

Special topics notes 2008

Chaitanya Ekanadham

1 Neural coding basics

1.1 Definitions

1. **Response function** $\rho(t) = \sum_j \delta(t - t_j)$ where t_j are the spikes
2. **Firing rate** $r(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \rho(s) \rangle ds$
3. **Spike count** $n = \int_0^T \rho(t) dt$
4. **Average firing rate** $\tilde{r} = \frac{1}{T} \langle \int_0^T \rho(t) dt \rangle = \frac{\langle n \rangle}{T}$
5. **Tuning curve** $f(s) = \text{avg firing rate given a parameterized stimulus } s$
(typically plotted as a histogram and then fit to some functional form)
6. **Stimulus/Response correlation** $Q_{rs}(\tau) = \frac{1}{T} \int_0^T r(t) s(t + \tau) dt$
7. **Spike-triggered average** $C(\tau) = \langle \frac{1}{n} \int_0^T \rho(t) s(t - \tau) dt \rangle \approx \frac{1}{\langle n \rangle} \int_0^T r(t) s(t - \tau) dt$
8. **Stimulus autocorrelation** $Q_{ss}(t) = \frac{1}{T} \int_0^T s(t) s(t + \tau) dt$
9. **Response autocovariance/autocorrelation** $Q_{\rho\rho}(t) = \frac{1}{T} \int_0^T \langle (\rho(t) - \tilde{r})(\rho(t + \tau) - \tilde{r}) \rangle dt$
10. **Rate code** - code in which all stimulus information is conveyed via the firing rate function $r(t)$ which does not vary at faster timescales than the stimulus $s(t)$ (Note: a rate code can still be very spike-time-dependent if the stimulus is very fast-varying thus inducing a fast-varying rate - the key is that if the stimulus is slow-varying than the rate function should not be so time-sensitive).
11. **Pulse/temporal code** - code in which information is carried in variations of the rate function that are on shorter timescales than the fastest stimulus variation.

12. **Independent spike assumption** - one neuron generates spikes independently of one another
13. **Correlation code** - spikes are not generated independently
14. **Independent neuron assumption** - neurons generate spike independently of one another
15. **Nonhomogeneous Poisson process** $N(t)$ with rate $\lambda(t)$:
 $N(0) = 0$ and $\forall t, s, \geq 0, N(t + s) - N(t) \sim \text{Poisson}(\int_t^{t+s} \lambda(r) dr)$.
16. **Fano factor** of a random process N_t over a time window $[0, T]$: $F(N_t) = \frac{\hat{\sigma}_{N_T}^2}{\hat{\mu}_{N_T}}$ (usually applied to spike counts)
17. **Coefficient of variation** of a random variable X : $C_v(X) = \frac{\hat{\sigma}_X}{\hat{\mu}_{N_T}}$ (usually applied to ISI)
18. **Volterra-Weiner series** given a stimulus $s(t)$:
 $\hat{r}(t) = r_0 + \int_0^T D_0(\tau) s(t - \tau) d\tau + \int_0^T D_1(\tau_1, \tau_2) s(t - \tau_1) s(t - \tau_2) d\tau_1 d\tau_2 + \dots$
19. **Entropy** of a probability distribution P is $-\sum_j P(j) \log P(j)$ i.e. the amount of 'uncertainty/variability.'
20. **Mutual information** between X and Y is $H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) = KL(P_X \times P_Y || P_{X,Y})$ i.e. the amount of uncertainty/variability in one variable that originates from the other (think of conditional entropy as the 'intrinsic variability' since this is a measure of uncertainty when the other variable is fixed or given).

1.2 Facts/derivations to know

1.2.1 Neural coding and spike train statistics

1. **Measuring $r(t)$** : divide $[0, T]$ into timebins and trial-average number of spikes in each bin to give discretized $r(t)$. Can generalize this by using a window function $\phi(\tau)$ to weight the affect of each spike: $r(t) =$

$$\sum_i \sum_{t_j} \phi(t - t_j^{(i)}) = \sum_i \int_0^T \rho^{(i)}(\tau) \phi(t - \tau) d\tau$$

Typical choices for $\phi(\tau)$ are indicator around 0, gaussian around 0, rectified decaying exponential (for causality need $\phi(\tau) = 0 \forall \tau < 0$).

