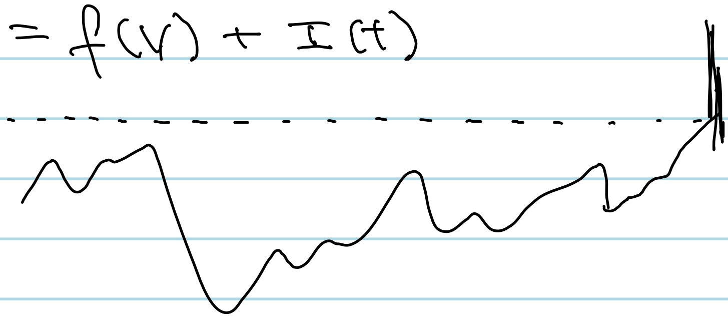
"ESCAPE NOISE" MODELS OF NEURAL FIRING AND GENERALIZED LINEAR MODELS.

- Start with... 1-D model for voltage

$$\dot{V} = f(V) + I(t)$$

V(t)  Integrate + fire w/ "hard" threshold... Spike when cross V_{th}.

- Try this experimentally...
  Fig 9.3 Gerstner et al text.

  See... tiny black histogram... is range

  almost invisible in plot of V values at timestep before measured spike.

  FACT: This small histogram is spread out.
  $\longrightarrow$
  Do not have single V_{th}.

Instead... define: $g(V - V_{th}) = \Pr(\text{fire} \mid V)$

loosely, spike "at end of" that timestep.

Gerstner et al Fig. 9.3 → deriving the "escape rate" of firing from data,

Plot in grey + (tiny) black histograms:

$p(v \mid \text{fire}) \cdot p(\text{fire})$ BLACK

$p(v \mid \text{no fire}) \cdot p(\text{no fire})$ GREY

} for same Δt from V measurement.

WANT! $$p(\text{fire} \mid v) = p(v \mid \text{fire}) \cdot \frac{p(\text{fire})}{p(v)}$$

Now: $$p(v) = p(v \mid \text{fire})\, p(\text{fire}) + p(v \mid \text{no fire})\, p(\text{no fire})$$

So... $$p(\text{fire} \mid v) = \frac{\text{BLACK}(v)}{\text{BLACK}(v) + \text{GREY}(v)}$$

See... $$f(v - v_{th}) \approx \frac{1}{\tau_0} \exp\left(\beta\,(v - v_{th})\right)$$

"Soft threshold"

Recover hard threshold as $\beta \to \infty$.

Say: spiking is a point process with intensity $f(v(t) - \Theta)$, ie $\text{Proba}(\text{spike} \in [t, t+dt])$

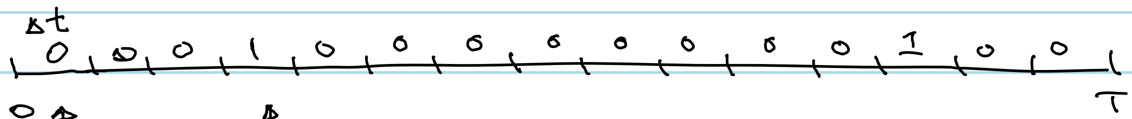
$$\approx f(v(t) - v_{th})\, dt$$

$$= \rho(t)\, dt,$$ in small dt limit.

- Further work relating I+F and point process in text: Arrhenius approximation to S.D.E., Plesser ('06'?)

IN FUTURE ADD MORE HERE !!!

SPIKE TRAIN: Temporal Sequence of 0's and 1's. Call this object $\mathcal{D}$ (for data)



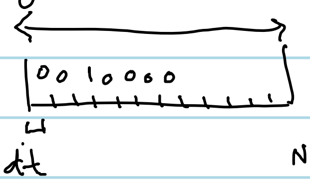
Proba ≈ $[1 - p(t_i)\Delta t]$ $\quad$ $p(t_k)\cdot\Delta t$

HAVE A SPIKE / DO NOT HAVE A SPIKE

$\Rightarrow$ **DEFINE: Likelihood of a spike train (Sec. 9.2)**

- Define # spikes in each bin = n_t

n_t follows Poisson distribution



$N = \frac{\Delta t}{dt}$ bins

$$p(n_t \; .) = \binom{N}{n_t} [dt \cdot p(t)]^{n_t} [1 - p(t)\cdot dt]^{(N - n_t)}$$

