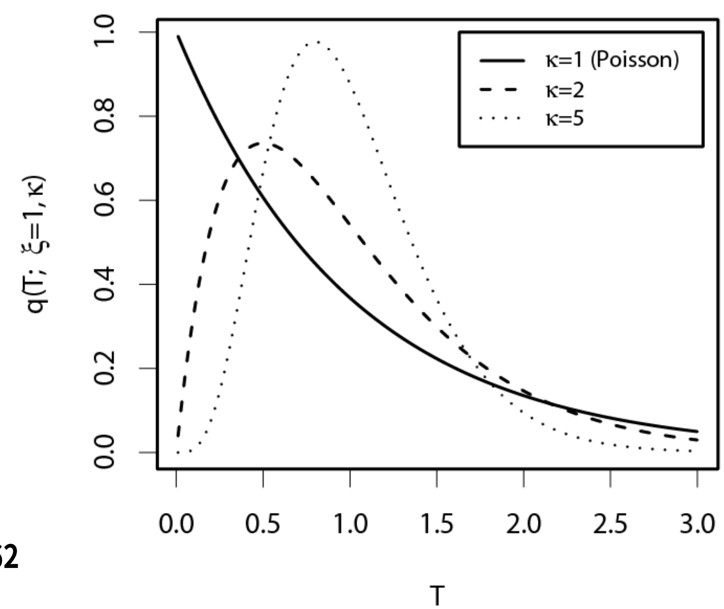
Gamma distribution

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

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

Two parameters $\begin{cases}\xi\text{: } \textbf{Firing rate}\\ \kappa\text{: } \textbf{Irregularity}\end{cases}$





Gamma distribution

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

$\kappa=1$: Poisson

$\kappa\to\infty$: regular

Integrate-and fire

Markov model

temporal correlations

O: $x(1)\,x(2)\,\mathrm{L}\;\, x(N)$

independent with firing probability $r(t)$

→ spike counts: Poisson
ISI: exponential

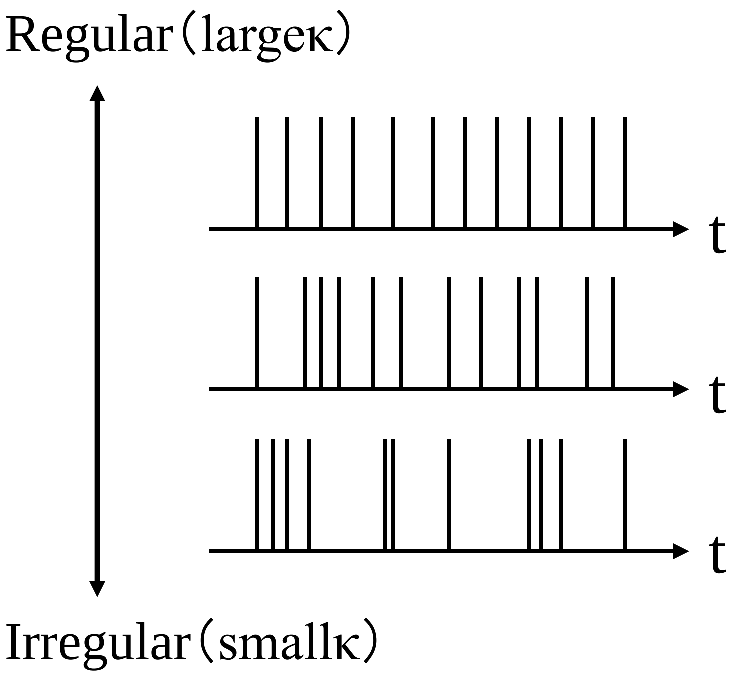
renewal:

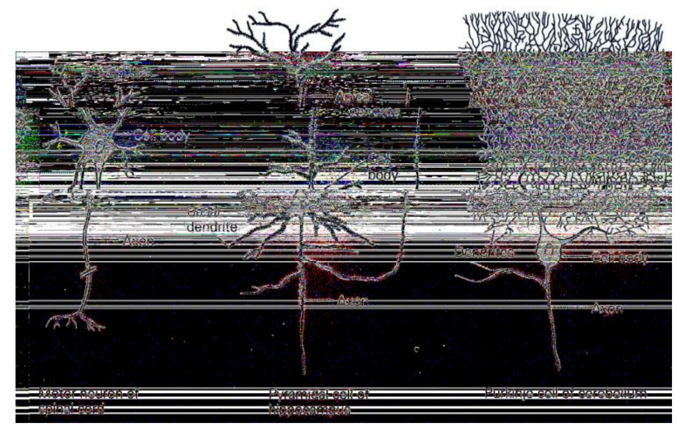
$$r(t) = k(t - t_i)$$: last spike

ISI distribution $f(t)$

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

Irregularity κ is unique to individual neurons.



