Peter Latham ①

11/11/04

Computing with Population Codes

OCNC

[1] Computing smooth functions in low dimensions (~10?) efficiently w/ pop codes.

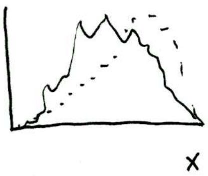
$z = \phi(x_1, x_2, \ldots x_D)$

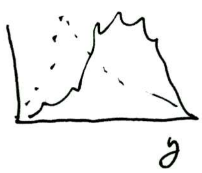
$\rightarrow z = \phi(x, y)$

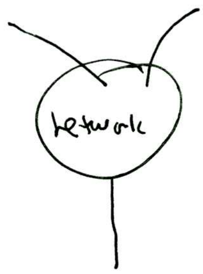
[Can follow along in book some of the figs are much better]

[Will scan/photocopy notes]

rate

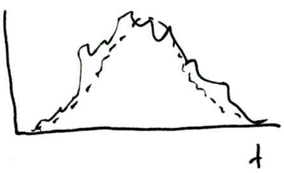






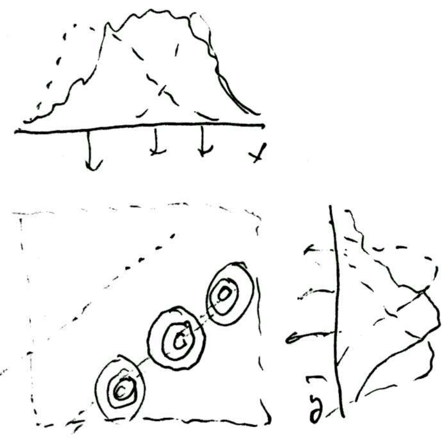
goal: understand how to build a network that computes arbitrary function.

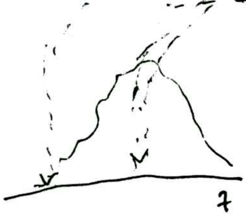
rate



$z = x + y$

[2]





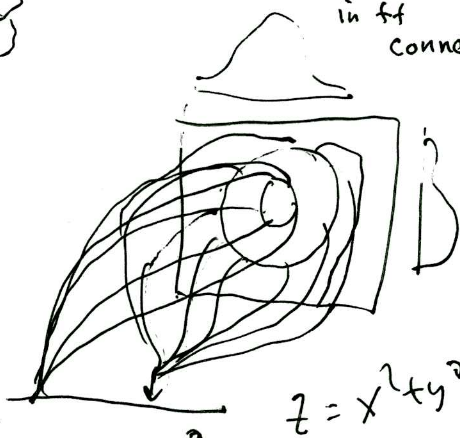
$z = x + y$

→ Key concept #1!!!

→ Only method we have for computing functions using population codes.

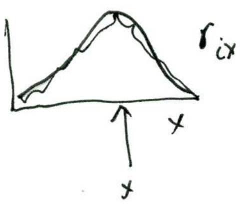
→ all action is in ff connections.

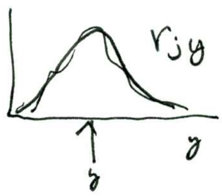
think about this as course continues.



$z = x^2 + y^2$

$\boxed{3}$ <u>Example</u>



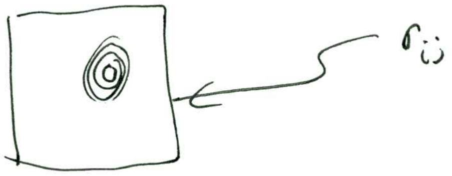


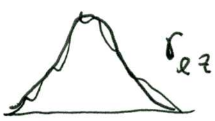
$$r_{ix} = \frac{e^{-\frac{1}{2}\frac{(x-x_i)^2}{\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

$i\Delta x$

$$r_{jy} = e^{-\frac{1}{2}(y-y_j)^2/\sigma^2}$$

$j\Delta y$





$$r_{ij} = r_{ix}\, r_{jy}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2+(y-y_j)^2\right]}}{2\pi\sigma^2}$$

$\boxed{4}$

$$r_{ez} = \frac{1}{N}\sum_{ij} \underbrace{\frac{e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2+(y-y_j)^2\right]}}{2\pi\sigma^2}}_{r_{ij}} \underbrace{\delta(z_e - \phi(x_i, y_j))}_{C_{eij}}$$

(linear!!!)

$$\langle z_e \rangle = \frac{1}{N}\sum_e z_e\, r_{ez}$$

$$= \frac{1}{N^2}\sum_{ij} \phi(x_i, y_j)\, \frac{e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2+(y-y_j)^2\right]}}{2\pi\sigma^2} \approx \phi(x,y)$$

Peaks when $x_i = x,\ y_j = y,\ z_e = \phi(x_i, y_j) = \phi(x,y)$

example: $\phi(x,y) = x+y$

$$r_{ez} = \frac{1}{N}\sum_i \frac{e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2 + (y-(z_e-x_i))^2\right]}}{2\pi\sigma^2}$$