(Binomial distribution)

$$= e^{-p(t)\Delta t} \cdot \frac{(p(t)\Delta t)^{n_t}}{n_t!} \quad \text{in } \lim dt \to 0.$$

In particular,

$$\begin{cases} p(n_t = 0) = e^{-p(t)\Delta t} \\ p(n_t = 1) = e^{-p(t)\Delta t} \cdot p(t)\Delta t \end{cases}$$

And, for future use:

$$\log p(n_t) \; . = -p(t)\Delta t + n_t \log(p(t)\cdot\Delta t) - \log(n_t!)$$

Next, given $\rho(t)$,

Likelihood of whole spike train $= P(D|\rho(t)) = \prod_t p(n_t)$

Log-Likelihood " $= \log P(D|\rho(t)) = \sum_t \log p(n_t)$

$$= -\sum_t \rho(t)\Delta t + n_t \log \rho(t) + \underbrace{n_t \log \Delta t - \log(n_t!)}_{\text{INDEP. OF } \rho(t)}$$

$$= -\sum_t \rho(t)\Delta t + n_t \log \rho(t) + c$$

In limit $\Delta t \to 0$,

$$\log P(D|\rho(t)) = -\int_0^T \rho(t)\, dt + n_t \log \rho(t) + c$$

$n_t = 1$ for $t \in \{t_j\}$, set of spike times

$n_t = 0$ $t \notin \{t_j\}$

$$\boxed{\log P(D|\rho(t)) = -\int_0^T \rho(t)\, dt + \sum_j \log \rho(t_j) + c} \quad (10.22)$$

PARAMETER ESTIMATION FOR SPIKING MODELS

<u>Goal:</u> Given D, and form of $\underbrace{p(t)}_{\theta}$ that depends on parameter list θ of spiking model.

<u>E.g.</u> $p(t) = f\Big(\underbrace{V(t)}_{\text{depends on }\theta\text{ via ODE solution}} - V_{th}\Big)$, above

then $p(D \mid p(t)) = p(D \mid \theta)$

<u>Find</u> most likely parameters θ, ie

$$\underset{\theta}{\operatorname{argmax}}\ p(\theta \mid D) = \underset{\theta}{\operatorname{argmax}}\ \frac{p(D \mid \theta) \cdot p(\theta)}{p(D)}$$ Bayes Rule

$$= \underset{\theta}{\operatorname{argmax}}\ p(D \mid \theta) \cdot p(\theta)$$

<u>For</u> "Flat prior case: $p(\theta) = \text{const.}$"

$$= \underset{\theta}{\operatorname{argmax}}\ p(D \mid \theta)$$

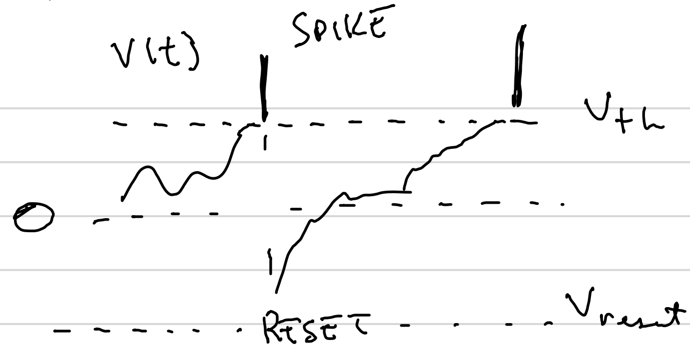
For later:

$$= \underset{\theta}{\operatorname{argmax}}\ \log p(D \mid \theta)$$

PARAMETERIZING SPIKING NEURON MODELS VIA $\vec{\theta}$

(6.4) Gerstner et al text: SPIKE RESPONSE MODEL (SRM)

Think...



I+F Model $\qquad \tau \dot{V} = -V + I(t)$

Before spike, sol^{n} $V(t) = \int_0^{\text{Time to last spike}} e^{-s/\tau} I(t-s)\,ds$, together with:

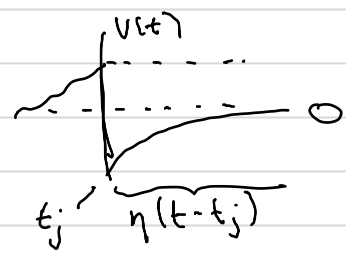
After spike: $V(t) \longrightarrow V(t) - (V_{th} - V_{reset})$, and decay $\to 0$