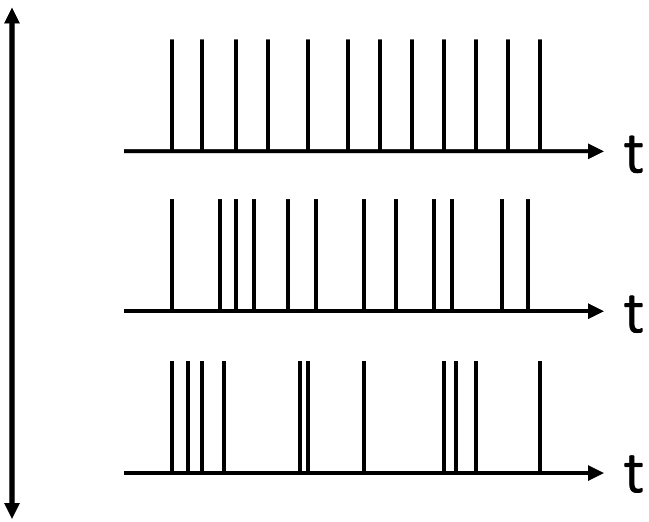


Irregularity varies among neurons.
(Baker & Lemon 2000; Shinomoto et.al., 2003)

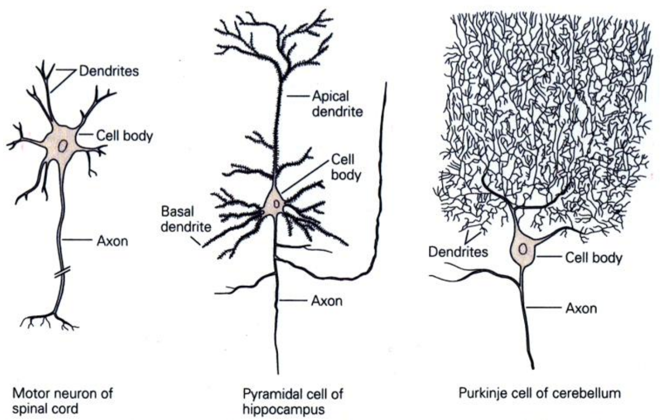
→ **<u>We assume that κ is independent of time.</u>**

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 $y(T,\kappa)$:**

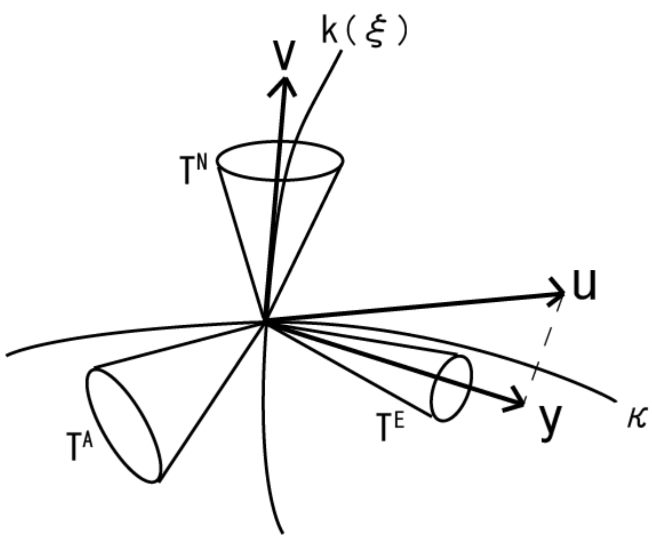
$$\longrightarrow \sum_{l=1}^{N} y(T_l;\kappa) = 0$$

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)

How to obtain an estimating function *y*:

Score functions

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

$$\longrightarrow \quad y = u - \frac{\langle u\cdot v\rangle}{\langle v\cdot v\rangle} v$$

temporal correlations

O: $x(1)\,x(2)\,\mathrm{L}\;x(N)$

independent with firing probability $r(t)$

→ spike counts: Poisson
ISI: exponential

renewal:

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

ISI distribution $f(t)$

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

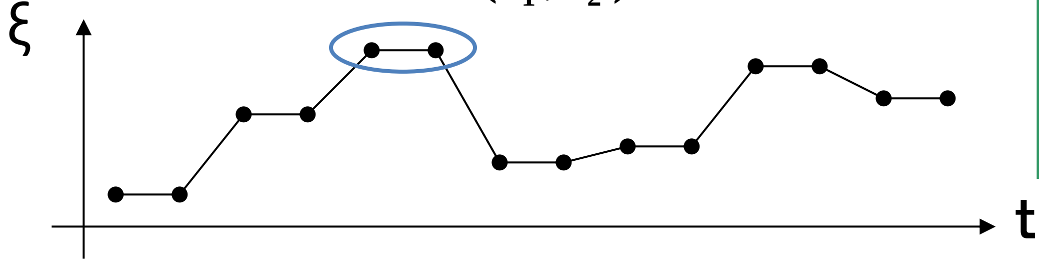
Estimation of κ by estimating functions

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

Example: m=2

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

***l*-th set:** (T_1^l, T_2^l)

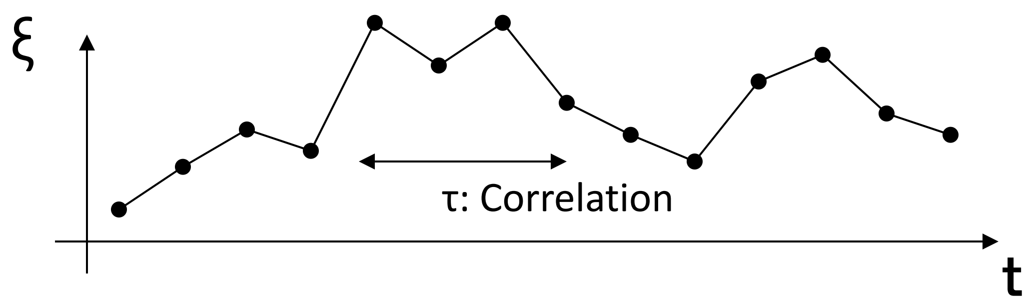


Estimating function: (E[y]=0)