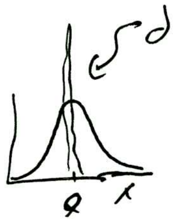
peaks when $x_i = x$

Peaks when $z_e = x_i + y$

peaks when $z_e = x+y$

spreads out connections

$$\boxed{5}\quad r_{lz} = \frac{1}{N}\sum_{ij} \frac{e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2+(y-y_j)^2\right]}}{2\pi\sigma^2}\, \underbrace{\Delta\left(z_l - \phi(x_i,y_j)\right)}_{\Delta(x)}$$



$$\langle z\rangle = \frac{1}{N^2}\sum_{ij,l} e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2+(y-y_j)^2\right]}\, \Delta\left(z_l - \phi(x_i,y_j)\right)\left[z_l - \phi(x_i,y_j) + \phi(x_i,y_j)\right]$$

$$\sum_x \Delta(x)\,x = 0$$

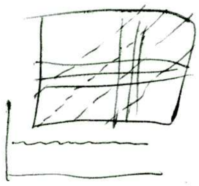
$$\frac{1}{N}\underbrace{\left(\sum_x \Delta(x)\right)}_{1} \phi(x_i,y_j)$$

$$= \frac{1}{N}\sum_{ij} e^{-\frac{1}{2\sigma^2}\left[(x-x_i)^2+(y-y_j)^2\right]}\, \phi(x_i,y_j)$$

$\boxed{6}$ nonlinearity is necessary:

$$r_{ij} = r_{ix} + r_{jy}$$

$$e^{-\frac{1}{2\sigma^2}(x-x_i)^2} \qquad e^{-\frac{1}{2\sigma^2}(y-y_j)^2}$$



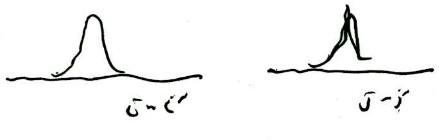
$$r_{lz} = \frac{1}{N^2}\sum_{ij}\left[r_{ix} + r_{jy}\right]\delta\left(z_l - \phi(x_i,y_j)\right) = \frac{1}{N}\sum_{ij} e^{-\frac{1}{2\sigma^2}(x-x_i)^2}\,\delta\left(z_l - \phi(x_i,y_j)\right) + \frac{1}{N}\sum_{ij} e^{-\frac{1}{2\sigma^2}(y-y_j)^2}\,\delta\left(z_l - \phi(x_i,y_j)\right)$$

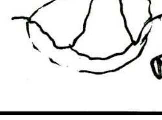
$$\langle z\rangle = \frac{1}{N^2}\sum_{ij}\left[r_{ix}\,\phi(x_i,y_j) + r_{jy}\,\phi(x_i,y_j)\right]$$

$$= \frac{1}{N}\sum_i r_{ix}\,\underbrace{\frac{1}{N}\sum_j \phi(x_i,y_j)}_{\text{not much info}} + \frac{1}{N}\sum_j r_{jy}\,\underbrace{\frac{1}{N}\sum_i \phi(x_i,y_j)}_{''} \sim \text{constant}$$

don't need to multiply

$$r_{ij} = \psi\left(\sum_{i'} W^x_{i,i'}\, r_{i'x} + \sum_{j'} W^y_{j,j'}\, r_{j'y} + \sum_{i'j'} w_{ij,i'j'}\, r_{i'j'}\right)$$





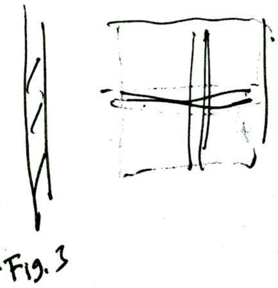
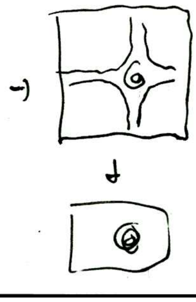
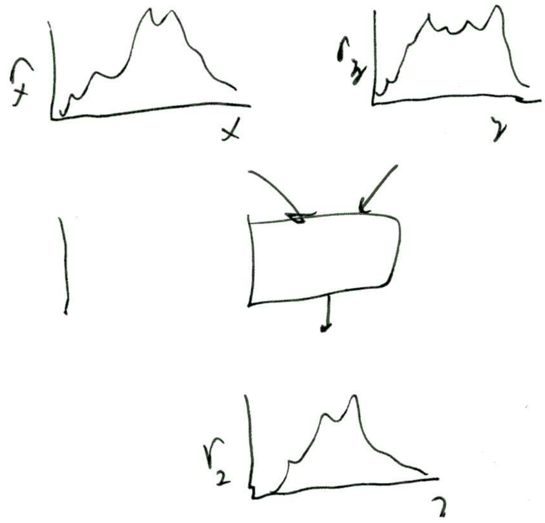


Fig. 3



7



$$I_{in} = -\left\langle \frac{\partial^2 \log P(\mathbf{r}_x, \mathbf{r}_y|\theta)}{\partial\theta^2} \right\rangle$$

$$\int dx\,dy\, P(\mathbf{r}_x, \mathbf{r}_y|x,y)\,P(x,y|\theta)$$

$$\uparrow$$

$$\delta(\theta-\phi(x,y))$$

$$I_{out} = -\left\langle \frac{\partial^2}{\partial\theta^2} \log P(\mathbf{r}_z|\theta) \right\rangle \leftarrow \phi(x,y)$$

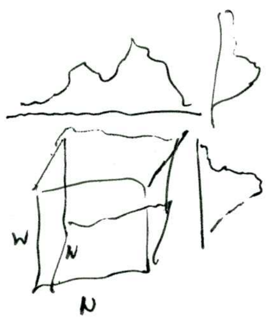
Want I_{out} as close as possible to I_{in}!!!

networks w/ $I_{out}/I_{in} \ll 1$ are __not__ __biologically__ __plausible__!!!