SRM Captures these effects via:

$$V(t) = \int_0^{\infty} \underbrace{K(s)}_{\text{general filter}} I(t-s)\,ds \;+\; \sum_{t_j} \underbrace{\eta(t-t_j)}_{\text{After-spike voltage jump}}$$

For IF rule above,

$\eta(t-t_j) = -(V_r - V_{th})\, \underbrace{H(t-t_j)}_{\text{Heaviside}} \cdot e^{-(t-t_j/\tau)}$

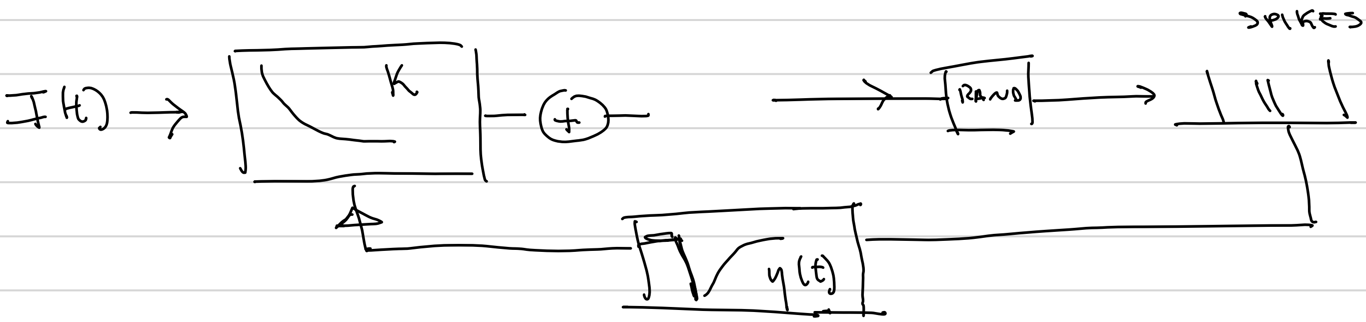


Some differences from I+F... (1) ALLOW .... post-spike effects to build up across several spikes.

(2) $\longrightarrow$ Upper limit on $K(s)$ is ∞... allow stimulus history to impact voltage across several spikes.

BUGS or FEATURES? Depends on how orig. gating variables evolve!

N.B. 6.4.1 of Gerstner text: Can incorporate time-dep. threshold into η: shift voltage relative to fixed V_{th}.



SPIKES
I(t)
K
+
RAND
y(t)

For SRM, $\Theta = \{ K(s), \eta(s) \}$.

Discrete time formulation: (10.2.1, Gerstner)

Time bins dt
Limit $K(s)$, $\eta(s)$ to be supported over $J \cdot dt$ time units.

$$\vec{\Theta} = \{ K(dt), \ldots, K(J \cdot dt), \eta(dt), \ldots, \eta(J \cdot dt) \}$$

Defining...

$$\vec{x}_t = \left(I_{t-dt} \cdot dt, \ldots, I_{t-J \cdot dt} \cdot dt, n_{t-dt}, \ldots, n_{t-J \cdot dt} \right)$$

0 or 1 spike count

$$p(t) = f(\vec{x}_t \cdot \vec{\Theta}).$$

Notation: $\vec{\Theta} = \vec{k}$ in text

$$p(t) = f(\vec{x}_t \cdot \vec{k})$$

Then...

$$\log p(D \mid \vec{\Theta}) = c + -\sum_t f(\vec{x}_t \cdot \vec{k}) \cdot dt + \sum_t n_t \log f(\vec{x}_t \cdot \vec{k})$$

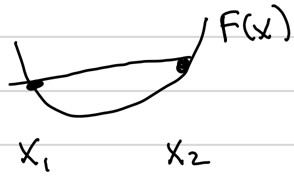
(10.32)

Aside: Convexity (Boyd + Vandenberghe, Convex optimization)

Consider $F(\vec{x})$, scalar-valued f^n.

F is convex if ... $F(\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2) \leq \lambda F(\vec{x}_1) + (1-\lambda) F(\vec{x}_2)$

for $\lambda \in [0,1]$



- Equivalent condition:
  F is convex IFF Hessian Matrix $\left[\left[\frac{\partial^2 F}{\partial x_i \partial x_j} \right]\right]$ is pos. def.