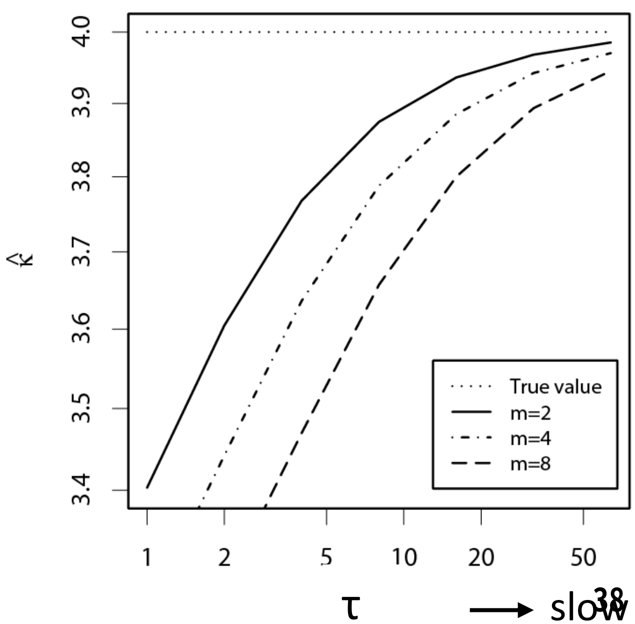
$$y=\log\frac{T_1T_2}{(T_1+T_2)^2}+2\varphi(2\kappa)-2\varphi(\kappa)$$

$$\bar{y}=\frac{1}{N}\sum_{l=1}^{N}\log\frac{T_1^lT_2^l}{\left(T_1^l+T_2^l\right)^2}+2\varphi(2\kappa)-2\varphi(\kappa)=0$$

Case of m=1
(spontaneous discharge)



Firing rate continuously changes and is driven by Gaussian noise.

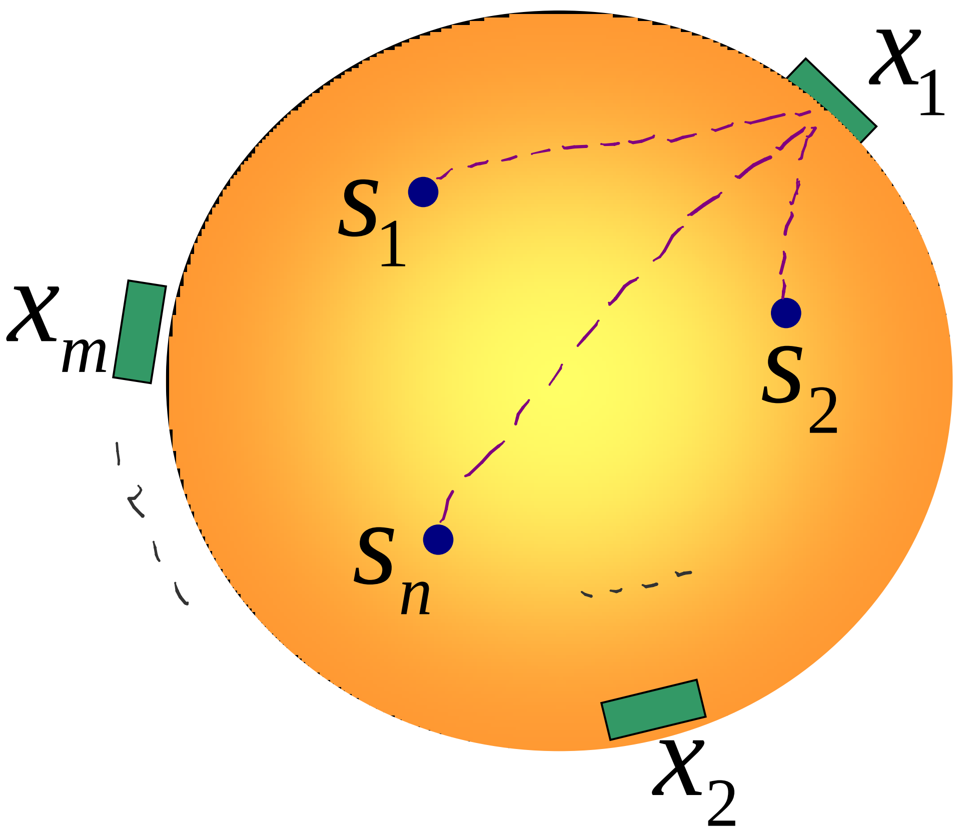


<u>κ can be approximated if the firing rate changes slowly.</u>

Estimating function (m=2):

$$\bar{y} = \frac{1}{N}\sum_{l=1}^{N}\log\frac{T_1{}^{l}T_2{}^{l}}{\left(T_1{}^{l}+T_2{}^{l}\right)^2} + 2\varphi(2\kappa) - 2\varphi(\kappa) = 0$$