E.G. Sharpening

Critical Points → also: computations
a) throw away info
b) re-represent

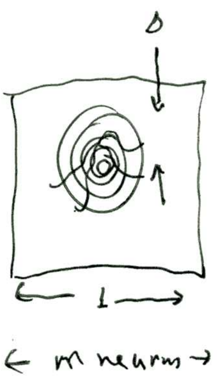
8 curse of dimensionality (?)



$\sim N^3$ neurons. D-dimensions: N^D

- is it really this bad???

Intuition:



\# neurons $\sim \left(\frac{\Delta}{L}\right)^2 M^2$
$\uparrow$
N

$\left(\frac{\Delta}{L}\right)^D M^D \sim N_1$

$M^D \sim N_1 \left(\frac{L}{\Delta}\right)^D$

broad is good!!!

\# neurons in D-dim: $\left(\frac{\Delta}{L}\right)^D M^D$
$\uparrow$
N

$N \sim \left(\frac{L}{\Delta}\right)^D N_1$

narrow in 1 direction helps, but hurts in other directions

Also: every time you add stim dimension, you lose some fraction of neurons.

$\boxed{9}$ $$p(r_{ijk\ldots} \mid x, y, z, \ldots) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{ijk\ldots}\left(r_{ijk\ldots D} - f\left(\frac{x-x_i}{\Delta}, \frac{y-y_j}{\Delta}, \ldots\right)\right)^2\right]$$

$$I_{xx} = -\left\langle \frac{\partial^2 \log p}{\partial x^2} \right\rangle = \left\langle \frac{1}{2\sigma^2} \frac{\partial^2}{\partial x^2} \sum_{ijk\ldots D} \left(\qquad \right)^2 \right\rangle$$

$$= \frac{1}{\sigma^2} \sum_{ijk\ldots} \frac{\partial^2 f}{\partial x^2}\left(\frac{x-x_i}{\Delta}, \frac{y-y_j}{\Delta}, \ldots\right)$$

$$\sum_i \to \int_1^M di \to \int_0^L \frac{di}{dx_i} dx_i = \frac{M}{L}\int dx_i$$

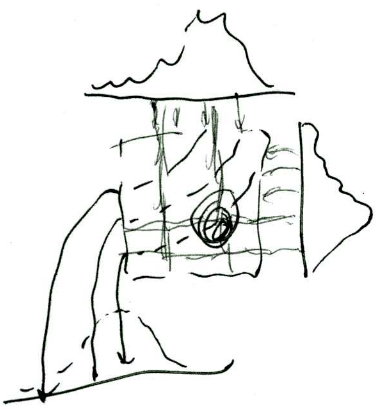
$$I_{xx} = \left(\frac{M}{L}\right)^D \int_0^L dx_i \int_0^L dy_j \ldots \frac{\partial^2 f}{\partial x^2}\left(\frac{x-x_i}{\Delta}, \frac{y-y_j}{\Delta}, \ldots\right)$$

$$\frac{x-x_i}{\Delta} \to x' \qquad dx_i \to \Delta\, dx' \qquad \frac{\partial}{\partial x} \to \frac{1}{\Delta}\frac{\partial}{\partial x'}$$

$$I_{xx} = \left(\frac{\Delta}{L}\right)^D \frac{1}{\Delta^2} M^D \underbrace{\int_0^{L/\Delta} dx'\, dy'\, \partial_{x'x'} f}_{\gamma^D} = \left(\frac{\Delta\gamma}{L}\right)^D \frac{1}{\Delta^2} \underbrace{M^D}_{N}$$

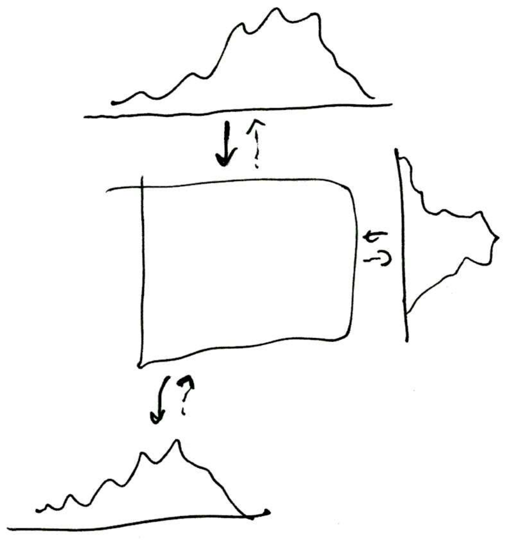
$$N \sim N_1 \Delta^2 \left(\frac{\Delta\gamma}{L}\right)^D$$

$\boxed{10}$ Summary



1. need nonlinearity to get a bump.
2. all the action is in ff connections
3. can compute I_{in}, I_{out}. MUST have $I_{out} \sim I_{in}$
4. Curse of dimensionality isn't too bad: can handle ~ 5-10...? dimensions.
5. no notion of time!!!