- Ex. - $\text{Exp}(\vec{x})$ is convex.
  - $\vec{k} \cdot \vec{x} + c$ (linear - affine) is convex
  - If $F(x)$ is convex then so is $F(\vec{k} \cdot \vec{x} + c)$
  - Sum of convex functions is convex.

Consequence. If F convex,
For unconstrained optimization problem $\underset{x}{\text{argmin}}\, F(x)$,

any point x_0 s.t. $\nabla F(x_0) = 0$ is global min.

$\rightarrow$ Can find a global optimum, if exists, via grad. descent type methods.

F is concave if $-F$ is convex. Same as above . . .

if F is concave, ... is global maximum.

$$\log p(D|\vec{\theta}) = c + -\underbrace{\sum_t f(\vec{x}_t \cdot \vec{k}) \cdot dt}_{\text{concave}} + \sum_t n_t \underbrace{\log f(\vec{x}_t \cdot \vec{k})}_{\text{concave + convex (linear)}}$$

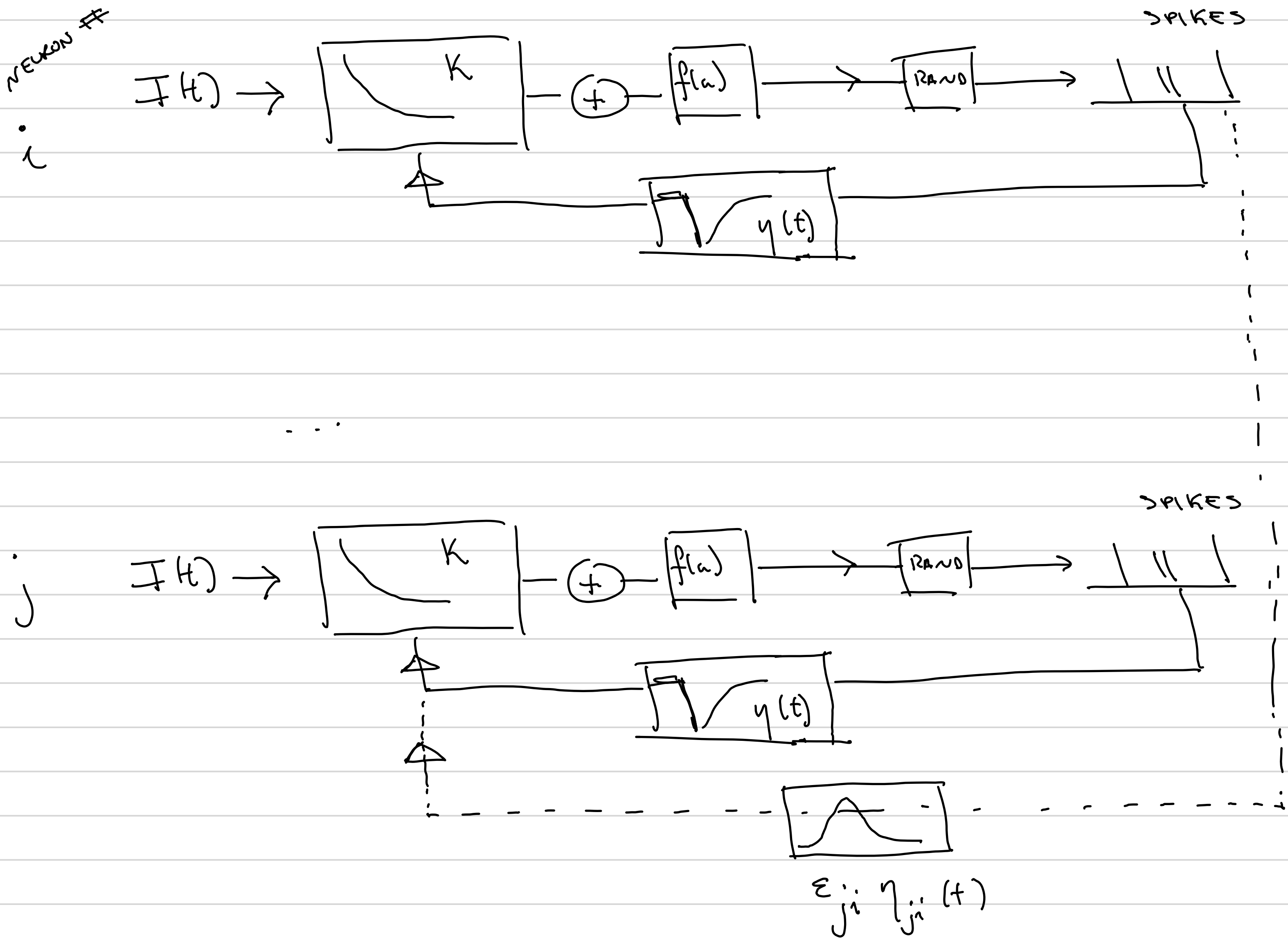
Say $f(x) = \exp(x)$. Then...

or... $f(x)$ concave + log convex (same answer).