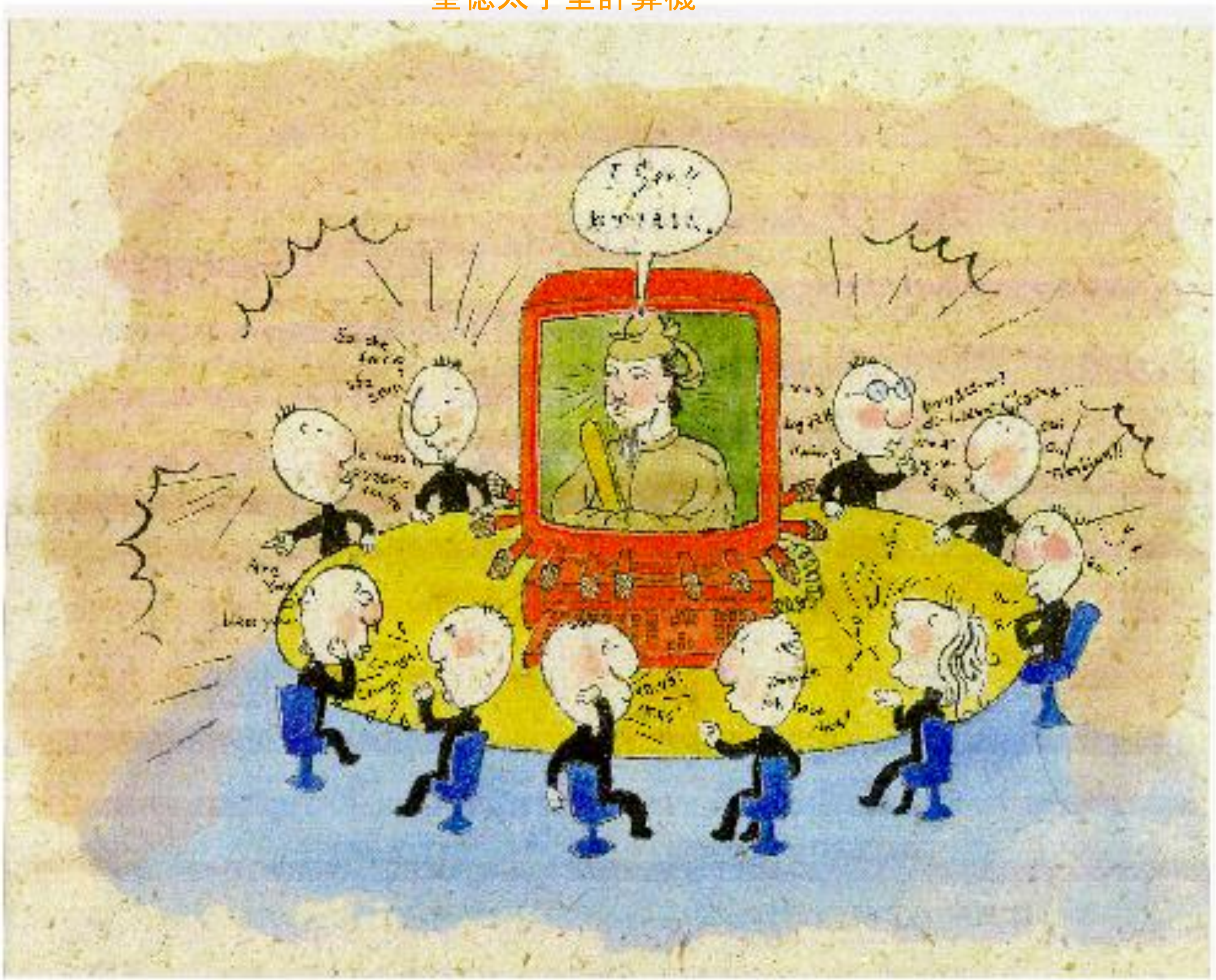
mixture and unmixture of independent signals



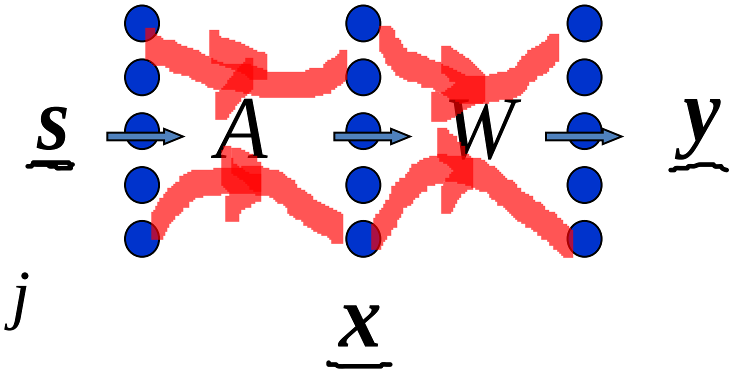
$$x_i = \sum_{j=1}^{n} A_{ij} s_j$$

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

聖徳太子型計算機



Independent Component Analysis



$$\boldsymbol{x} = A\boldsymbol{s} \qquad x_i = \sum A_{ij} s_j$$

$$\boldsymbol{y} = W\boldsymbol{x} \qquad W = A^{-1}$$

observations: x(1), x(2), …, x(t)
recover: s(1), s(2), …, s(t)

$$\mathbf{x=As} \Rightarrow \mathbf{y=Wx} \quad : \quad \mathbf{W=A^{-1}}$$

observations: x(1), x(2), …, x(t)