2. **$r(t)$ is like a trial-averaged density of $\rho(t)$:** $\forall t \in [0, T]$ and $\forall h(s)$ nice, $\int_0^T h(s) \langle \rho(t-s) \rangle ds = \int_0^T h(s) r(t-s) ds$

3. **The spike-triggered average is reverse stimulus/response correlation scaled by avg firing rate:** $C(\tau) = \frac{Q_{rs}(-\tau)}{\bar{r}}$

4. **Measuring response autocorrelation:** Divide $[0, T]$ into timebins $I_m = [(m-0.5)\Delta t, (m+0.5)\Delta t]$. Define N_m be the number of spike pairs (including self-pairs) separated by time $t \in I_m$, and let $H_m = \frac{N_m}{T} - \frac{n^2 \Delta t}{T^2}$. The autocorrelation $Q_{\rho\rho}(\tau)$ can be estimated by looking at $\lim_{\Delta t \rightarrow 0} \frac{H_{\tau/\Delta t}}{\Delta t}$.

5. **Weiner-Kintchine theorem: the correlation function of a stationary ergodic signal is the fourier transform of power spectrum:** $\langle f(t)f(t+\tau) \rangle = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S(\omega) e^{-i\omega\tau}$ where the power spectrum $S(\omega) = \lim_{T \rightarrow \infty} T \sigma^2(\omega)$.

1.2.2 Theory of point processes

1. **Useful facts about general point process $N(t)$**

- (a) **The conditional intensity function** (hazard function) is given by:

$$\lambda(t|H_t) \equiv \lim_{\Delta t \rightarrow 0} \frac{P(N(t+\Delta t) - N(t) = 1 | H_t)}{\Delta t}$$

(i.e. the instantaneous probability of spiking at time t given history up to time t)

- (b) **The ISI or spike time density:**

$$f(t|H_t) \equiv \lim_{\Delta t \rightarrow 0} \frac{P(t_{N(t)+1} \in [t, t+\Delta t] | H_{t_{N(t)}})}{\Delta t}$$

(i.e. the instantaneous probability of spiking at time t for the first time since the last spike, given the history up till the last spike)

- (c) **Connection between CIF and ISI density:**

- i. $f(t|H_t) = \frac{1}{\Delta t} P(N(t+\Delta t) - N(t) = 1 | H_t) P(N(t) - N(t_{N(t)}) = 0) = \lambda(t|H_t) \exp\left\{-\int_{t_{N(t)}}^t \lambda(s|H_s) ds\right\}$.

(i.e. prob. of spiking at t and not spiking anytime in the interval $(t_{N(t)}, t)$).

$$\text{ii. } \lambda(t|H_t) = \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) = 1 | H_t, t_{N(t)+1} > t) = \frac{f(t|H_t)}{1 - \int_{t_{N(t)}}^t f(s|H_s) ds}$$

(d) **Joint dbn of an event sequence:**

$$P(0 \leq t_1 < \dots < t_n \leq T \cap N(T) = n) = \left(\prod_{j=1}^n \lambda(t_j | H_{t_j}) \right) e^{-\int_0^T \lambda(s|H_s) ds}$$

(Can derive this by taking n independent realizations of the ISI distribution).

(e) **Conditional prob. of n event times given number of events** can be obtained by dividing the above by

$$P(N(t) = n) = \frac{e^{-\Lambda} \Lambda^n}{n!} \text{ where } \Lambda = \int_0^T \lambda(s|H_s) ds. \text{ This gives}$$

$$P(0 \leq t_1 < \dots < t_n \leq T | N(t) = n) = n! \left[\prod_{j=1}^n \lambda(t_j | H_{t_j}) \right] \Lambda^{-n}.$$

2. Special properties of homog. Poisson PP (rate λ):

(a) **Density of the ISI is the Exponential(λ) density** $f(x) = \lambda e^{-\lambda x}$.

(b) **Distribution of an entire event sequence:**

$P(0 \leq t_1 < \dots < t_n \leq T \cap N(T) = n) = \lambda^n e^{-\lambda T}$ (independent of the event times).