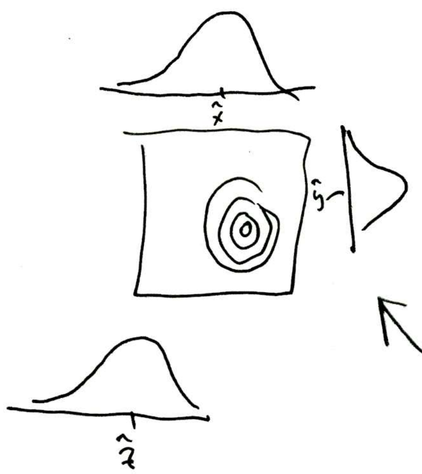
11 What if all variables tell you something? e.g.: auditory + vision + eye position



add feedback

~~Main~~ approach: consider network dynamics

- noisy hills act as I.C.
- dynamics causes system to evolve onto smooth hill(s)



- readout is easy!!!
- ANN (2-D)

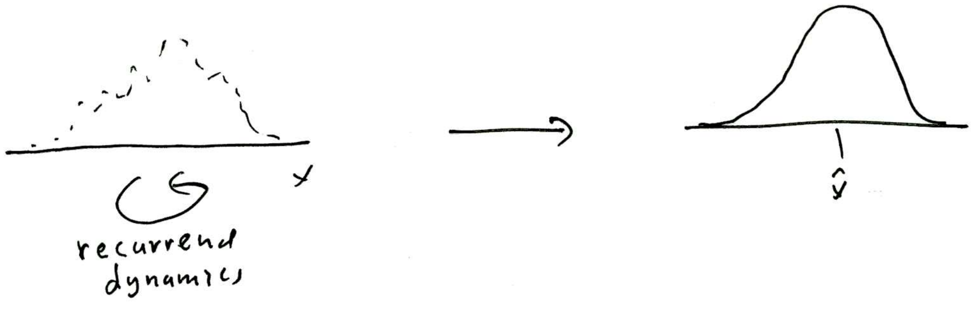
$z = f(x_1, x_2 .. x_D)$ D-dim attractor

half-shifting tuning curve

~~[illegible]~~

12 - ~~[illegible]~~

- if we understand the 1-D case, we understand everything.

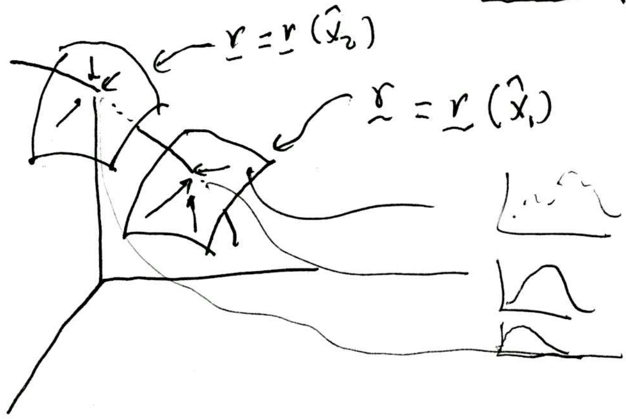


[different dots lead to different hills]

[explain why this is a 1-D ANN]

13

$\hat{x} = \hat{x}(\underset{\sim}{r}) \Rightarrow \underset{\sim}{r} = \underset{\sim}{r}(\hat{x})$ ← N-1 dim. manifold



- estimation ⇔ finding which manifold we're on!

$\dot{\underline{x}} = \underline{F}(\underline{x})$ [e.g. $\dot{x}_i = \psi(\sum_j w_{ij} x_j)$]

$\frac{d}{dt}\hat{x}(\underline{r}) = \frac{\partial \hat{x}}{\partial \underline{r}} \cdot \dot{\underline{r}} = \underset{\sim}{F} \cdot \frac{\partial \hat{x}}{\partial \underline{r}}$

gradients } perpendicular
dynamics

- this specifies dynamics - all we have to do is find a network that implements it!!!

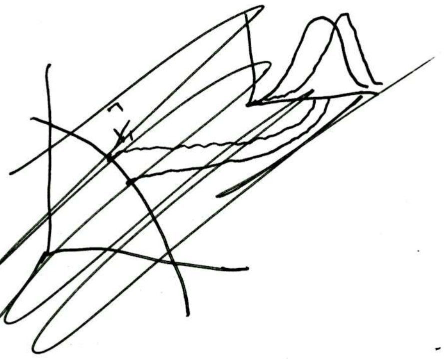
14 Formal Analysis

$\dot{r}_i = \frac{dr_i}{dt} = \cancel{\quad} F_i(\underline{r})$ [$\dot{r}_i = F(\sum_j w_{ij} r_j)$]

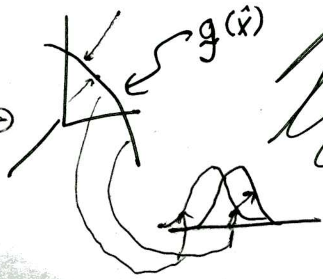
$r_i(0) = f_i(x) + \cancel{\quad} \eta_i$ ← $\langle \eta_i \eta_j \rangle = R_{ij}$