A: unknown matrix

s: unknown

$$r(\mathbf{s}) = r_1(s_1)r_2(s_2)\ldots r_n(s_n)$$

independent distribution

r(s): unknown

$$E[\mathbf{s}] = 0$$

$$\Delta \boldsymbol{W} = -\eta \frac{\partial l(\mathbf{y}, \boldsymbol{W})}{\partial \boldsymbol{W}}$$

cost function:
degree of non-independence

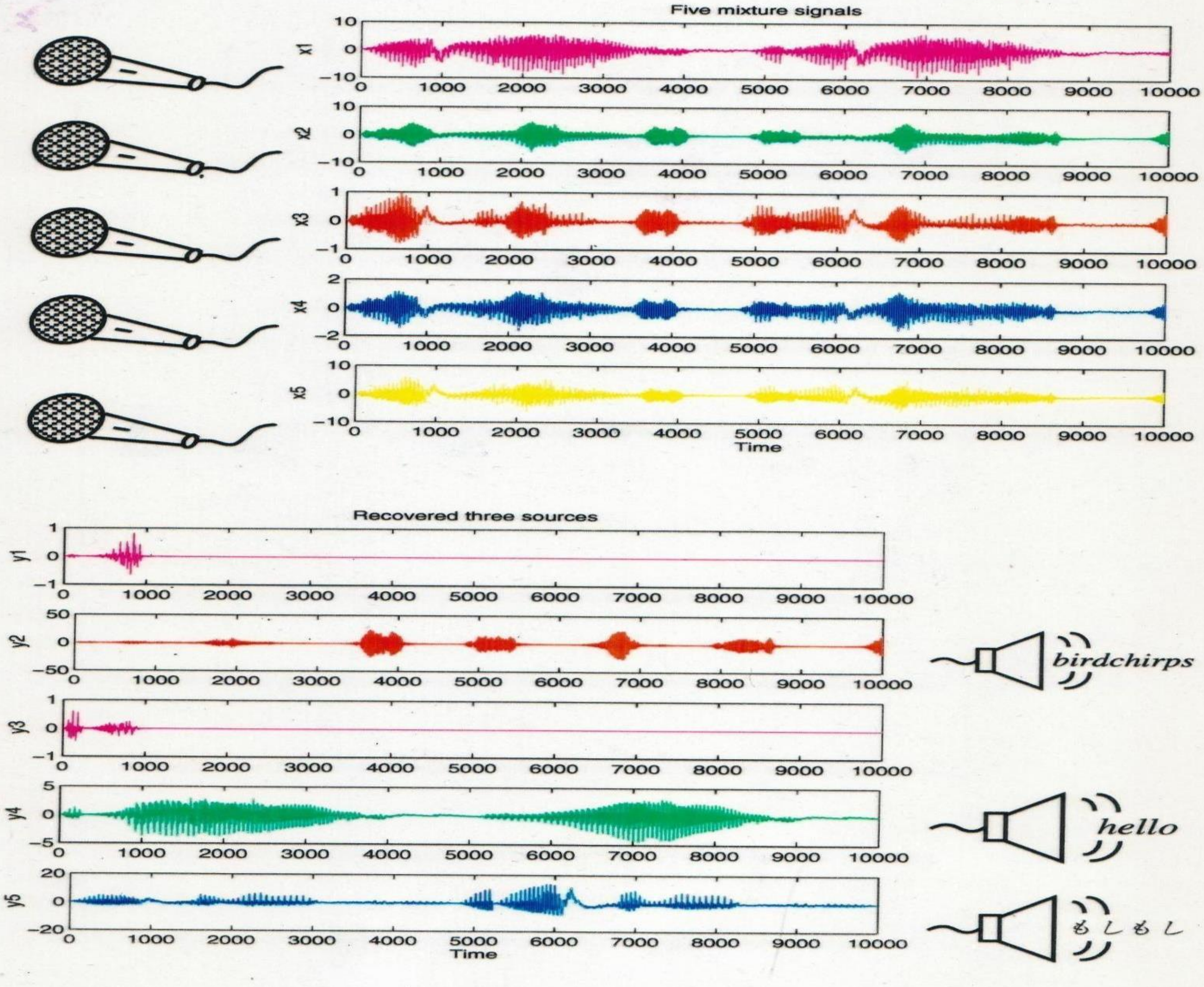
Semiparametric Statistical Model

$$p(\mathbf{x};\mathbf{W},r) = |\mathbf{W}|\, r(\mathbf{W}\mathbf{x})$$

$\mathbf{W} = \mathbf{A}^{-1}$, $r(s)$: unknown $\qquad r = \prod r_i$

x(1), x(2), …, x(t)