(c) **Conditional prob. of n event times given number of events is the uniform density** $\frac{n!}{T^n}$ on $\{0 \leq t_1 < \dots < t_n \leq T\} \subseteq \mathbb{R}^n$.

(d) **The Fano factor for events is $F = 1$, Coeff. var. for ISI is $C_v = 1$, $Q_{\rho\rho}(\tau) = \lambda \delta_\tau$** (easy).

(e) **Limit of a Bernoulli process with parameter $\lambda \Delta t$:**

$$P(N(T) = j) = \binom{\lceil T/\Delta t \rceil}{j} (\lambda \Delta t)^j (1 - \lambda \Delta t)^{\lceil T/\Delta t \rceil - j}$$

One can show that:

$$\text{i. } \binom{\lceil T/\Delta t \rceil}{j} \rightarrow \frac{1}{j!} \left(\frac{T}{\Delta t} \right)^j$$

$$\text{ii. } (1 - \lambda \Delta t)^{\lceil T/\Delta t \rceil - j} \approx \left(1 + (-\lambda \Delta t) \right)^{\left(-\frac{1}{\lambda \Delta t} \right) (-\lambda T + j \lambda \Delta t)} \rightarrow e^{-\lambda T}$$

Plugging these limits in the mass function gives $P(N(T) = j) = \frac{1}{j!} \frac{T^j}{(\Delta t)^j} \lambda^j (\Delta t)^j e^{-\lambda T}$, which is the Poisson mass function

3. **Time-rescaling theorem:** Suppose $N(t)$ is the realization of a rate process with CIF $\lambda(t|H_t)$ on $t \in [0, T]$ with event times $0 \leq t_1 < \dots < t_{N(T)} \leq T$. Then if we make the change of variables $\Lambda(t) = \int_0^t \lambda(s|H_s)ds$, then $\{\Lambda(t_j)\}$ are event times of a homogeneous PP with rate 1.

Proof idea: Show that the transformed ISI's $\tau_k = \int_{t_k}^{t_{k+1}} \lambda(s|H_s)ds$ are i.i.d. $Exponential(1)$ r.v.'s by computing their joint density:

$$f(\tau_1, \dots, \tau_n \cap \tau_{n+1} > T) = f(\tau_1, \dots, \tau_n)P(\tau_{n+1} > T|\tau_1, \dots, \tau_n).$$

The second term is just $exp(-\int_{t_n}^T \lambda(s|H_s)ds)$ and the first term, when placed in an integral, can be rewritten by a change of variables as $|J|f(t_1, \dots, t_n \cap N(t_n) = n)$ where J is the jacobian of Λ . Since τ_k only depends on $\{t_j\}_{j=1}^k$, J is lower triangular and $|J| = \prod_{j=1}^n \lambda(t_k|H_{t_k})^{-1}$.

Thus the joint density becomes $[\prod_{j=1}^n e^{-\tau_k}]e^{-\int_{t_n}^T \lambda(s|H_s)ds}$.

1.2.3 Encoding/decoding

1. **Optimal linear Volterra-Weiner kernel** $D^*(\tau) = \arg \min_{D(\tau)} \|r(t) - \hat{r}(t)\|_2^2$

satisfies:

$$\forall \tau \int_0^T D^*(\tau') Q_{ss}(\tau - \tau') d\tau' = Q_{rs}(-\tau) \quad (1)$$

One can show this by writing the squared L_2 error in discrete time:

$$\|r(t) - \hat{r}(t)\|_2^2 = \sum_i \Delta t_i (r(t_i) - r_0 - \sum_j D(t_j) s(t_i - t_j) \Delta t_j)^2.$$

Setting $\nabla D(\tau_k) = 0$ and rearranging gives

$$\sum_i \Delta t_i \sum_j \Delta t_j D(t_j) s(t_i - t_j) s(t_i - t_k) = \sum_i \Delta t_i (r(t_i) - r_0) s(t_i - t_k).$$

Going back to continuous time and assuming the stimulus has mean 0, we can make a change of variables in the inner integral on the LHS to get the result.