← $f(x - x_i)$

$r_i(t=\infty) = g_i(\hat{x}) = g(x - \hat{x})$



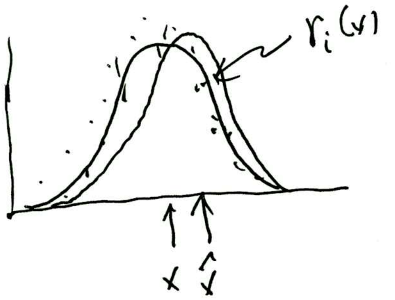
$\boxed{F_i(g(\hat{x})) = 0}$ ⟷



goal

$\boxed{\text{Compute } p(\hat{x}|x)}$

$\boxed{15}$ $f = g$



[show multiple $f(\hat{x}-x_i)$]

this is a function of the network.

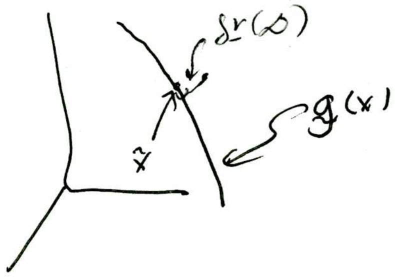
goal: want $\mathrm{Var}[\hat{x}] = \dfrac{1}{I}$

$$I = + \sum_{ij} f_i'^{T}(x-x_i) R^{-1}_{ij} f_j'(x-x_j) + \frac{1}{2}\mathrm{tr}\left\{R^{-1}R' R^{-1} R'\right\}$$

Approach: linearize

$\boxed{16}$ $\dot{\underset{\sim}{r}} = \underset{\sim}{F}(\underset{\sim}{r})$

$\underset{\sim}{r}(t) = g(\hat{x}) + \delta \underset{\sim}{r}(t)$



$$\delta \dot{\underset{\sim}{r}} = F(g(\hat{x}) + \delta r)$$

$$= \cancel{F(g(\hat{x}))}^{\,0} + \frac{\partial F}{\partial g} \cdot \delta r$$

$$= \underset{\sim}{J}(\hat{x}) \cdot \delta r$$

$$\delta \underset{\sim}{r}(t) = e^{\underset{\sim}{J}(\hat{x})t} \cdot \delta \underset{\sim}{r}(0) = e^{\underset{\sim}{J}(\hat{x})t} \cdot [\underset{\sim}{r}(0) - g(\hat{x})]$$

$$= e^{\underset{\sim}{J}(\hat{x})t} \cdot [\underset{\sim}{f}(x) + \eta(x) - g(\hat{x})]$$

$$\underset{\sim}{J}(x) = \sum_k \lambda_k \underset{\sim}{v}_k \underset{\sim}{v}_k^{+} \qquad \underset{\sim}{J} \cdot \underset{\sim}{v}_k = \lambda_k \underset{\sim}{v}_k \qquad \underset{\sim}{v}_k^{+} \cdot \underset{\sim}{J} = \lambda_k \underset{\sim}{v}_k^{+}$$

$$e^{t \sum_k \lambda_k \underset{\sim}{v}_k \underset{\sim}{v}_k^{+}} = \sum_k e^{\lambda_k t} \underset{\sim}{v}_k \underset{\sim}{v}_k^{+} \rightarrow \underset{\sim}{v}_0 \underset{\sim}{v}_0^{+}$$

$$\delta \underset{\sim}{r}(\infty) = \underset{\sim}{v}_0^{(\hat{x})} \underset{\sim}{v}_0^{+(\hat{x})} \cdot [\underset{\sim}{f}(x) + \eta(x) - g(\hat{x})] \Rightarrow$$

[17] Algebra algebra algebra...

can compute $\delta x = \hat{x} - x = \dfrac{V_0^+(x)\cdot \xi(x)}{V_0^+(x)\cdot f'(x)}$

$\langle \delta x \rangle = 0$

⟨33⟩

$\langle \delta x^2 \rangle = \dfrac{V_0^+ \cdot R \cdot V_0^+}{(V_0 \cdot f')^2}$

Optimal (min. variance) when $\boxed{V_0^+ = R^{-1}\cdot f'}$

everything about the network is in here.

$\langle \delta x^2 \rangle = \dfrac{1}{f'\cdot R^{-1}\cdot f'} = \dfrac{1}{I_{network}}$

[18] $I_{network} = f'\cdot R^{-1}\cdot f'$; $I = f'\cdot R^{-1}\cdot f' + \frac{1}{2}\,\mathrm{tr}\{R^{-1}\cdot R'\cdot R^{-1}\cdot R'\}$

can be $O(1)$ — $O(N)$

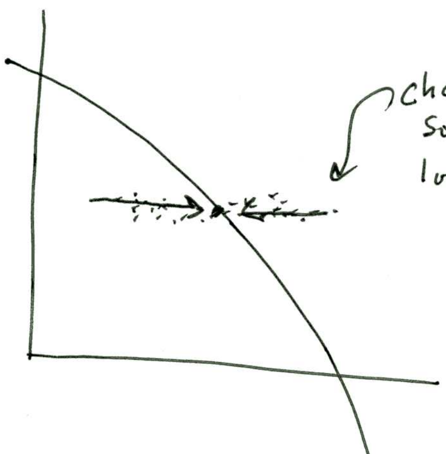
optimal when $R' = 0$!!!