→ $\log p(D|\vec{\theta})$ is <u>concave</u> in $\vec{k}$.

AND... easily computable gradient + Hessian ⟶ standard algorithms for minimization eg. Newton - or Newton - Raphson.

Allowing more neural interactions...



Redefine $\vec{X}_t = (\vec{I}, \vec{n}_1, \ldots, \vec{n}_N)$

neuron 1

$\vec{\theta}_j := (\vec{K}_j, \vec{\eta}_j, \varepsilon_{j1}\vec{\eta}_{j1}, \varepsilon_{j2}\vec{\eta}_{j2}, \ldots)$

AND... problem STAYS CONCAVE

WOAH! Lots of parameters. Protect against overfitting ...

(1) Re-introduce prior; to enforce "smooth" filters or other CONSTRAINTS

$$\log\left[p(\vec{\theta}|D)\, p(\vec{\theta})\right] = \underbrace{\log p(\vec{\theta}|D)}_{\text{concave}} + \underbrace{\log p(\vec{\theta})}_{\substack{\text{if this is}\\ \text{concave ...}\\ \text{whole problem stays}\\ \text{concave.}}}$$

(2) Allow for parameterized form of filters, eg via Basis Functions.

[Shizhe / Witten / Shojaie; Pillow et al 2008] e.g.,

$$k(s) \longrightarrow \sum_{j=1}^{B} \alpha_j \beta_j(s) \qquad \text{for } B \text{ basis vectors, } B \ll J$$

$$\vec{x}_t \longrightarrow \left(\int I(t-s)\, \beta_1(s)\, ds, \ldots, \int I(t-s)\, \beta_B(s)\, ds, \ldots \right)$$

$$\vec{k} \to (\alpha_1, \ldots, \alpha_B, \ldots)$$

and likewise for "η" spike interaction filters.

SAME form of $p(\theta|D) \longrightarrow$ stays concave!

CONNECTIONS TO RECEPTIVE FIELD MODELS: (11.2.1, Text)

- Allow stimulus to have spatial component: $I(t) \to I(x, y, t)$

  + filters $K(s) \to K(x, y, s)$

- (often): give complex spatiotemporal gaussian noise I to fit filters
- The resulting $K(x,y,s)$ is est. of receptive field...

  Paninski et al 2003. Compare with "LN" models

$$\rho(t) = N\left(\underset{\substack{\uparrow \\ \text{fit by "SPIKE TRIGGERED AVERAGE."}}}{L(x,y,s)} \cdot I(x,y,s) \right)$$

  Result:

  $K(x,y,s)$ is equivalent to a "weighted" spike triggered average, with $N = f$.

  via $f'(\cdot)$

- So... can be similarly used to discover receptive fields.

- CONNECTION TO RIEKE LECTURE:

  - Identify receptive fields for retinal ganglion cells.
  - Q: Get better (more predictive on "test" data) fit for an additional nonlinear stage of processing...

$I(x, y, s) \longrightarrow N^{upstream}\left(I(x, y, s)\right)$

- $N^{upstream}$ was postulated: can it be <u>FIT</u>?

  Freeman + Simoncelli – Variational approach

  OR, fit FFW cascade of GLM models ?
  SRM

DECODING (Ch. 11.3, Text). From spikes to stimulus.

- Fit $\vec{k}$ on "test" set of data $\vec{X}$.
- For held-out "test" data $\vec{X}$, consider

$$p\big(D \mid \{\underbrace{\vec{I}}_{\text{Unknown stim}}, \underbrace{\vec{n}}_{\text{spikes, known}}\}, \underbrace{\vec{k}}_{\text{known filters, etc.}}\big) \quad \text{as function of } \vec{I}.$$

By exact same arguments as above, this is concave function of $\vec{I}$.

So... do Max-Likelihood DECODING / PREDICTION of $\vec{I}$.

(Fig. 11.12, Text).