In the white noise case, this is basically the STA since $Q_{ss}(\tau) = \sigma^2 \delta(\tau)$, we have: $D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma^2} = \frac{\tilde{r}^C(\tau)}{\sigma^2}$

2. **The optimal stimulus $s(t)$ with bounded energy ($\|s(t)\|_2^2 \leq E$) is proportional to $D^*(\tau)$:** One can show this by formulating the problem

with lagrange multipliers and applying the same time-discretization trick.

3. **Bussgang Theorem:** Consider the application of a static nonlinearity $F(\cdot)$ to the 1st-order estimate $\hat{r}(t) = \int_0^T D(\tau)s(t-\tau)d\tau$ and the case where $s(t)$ is white noise. We want to know whether the optimal linear kernel $D^*(\tau)$ (derived above) satisfies the 'consistency' condition:

$$D^*(\tau) = \frac{1}{T\sigma_s^2} \int_0^T \hat{r}(t)s(t-\tau)dt = \frac{1}{T\sigma_s^2} \int_0^T F(\hat{r}(t))s(t-\tau)dt \quad (2)$$

i.e. the relationship between kernel and stimulus/estimated response correlation does not change. It turns out that this holds if $s(t)$ is *Gaussian* white noise. This can be argued as follows:

- (a) For an r.v. X with dbn $\mathcal{N}(0, \sigma^2)$ and a constant $\alpha \in \mathbb{R}$, show that $\mathbb{E}[XF(\alpha X)] = \sigma^2\alpha\mathbb{E}[F'(\alpha X)]$ (this is just integration by parts for a Gaussian integral).
- (b) Extend to \mathbb{R}^n , i.e. if \vec{X} has dbn $\mathcal{N}(\vec{0}, \sigma^2 I)$, then for any $\vec{\alpha} \in \mathbb{R}^n$, show that $\mathbb{E}[\vec{X} * F(\vec{\alpha} * \vec{X})] = \sigma^2\vec{\alpha}\mathbb{E}[F'(\vec{\alpha} * \vec{X})]$ (where F is applied component-wise to a vector. (Just look at each component and integrate one of the n integrals by parts).
- (c) Extend to functionals, i.e. if $X(t)$ is a stationary gaussian signal and $D(\tau)$ is some kernel then:

$$\frac{1}{T} \int_0^T s(t-\tau)F((D*s)(t))dt = \frac{1}{T}\sigma_s^2 D(\tau) \int_0^T F'((D*s)(t))dt \quad (3)$$

(Express the integrals as sums, use previous step, and take $\Delta t \rightarrow 0$).

1.2.4 Entropy and mutual information

1. **Bounding the entropy of a spike times** We can easily calculate the entropy of a spike train under the following (very strong) assumptions:

- (a) Time is divided up into bins of size $\frac{T}{\Delta t}$ in which there is at most 1 spike
- (b) The probability $p = \tilde{r}\Delta t$ of a spike in any bin is the same and independent of all the other bins (they are like i.i.d Bernoulli r.v.'s)

In this case we know that:

$$H(D) = -\frac{T}{\log 2\Delta t} [[\tilde{r}\Delta t \log(\tilde{r}\Delta t) + (1 - \tilde{r}\Delta t) \log(1 - \tilde{r}\Delta t)] \quad (4)$$

Some important remarks:

- (a) Using an approximation of log for small Δt we can approximate the **entropy rate** as: $H(D)/T \approx \tilde{r} \log_2(\frac{e}{\tilde{r}\Delta t})$ bits/sec
- (b) $\Delta t \rightarrow 0$, the entropy goes to ∞ , which makes sense we can then encode infinitely any possibilities
- (c) Typically ISI's will be distributed over approximately $1/\tilde{r}$ seconds/spike, it follows that **each interval (each spike) carries approximately $\log_2(\frac{1}{\tilde{r}\Delta t})$ bits/spike of entropy.**
- (d) The assumptions only increase the entropy, since in reality spikes are correlated, thus reducing the joint entropy of spike times. Here we calculate entropy by simply counting the log of the number of possibilities given a resolution Δt .