$x_i = \mu + \xi_i$

$\hat{x} = \frac{1}{N}\sum_i x_i$

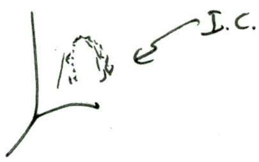
$\langle \xi \rangle = 0$

$\langle \xi^2 \rangle = \frac{\sigma^2}{N}\left[1+(N-1)\rho\right]$



choose network so trajectories look like this.

what can go wrong

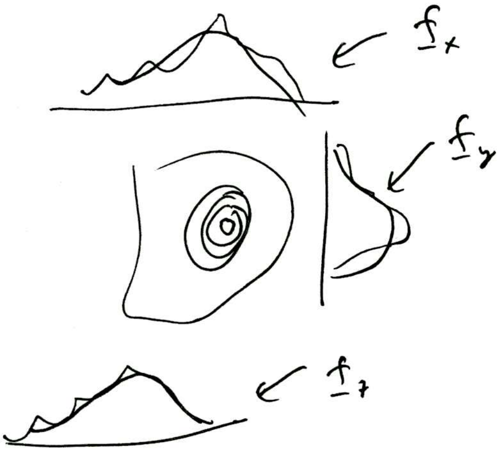


$R' \neq 0$

$\boxed{10}$

$$V_0^+ = \begin{bmatrix} \boxed{x} & & \\ & \boxed{y} & \\ & & \boxed{z} \end{bmatrix} \begin{bmatrix} f_x' \\ f_y' \\ f_z' \end{bmatrix}$$

$\boxed{19}$ more interesting cases



$$\underline{V}_{0x}^+ = \underline{\underline{R}}_x^{-1} \cdot \underline{f}_x$$

$$\underline{V}_{0y}^+ = \underline{\underline{R}}_y^{-1} \cdot \underline{f}_y$$

$$\underline{V}_{0z}^+ = \underline{\underline{R}}_z^{-1} \cdot \underline{f}_z$$

what happens to $\underline{V}_0^+$ if we ~~scale~~ change amplitude?

$$\underline{f} \to a\underline{f}$$

$$\underline{\underline{R}} \to a\underline{\underline{R}} \leftarrow \text{Poisson (only)}$$

$$\underline{\underline{R}}^{-1} \cdot \underline{f} \to \underline{\underline{R}}^{-1} \cdot \underline{f} \leftarrow \text{same network!!!}$$

$\boxed{\text{DO } 20 + 21 + 22 \text{ first!!}}$

$\boxed{\text{Non-Poisson: need a new network for each amplitude}}$

$\boxed{20}$ - Cue combination

$$p(\underset{\sim}{r}_1|x) \sim e^{-\frac{1}{2}(x-\hat{x}_1(\underline{r}))^2/\sigma_{x_1}^2}$$

$$p(\underset{\sim}{r}_2|x) \sim e^{-\frac{1}{2\sigma_{x_2}^2}(x-\hat{x}_2(\underline{r}))^2}$$

$$p(x|\underline{r}_1,\underline{r}_2) = p(\underline{r}_1|x)p(\underline{r}_2|x)p(x) \sim e^{-\frac{1}{2}\left[\frac{(x-\hat{x}_1(\underline{r}))^2}{\sigma_{x_1}^2} + \frac{(x-\hat{x}_2(\underline{r}))^2}{\sigma_{x_2}^2}\right]}$$

$$\hat{X}_{ML} = \frac{\hat{x}_1/\sigma_{x_1}^2 + \hat{x}_2/\sigma_{2x}^2}{\frac{1}{\sigma_{1x}^2} + \frac{1}{\sigma_{2x}^2}}$$

optimal estimate depends on variance!!!

$p(\underline{r}_x|x)$
$p(\underline{r}_y|y)$
$p(\underline{r}_z|z)$

$$p(x,y,z|\underset{\sim}{r}) \sim \cancel{e^{-\frac{1}{2}\left[\frac{(x-x_i)^2}{\sigma_x^2} + \frac{(y-y_i)^2}{\sigma_y^2} + \frac{(z-z_i)^2}{\sigma_z^2}\right]}} \quad \phi(x,y)$$

$$\exp\left[-\frac{1}{2}\left(\frac{(x-\hat{x})^2}{\sigma_x^2} + \frac{(y-\hat{y})^2}{\sigma_y^2} + \frac{(z-\hat{z})^2}{\sigma_z^2}\right)\right] \quad \phi(x,y)$$