Cocktail party experiment



- **5 microphones (sensors) and only 3 speakers**

Example of color image separation :

Five original images (but unknown to the neural net)

Five mixed images for separation

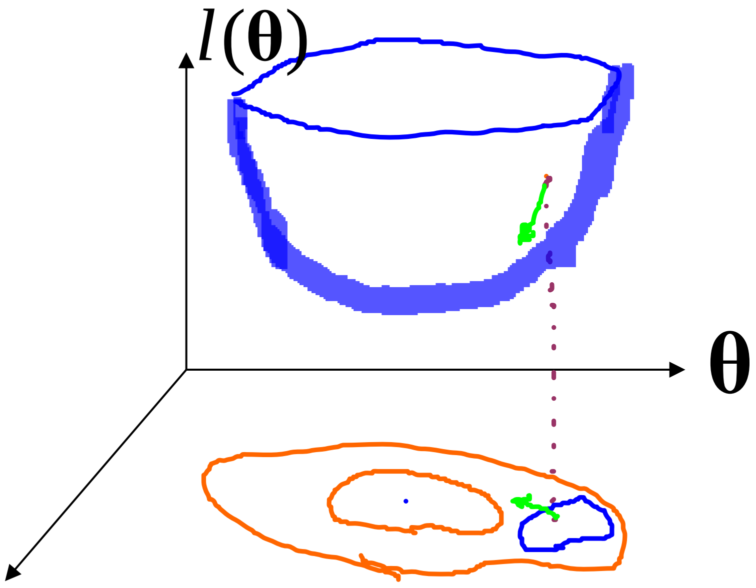
Final (stable states) of five separated images

Natural Gradient

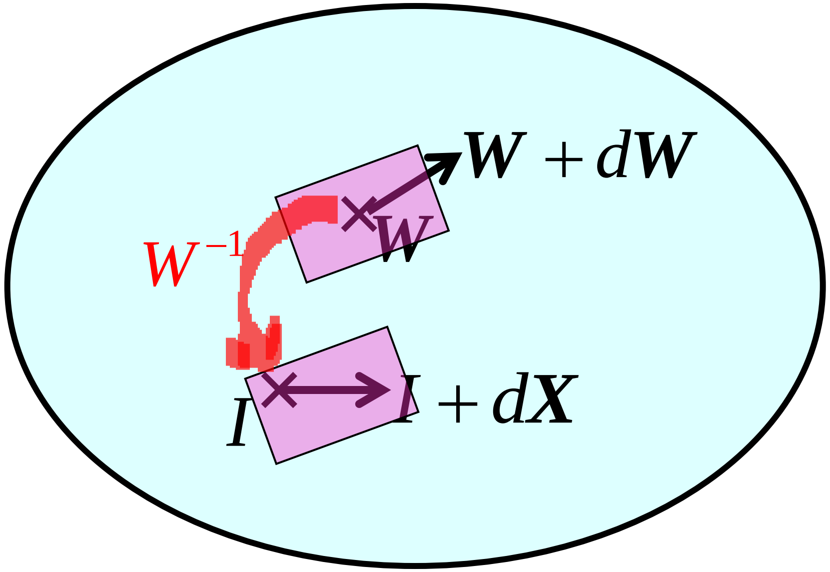
$$\max \qquad dl = l(\boldsymbol{\theta} + d\boldsymbol{\theta}) - l(\boldsymbol{\theta})$$

$$|d\boldsymbol{\theta}|^2 = \varepsilon$$

$$\nabla l = G^{-1}(\boldsymbol{\theta}) \nabla l$$



Space of Matrices : Lie group



$$d\boldsymbol{X} = d\boldsymbol{W}\boldsymbol{W}^{-1}$$

$$\left|d\boldsymbol{W}\right|^2 = \mathrm{tr}\left(d\boldsymbol{X}d\boldsymbol{X}^T\right) = \mathrm{tr}\left(d\boldsymbol{W}\boldsymbol{W}^{-1}\boldsymbol{W}^{-T}d\boldsymbol{W}^T\right)$$

$$\nabla l = \frac{\partial l}{\partial \boldsymbol{W}}\boldsymbol{W}^T\boldsymbol{W}$$

dX : **non-holonomic basis**

Natural Gradient

$$\Delta \boldsymbol{W} = -\eta \frac{\partial l\left(\boldsymbol{y}, \boldsymbol{W}\right)}{\partial \boldsymbol{W}} \boldsymbol{W}^T \boldsymbol{W}$$