2. Bounding the entropy of spike counts with entropy maximization

Which distribution $P(n)$ on spike counts yields the highest entropy

$$H(N) = -\sum_n P(n) \log_2 P(n) \text{ given the observed mean spike count}$$

$$\langle n \rangle = \sum_n nP(n). \text{ Formulating this as a constrained optimization problem (with 2 constraints - normalization and fixed mean) and setting the 'functional derivative' to 0, we get that the solution } P(n) = \frac{1}{Z} e^{-\alpha n} \text{ is the exponential distribution with rate } \alpha \text{ satisfying } e^\alpha = \frac{1+\langle n \rangle}{\langle n \rangle}.$$

The resulting maximal entropy is $\log_2(1 + \langle n \rangle) + \langle n \rangle \log_2(1 + 1/\langle n \rangle)$ bits. Important remarks:

- (a) As the time resolution decreases ($\delta t \uparrow$), the *entropy per spike* (divide entropy by $\langle n \rangle$) decreases since $\langle n \rangle$ becomes larger.
- (b) Even if our resolution is exactly the ISI (so that $\langle n \rangle = 1$), our entropy per spike is exactly 2 bits.
- (c) If we take the time window (T) very small, the spike count entropy is approximately the spike time entropy (this where a rate and timing code are essentially the same).

3. Bounding the mutual information between stimulus s and response y with entropy maximization

The mutual information is

the amount of variability in the response that comes from the variability in the stimulus. If we make the assumption that $y = gs + \eta$ where η is Gaussian noise with known variance, then $I(s; y) = H(y) - H(\eta)$, which is maximized (with a power constraint) when y has a Gaussian distribution iff s has a Gaussian distribution. Thus $I(s; y) \leq I(X; Y)$ where $X \sim \mathcal{N}(0, \sigma_s^2)$ and $Y = gX + \mathcal{N}(0, \sigma_\eta^2)$. This can be calculated to be:

$$I(X; Y) = \frac{1}{2} \log_2 \left[1 + \frac{\sigma_s^2}{\sigma_\eta^2 / g^2} \right] \quad (5)$$

Usually we define 'effective noise' as $\eta_{eff} = \eta/g$ and $\frac{\sigma_s^2}{\sigma_{\eta_{eff}}^2}$ is the **signal-to-noise ratio**.

2 Spiking models of neurons

2.1 Definitions

1. **Membrane potential** $V(t)$ = difference in voltage of cell (inside - outside). Resting potential is around $-60mV$. Can also think of voltage as 'amount of work' needed to transfer charge from one place to another.
2. **Depolarization/Hyperpolarization** - decrease/increase in $|V(t)|$ (usually an increase/decrease in $V(t)$), which could trigger/inhibit an action potential
3. **Membrane capacitance** $C_M = \frac{Q(t)}{V(t)}$ - measure of the amount of charge $Q(t)$ stored for a given electric potential $V(t)$. Given by $c_M A$ where c_M is dimensionless capacitance and A is surface area. Also relates to capacitative current (see below) via the equation $I_C(t) = C_M \frac{dV}{dt}$
4. **Membrane resistance** $R_M(V, t)$ - measure of how much the membrane blocks current. For the membrane it is given by $\frac{r_M}{A}$ where r_M is the dimensionless resistivity of the material and A is the surface area.
5. **Membrane conductance** $G_M(V, t) = \frac{1}{R_M(V, t)}$ - inverse of resistance. Given also by $A g_M$ where g_M is dimensionless conductivity and A is the surface area.
6. **Membrane current** $I_M(t)$ - rate of change of charge ($Q(t)$). This is usually modeled as the sum of the current through the various

ion channels using $V = I_M R_M$: $I_M(t) = \sum_j I_j(t)$ where $I_j(t) = \frac{V_j(t)}{R_j} = G_j(V, t)V_j(t)$ is the current passing through channel j with resistance $R_j(V, t)$ (or conductance $G_j(V, t)$), which could be voltage/time-dependent.

7. **Capacitive current** $I_C(t)$ - amount of current needed to change voltage at a certain rate $(\frac{dV}{dt})$ given fixed membrane capacitance C via the equation $Q(t) = C_M V(t)$.
8. **Nernst potential (Reversal potential)** of a particular ion - membrane potential at which there is no net flow of ions from one side to the other. Perturbation of the membrane potential on either side of the Nernst potential reverses the net direction of ion flow. This is a generalization to the equilibrium potential which says the flux of ions are all 0 (equivalent to Nernst in the single-ion case) The Nernst potential is derived by the following equation:

$$E = \frac{V_T}{z} \log\left(\frac{O}{I}\right) \quad (6)$$

where V_T is derived by some constants, z is the charge of the ion (charge = zq), and O and I are outside and inside ion concentrations, respectively.