21 Poisson:

$$P(\underset{\sim}{r}|x) = \prod_i \frac{f(x-x_i)^{r_i} \, e^{-f(x-x_i)}}{r_i!}$$

$$\propto e^{-\sum_i f(x-x_i) + r_i \log f(x-x_i)}$$

$$f(x-x_i) = e^{-\frac{1}{2\sigma^2}(x-x_i)^2}$$

$$\propto e^{-\frac{1}{2\sigma^2}\sum_i r_i (x-x_i)^2}$$

$$\propto e^{-\frac{N}{2\sigma^2}\left[\bar{r}x^2 - 2x\overline{rx} + \overline{rx^2}\right]}$$

$$\propto e^{-\frac{N\bar{r}}{2\sigma^2}\left[x - \frac{\overline{rx}}{\bar{r}}\right]^2}$$

Ind. Pom

$I = \sum \frac{f'^2}{f}$

$f \to af$

$I \to aI$

$Var \sim \frac{1}{a}$

$a \sim \frac{1}{Var}$

$var(x) = \frac{\sigma^2}{N\bar{r}}$ $\quad \bar{r} \sim f \quad \Rightarrow \quad Var \sim \frac{1}{amplitude}$, $\frac{1}{var} \sim$ amplitud $\to$ just right!!

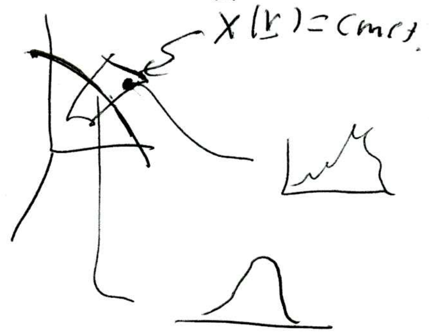
22 Poisson encodes variance through amplitude!!!

- ANN's automatically use this fact!!!

23 Summary

Forward model: $p(\underline{r}|x) \to p(x|\underline{r}) \propto p(\underline{r}|x)\,p(x)$

$$\Rightarrow \hat{x}(\underline{r})$$



- Network figures out which manifold you live on!!!
- ~~Nem~~ Formalism allows you to find network eqns.
- May or may not be biologically plausible.

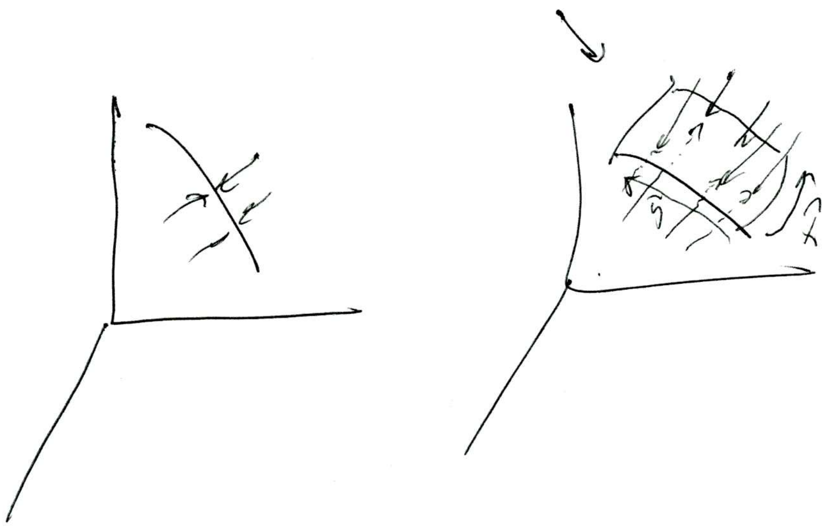
Generalizes: $p(\underline{r}|x,y,z) \neq p(\underline{r}_x|x)\,p(\underline{r}_y|y)\,p(\underline{r}_z|z)$

$$\to p(x,y,z|\underline{r}) \propto p(\underline{r}|x,y,z)\,p(x,y,z)$$

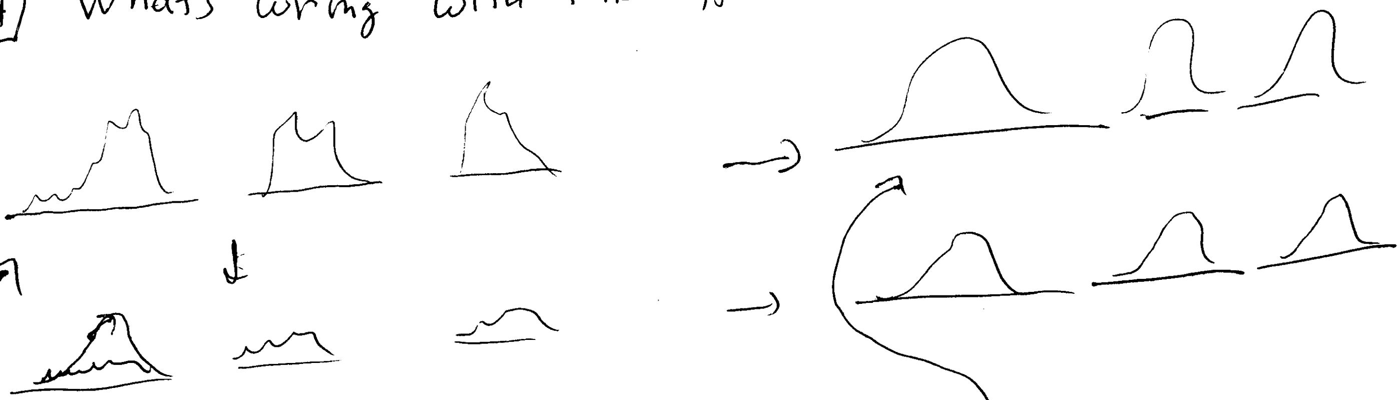
$$\delta(z-\phi(x,y))$$

$$\Rightarrow \hat{x}(\underline{r}),\ \hat{y}(\underline{r}),\ \hat{z}(\underline{r})$$