Natural Gradient Learning Algorithm

$$\Delta \boldsymbol{W} = -\eta \boldsymbol{F}(\boldsymbol{y})$$

$$\boldsymbol{F}(\boldsymbol{y}) = \{I - \boldsymbol{\varphi}(\boldsymbol{y})\boldsymbol{y}^T\}\boldsymbol{W} \quad ; \quad \boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$$

$$\Delta W_{ij} = -\eta \sum_k \{\delta_{ik} - \varphi_i(y_i) y_k\} W_{jk}$$

$$\varphi_i(y_i) = -\frac{d}{dy} \log q_i(y_i)$$

Estimating Functions

$$\Delta \boldsymbol{W} = -\eta \boldsymbol{F}(\boldsymbol{y}, \boldsymbol{W})$$

$$E_{W,r}\left[\boldsymbol{F}\left(\boldsymbol{y}, \boldsymbol{W}'\right)\right] \begin{cases} = 0, & \boldsymbol{W}' = \boldsymbol{W} \\ \neq 0, & \boldsymbol{W}' \neq \boldsymbol{W} \end{cases}$$

estimating equation

$$\sum_t \boldsymbol{F}\left(\boldsymbol{y}_t, \boldsymbol{W}\right) = 0 \qquad \boldsymbol{y}_t = \boldsymbol{W}\boldsymbol{x}_t$$

learning

$$\Delta \boldsymbol{W}_t = -\eta \boldsymbol{F}\left(\boldsymbol{W}_t \boldsymbol{x}_t\right)$$

Admissible class

$$\boldsymbol{F}(\boldsymbol{y}, \boldsymbol{W}) = \{\boldsymbol{I} - \boldsymbol{\varphi}(\boldsymbol{y})\boldsymbol{y}^T\}W$$

$$\tilde{\boldsymbol{F}} = \boldsymbol{R}(\boldsymbol{W})\boldsymbol{F}(\boldsymbol{y}, \boldsymbol{W}) \qquad \{\alpha\varphi(y_i)y_j - \varphi(y_j)y_i\}W$$

$\sum \boldsymbol{R}(\boldsymbol{W})\boldsymbol{F}(\boldsymbol{W}\boldsymbol{x}_t) = 0$: estimating equation

on-line learning : $\Delta \boldsymbol{W}_t = -\eta \boldsymbol{R}(\boldsymbol{W}_t)\boldsymbol{F}(\boldsymbol{W}_t\boldsymbol{x}_t)$

canonical estimating function: $\dfrac{\partial \tilde{F}}{\partial W} = I$

Canonical Estimating Function ---Newton Method

$$\delta \boldsymbol{W}_t = -\eta \left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{W}} \right)^{-1} \boldsymbol{F}(\boldsymbol{y}_t)$$

$$F = \nabla l$$

$$(\nabla\nabla l)^{-1} \nabla l : \quad \text{Newton Method}$$

Stabilization—Newton Method

$$\Delta W_{ij}(t) = \eta_t \frac{1}{\sigma_i^2 \sigma_j^2 k_i k_j - 1}\left[\sigma_i^2 k_j \varphi(y_i) y_j - \varphi(y_j) y_i\right] W$$

$$= \eta_t \{\alpha \varphi(y_i) y_j - \varphi(y_j) y_i\} W$$

$$\sigma_i^2 = E[y_i^2], \quad \kappa_i = E[\varphi_i'(y_i)]$$

efficient algorithm

$\Rightarrow$

adaptive method

Stability Analysis

$$\dot{\boldsymbol{W}}(t) = -\eta E\left[\boldsymbol{F}(\boldsymbol{y})\right]$$

$$\delta\dot{\boldsymbol{W}}(t) = -\eta E\left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{W}}\right]\delta\boldsymbol{W}$$

$$\boldsymbol{R} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{W}} \quad : \quad R_{ij,kl}$$

$$R_{AA'} = \begin{array}{cc} & \begin{array}{cc} A & A' \end{array} \\ \begin{array}{c} A \\ A' \end{array} & \begin{bmatrix} \kappa_i & 1 \\ 1 & \kappa_j \end{bmatrix} \end{array}$$

$$A = (ij)$$
$$A' = (ji)$$

$$\kappa_i = E\left[\varphi_i^{'}(s_i)\right]$$

$$\kappa_i \kappa_j > 1 \quad , \quad \kappa_i > 0$$

Estimating function

$$y = \mathbf{w} \cdot \mathbf{x}$$

$$\mathbf{f}(\mathbf{x}, \mathbf{w}) = \varphi(y)\mathbf{x} - y\varphi(y)\mathbf{w}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \{\varphi(y_t)\mathbf{x}_t - y_t\varphi(y_t)\mathbf{w}_t\}$$

Newton's Method

FastICA

Estimating Function in Noisy Case

$$\boldsymbol{x} = A\boldsymbol{s} + \boldsymbol{n}$$

$$\boldsymbol{F} = \left(F_{ab} \right)$$

$$F_{ij} = y_i^3 y_j - 3v_{ii} y_i y_j - 3v_{ij} y_i^2 + 3v_{ii} v_{ij}$$

$$v_{ij} = E\left[n_i n_j \right]$$

$$\Delta v_{ij} = -c\left(v_{ij} - y_i y_j \right)$$

Temporal Correlation

$$\boldsymbol{s}(1), \boldsymbol{s}(2), \cdots, \boldsymbol{s}(t), \cdots$$

$$\boldsymbol{\Lambda}(\tau) = \overline{\boldsymbol{s}(t)\boldsymbol{s}(t-\tau)} = \mathrm{diag}[\lambda_1, \ \ldots, \ \lambda_n]$$

$$\boldsymbol{V}(\tau) = \overline{\boldsymbol{x}(t)\boldsymbol{x}(t-\tau)^T} = A\overline{\boldsymbol{s}(t)\boldsymbol{s}(t-\tau)}A^T$$

$$= A\boldsymbol{\Lambda}(\tau)A^T$$

Simultaneous Diagonalization

$$W\boldsymbol{V}(\tau)W^T = \boldsymbol{\Lambda}(\tau) \qquad \tau = 0, \ 1, \ 2, \cdots$$