9. **Time constant** $\tau_M = R_M C_M = r_M c_M$ sets the scale for changes in $V(t)$

2.2 Laws/Facts to know

1. **Ohm's Law:** Consider a cylindrical slice of cable of length Δx with cross-sectional area A . Suppose there is axial current I_L passing through the cable and that voltage $V(x)$ is independent of the position in each cross section. Then $V(x) - V(x + \Delta x) = I_L r_{slice} = I_L \rho \Delta x / A$ where ρ is the resistivity of the material. Taking $\Delta x \rightarrow 0$, we have:

$$\frac{dV}{dx} = -I_L \rho / A = -I_L R_L \quad (7)$$

2. **Conservation of current:**

- (a) Cable version: Consider a semi-infinite cable starting at position x_0 with axial current I_L . At any position $x \geq x_0$, the amount of current going in must be equal to the amount of current at x

+ the amount that was lost through the membrane (membrane current). That is, : $I_L(x_0) = \int_{x_0}^x I_M(s)ds + I_L(x)$. Taking $\frac{d}{dx}$ of both sides gives:

$$\frac{dI_L}{dx} = -I_M(x) \quad (8)$$

- (b) Space-clamped version (for a fixed x): The total membrane current $I_M(t)$ can be split up into the capacitative and ionic channel components, giving:

$$I_M(t) = I_C(t) + \sum_j G_j(t)(V_j(t) - E_j) = C_M \frac{dV}{dt} + \sum_j G_j(t)V_j(t) \quad (9)$$

if we model the ionic channel current as a sum of channel currents.

3. **Cable equation:** Combining Ohm's law and conservation of current (in a cable) gives:

$$\frac{d^2V}{dx^2}(x, t) = R_L I_M(x, t) \quad (10)$$

Typically $I_M(x, t)$, for given x , is expanded as the sum of time and voltage-dependent channel currents.

2.3 Hodgkin-Huxley

1. **Assumptions:**

- (a) Voltage is 'space clamped' i.e. it is independent of position x and is solely a function of time.
- (b) 3 channels (currents): Sodium (Na), potassium (K), leakage (L) with **constant** reversal potentials $E_K < E_L < 0 < E_{Na}$.
- (c) Conductance of channel j , $G_j(V, t)$ is given by $\bar{g}_j P_j$ where \bar{g}_j is a constant representing the maximum channel capacity, and P_j is the probability of the channel being open (the idea is that there are several of these channels so P_j is the proportion of channels open).
- (d) The potassium (K) channel has 4 independent subgates. Each subgate has a closed state (S) and open state (S^*), and the fraction of subgates of this type in state S^* is given by $n(V, t)$. Thus $P_K = n^4$

- (e) The sodium (Na) channel has 3 independent subgates of which the fraction in the open state is given by $m(V, t)$. There is also an independent 'blocker' type of gate of which the fraction in the open state is $h(V, t)$. So $P_{Na} = m^3 h$
- (f) The gating probabilities $x \in \{n, m, h\}$ are time and voltage dependent. A subunit in a channel of type x transitions from $S \rightarrow S^*$ at rate $(\alpha_x(V))$ and from $S^* \rightarrow S$ at rate $(\beta_x(V))$. The dynamics resulting from these ODE's (see latter 3 equations below) can be characterized by a decay to a (voltage-dependent) 'equilibrium probability' with a voltage-dependent 'time constant':

$$x(V, t) \rightarrow x_\infty(V) = \frac{\alpha_x(V)}{\alpha_x(V) + \beta_x(V)} \quad (11)$$

with time constant $\tau_x(V) = (\alpha_x(V) + \beta_x(V))^{-1}$. Typically, $n_\infty(V)$ and $m_\infty(V)$ increase with V and are sigmoidal in shape, while $h_\infty(V)$ decreases with V and is like the reverse sigmoid. The variables V and m are *fast*, while h and n are *slow*.