r



24 What's wrong with this approach?



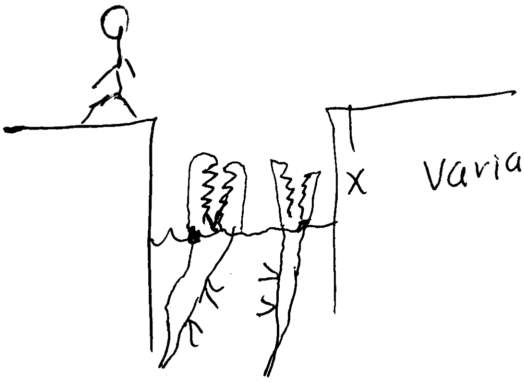
here we have a forward model: $p(\underline{r}|x) \Rightarrow p(x|\underline{r})$

$\Rightarrow$ information about variance

~~$p(\underline{r}|x)$~~ no info about actual variance.

We have ML estimator - as good as it gets - so why do we care?

25



variance is important!!!

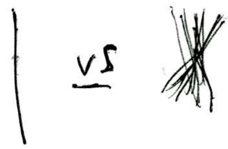
To make decisions, we need to estimate variance!!!

$\boxed{2.6}$ $p(\underline{r}|x) \Rightarrow p(x|\underline{r}) \propto p(\underline{r}|x)\,p(x)$

$\Rightarrow Var[x|\underline{r}] = \sigma_x^2(\underline{r})$

$\sigma_x^2 = \sigma_x^2(\underline{r}) \Rightarrow$ all the same formalism!!

Alternatively: $I = \sum_i \frac{f'^2(x-x_i)}{f'(x-x_i)}$



e.g. $f(x) = a \exp\left[\frac{\cos(x)-1}{\sigma^2}\right]$

$\Rightarrow I = \tilde{I}(a,\sigma)$

could compute $p(I|\underline{r}) \propto p(\underline{r}|I)\,p(I)$

$= \int da\,d\sigma\; p(\underline{r}|a,\sigma)\; \underbrace{p(a,\sigma|I)/p(I)}$

$p(I|a,\sigma)\,p(a,\sigma)$

$\uparrow$

$\delta(I-\tilde{I}(a,\sigma))$

$\hat{I} = \langle I \rangle = \int dI\; I\, p(I|\underline{r})$

$= \hat{I}(\underline{r}) \Rightarrow$ all the same formalism!!

$\boxed{27}$ General Bayesian Inference

θ : low level features

know $p(\theta|y)$
$\quad$ high level objects

know $p(\underline{r}|\theta)$

$\Rightarrow \; p(y|\underline{r}) \propto \cancel{\int d\theta \, p} \; p(\underline{r}|y)\, p(y)$

$= \int d\theta \; p(\underline{r}|\theta)\, p(\theta|y)\, p(y)$

$\Rightarrow \; \hat{y}(\underline{r}) \Leftarrow \underline{r}(\hat{y})$

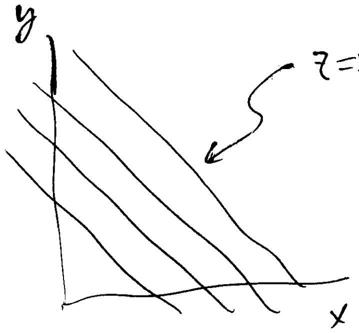
$\Rightarrow$ network equation:

$\dot{\underline{r}} = \underline{F}(\underline{r})$ $\quad \underline{F} \cdot \dfrac{\partial \hat{y}}{\partial \underline{r}} = 0$

Summary

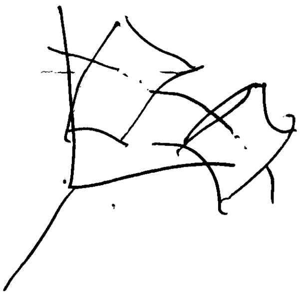
$\boxed{28}$ Main point: computation correspond to ~~[illegible]~~ finding which invariant manifold you're on.

Simple example: addition



z=x+y. which manifold are we in?

$$p(\underline{r}|x) \Rightarrow p(x|\underline{r}) \Rightarrow \hat{x}(\underline{r}) \Rightarrow \bar{\underline{r}}(t)$$



$$p(y|\underline{r}) \propto \int d\theta \, p(\underline{r}|\theta)\, p(\theta|y)\, p(y)$$

$$\Rightarrow \hat{y}(\underline{r})$$

29 Caveats.

1. Noise

2. $\vec{y}(\underline{v})$ can be hard to calculate

3. $\dot{\underline{v}}$ may not be biologically plausible

4. ANN's: end of line because we've lost info about variance. [variance encoded in amplitude + width]

— want to develop other methods!!!
[(4) is killer]