Estimating Function in Temporally Correlated Case

$$\boldsymbol{F}\left(\boldsymbol{y}, W\right) = I - \varphi\left(\tilde{\boldsymbol{y}}\right) \circ \boldsymbol{B}\left(z^{-1}\right) \boldsymbol{y}^{T}$$

$$: \varphi_i\left(\tilde{y}_i\right) B_i(z^{-1}) y_t$$

$$\tilde{y}_i = B_i(z^{-1}) y_i = \sum_k b_{ik} y_i\left(t-k\right)$$

class of admissible estimating functions: maximum likelihood; joint diagonalization

Spatial mixing

$$\begin{cases} x_1 = a_1 s_1 + \cdots + a_n s_n \\ x_2 = a_1^{'} s_1 + \cdots + a_n^{'} s_n \end{cases}$$

$$\cdots\cdots$$

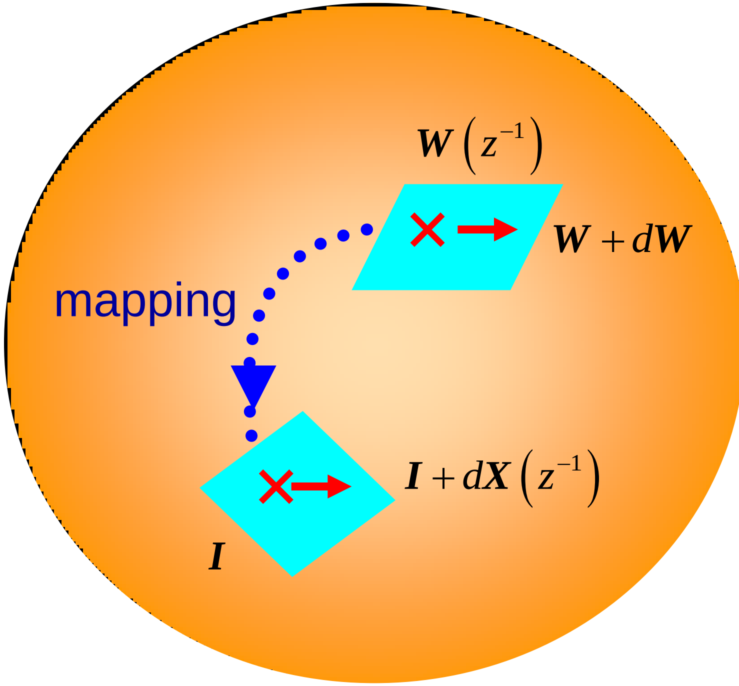
$$\boldsymbol{x} = A\boldsymbol{s}$$

Temporal mixing—convolution

$$x(t) = a_1 s(t) + a_2 s(t-1) + \cdots + a_n s(t-n)$$

$$\boldsymbol{x}(t) = \int \boldsymbol{A}(t-\tau)\boldsymbol{s}(\tau)\,d\tau$$

Manifold of Linear Systems



$$dX\left(z^{-1}\right)=dW\left(z^{-1}\right)W^{-1}\left(z^{-1}\right)$$

Metric Structure

$\left\{\begin{array}{l}\text{Lie group}\\ \text{Fisher information}\end{array}\right.$

$$\tilde{\nabla} f\left(\boldsymbol{W}\right)=\frac{\partial f}{\partial \boldsymbol{W}}\circ W^{T}\left(z\right)W\left(z^{-1}\right)$$

Manifold of Linear Systems

Topological Structure

$$\begin{cases} \boldsymbol{v}(t+1) = \boldsymbol{F}\boldsymbol{v}(t) + \boldsymbol{G}\boldsymbol{s}(t) \\ \boldsymbol{x}(t) = \boldsymbol{H}\boldsymbol{v}(t) \end{cases}$$

$$\boldsymbol{A}(z^{-1}) = \boldsymbol{H}(z\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{G}$$

$$\underbrace{\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{H}\}}_{\text{parameter}} / T$$

$$\begin{cases} \boldsymbol{T}\boldsymbol{F}\boldsymbol{T}^{-1} \\ \boldsymbol{H}\boldsymbol{T}, \quad \boldsymbol{T}^{-1}\boldsymbol{G} \end{cases}$$

Non-Holonomic Basis

$$dX = dW W^{-1}$$

$$\Delta W = -\eta \frac{\partial l}{\partial W} W^T W$$

$$\Delta X_t = -\eta \frac{\partial l}{\partial X}$$

$$\Delta X = -\eta \left(\frac{\partial^2 l}{\partial X \partial X} \right)^{-1} \frac{\partial l}{\partial X}$$