- (g) The leakage (L) channel has a fixed conductance $\overline{g_L}$.

2. Resulting ODE:

$$C_M \dot{V} = \overline{g_{Na}} m^3 h (V - E_{Na}) + \overline{g_K} n^4 (V - E_K) + \overline{g_L} (V - E_L) + I(t) \quad (12)$$

$$\forall x \in \{n, m, h\}$$

$$\dot{x} = \alpha_x(V)(1 - x) + \beta_x(V)x \quad (13)$$

3. **Fast time-scale analysis:** On a fast time-scale, h and n are assumed effectively constant and we analyze in the (V, m) plane. For $I_M \equiv 0$, the nullcline $\frac{dV}{dt} = 0$ is given by points (V, m) satisfying:

$$v = \frac{g_{Na} m^3 h E_{Na} + g_K n^4 E_K + g_L E_L}{g_{Na} m^3 h + g_K n^4 + g_L} \quad (14)$$

i.e. a weighted combination of the equilibrium potentials. The (effectively constant) values of n and h determine the number of points at which the nullclines $\frac{dm}{dt} = 0$ and $\frac{dV}{dt}$ intersect (1,2-borderline, or 3 points). In the 3-point case, we have 2 stable nodes R (rest), and E (excitatory), and one saddle node T (transitional) (to show this rigorously, linearize and look at eigenvalues). There is a separatrix dividing

the R-basin of attraction from the E-basin of attraction (this is the stable manifold of T). Jumps in voltage (corresponding to lateral shifts rightward in the (V, m) plane) that cross the separatrix lead to the excitatory state. Subsequently, the the slow variables n, h evolve so that the the points T and E collide, leaving R as the only stable node (saddle-node bifurcation).

4. **Slow time-scale analysis:** If we make 2 assumptions-

- (a) $m(t)$ is relatively instantaneous, i.e. $m(t) \equiv m_\infty(V)$
- (b) $\tau_n(t) \approx \tau_h(t)$, so that we can make the approximation $(b-h) \approx an$ for some constants a, b . More formally, we project the state pair $n(V, t), h(V, t)$ to the line $b-h=an$. Typically, (a, b) are chosen s.t. this line is tangent to the curve traced by $\{(n_\infty(V), h_\infty(V) : V \in \mathbb{R})\}$ at the point $(n_\infty(V_{rest}), h_\infty(V_{rest}))$.
- (c) $V(t)$ is also relatively instaneous, i.e. $C_M \frac{dV}{dt} = g_{Na}m_\infty^3(b-an)(V - E_{Na}) + g_K n^4(V - E_K) + g_L(V - E_L) = 0$. This defines the **slow manifold** in the (V, n) plane. Deviations (necessarily lateral since n is slow) from this manifold occur only on the fast time-scale.

We can analyze the (V, n) phase-plane. There is range of values of n (and corresponding h -values) for which there are 3 points on the slow manifold (this corresponds to those values for (n, h) for which there were 3 equilibria in the fast time scale). Thus, the slow manifold consists of 3 segments corresponding to the R, T, and E fast time-scale fixed points. There is a region in the plane where trajectories will always jump laterally to the R segment, and another region where they jump laterally to the E segment. The T segment is unstable. Evolution *on* the slow manifold is characterized by motion toward the nullcline $\frac{dn}{dt} = 0$. One can see then that if the nullcline $\frac{dn}{dt} = 0$ intersects the slow manifold only in the R branch, then the E branch can only be temporary (it cannot stay in the excited state indefinitely).

5. 2D models are classified into 2 types based on what type of bifurcation occurs as we vary the applied current I :

- (a) (Type I) When there is a saddle-node bifurcation on a limit cycle, i.e. a stable node and saddle point collide to form a stable limit cycle whose frequency starts at 0. (this is not in any of the examples here)

- (b) (Type II) When there is a subcritical Hopf bifurcation, i.e. within a bounded surface, a single stable fixed point loses stability, thus admitting a stable limit cycle inside the surface (by Poincare-Bendixson) whose frequency starts at some nonzero finite value.
- Simple example of Type II: Fitz-Nagumo**

$$\dot{V} = V - V^3/3 - w + I \quad (15)$$

$$\dot{w} = \epsilon(b_0 + b_1V - w) \quad (16)$$

6. Accounting for biological phenomena

- (a) **Spike (action potential) generation and negative overshoot:** Applied current \rightarrow increases $V(t) \rightarrow$ increases m, n and decreases h (but n, h are much slower than $m \rightarrow$ increases g_{Na} opening Na channels \rightarrow increases $V(t)$ (positive feedback). When $V(t)$ is very high h keeps on decreasing and n keeps on increasing \rightarrow decreases g_{Na} and g_K (potassium ions flow out) \rightarrow decreases $V(t)$.
- (b) **Periodic spiking:** by applying steady current we can translate the slow manifold upward so that the nullclines intersect on the T branch of the slow manifold. This is an example of a **subcritical Hopf bifurcation** in which the stable node loses stability (the bifurcation parameter being varied is the applied current). Poincare-Bendixson ensures the existence of a stable limit cycle (which indeed is the cycle corresponding to periodic spiking).
- (c) **Threshold effect:** the separatrix in the (m, V) (fast) phase plane shows that there is a minimum amount of jump in voltage needed in order to make a neuron spike from rest.
- (d) **Refractoriness and negative overshoot:** Just after an action potential, there is a return to the R branch of the slow manifold with volage slightly lower than the rest potential, where a very large voltage jump is needed in order to go the E branch again, since you have to cross the 'T' segment of the manifold in order to get back to the spiking state.
- (e) **Anode-break excitation:** if we clamp the voltage to some value for which n is below the fixed point and let the system evolve for a long time, then unclamp, we get an immediate spike. If we were to alter the voltage immediate from rest state, a spike would not occur - we need to let the system evolve beforehand.

7. **Connecting with cable equation:** Plugging the equation for I_M into the cable equation, we can replace the first equation above with:

$$\frac{a}{2\rho} \frac{d^2V}{dx^2} = C_M \frac{dV}{dt} + \overline{g_{Na}} m^3 h (V - E_{Na}) + \overline{g_K} n^4 (V - E_K) + \overline{g_L} (V - E_L) \quad (17)$$

It turns out that this partial-differential equation admits a 'traveling wave' solution $V(x, t) = V(t - x/\theta)$ (think of a 'blip' in voltage traveling along the cable). This is essential in the theory of how information is transmitted in the brain. The basic steps for showing this are outlined as follows:

- (a) Introduce a parameter ϵ into the first and second equations wherever we consider partial derivatives of the fast variables V, m . In particular, multiply the $\frac{d^2V}{dx^2}$ term by ϵ^2 , and the $\frac{dV}{dt}$ and $\frac{dm}{dt}$ terms by ϵ . We want the behavior of the system as $\epsilon \rightarrow 0$ to tell us something about when $\epsilon = 1$.
- (b) Make some simplifications:
 - i. Assume $m_\infty(V) = n_\infty(V) = 1 - h_\infty(V)$ is just a step fn at some $V^* \in (E_L, 0)$.
 - ii. τ_m, τ_h, τ_n are all indep. of V , $\tau_m = 0$ (so $m \equiv m_\infty$).
- (c) Guess a solution of the form where V, m, h, n are functions of one variable $T = t - \frac{x}{\theta}$ instead of (x, t) . We guess explicit solutions for m, n, h as follows:
 - i. $m(T)$ is 0 except for $T \in (0, T_1)$ for some T_1
 - ii. $h(T)$ is 1 on $(-\infty, 0)$, exponentially decays on $(0, T_1)$, and then sigmoidally recovers back to 1 on (T_1, ∞) .
 - iii. $n(T)$ is 0 on $(-\infty, 0)$, increases on $(0, T_1)$, and decays exponentially back to 0 on (T_1, ∞) .
- (d) Solve an 'outer solution' for $V(T)$ valid away from $T = 0$ or 1 by letting $\epsilon \rightarrow 0$, and plugging in our fixed forms for $m(T), h(T), n(T)$.
- (e) For T near 0, make a change of variables $T = \epsilon S$ and let $U(S) = V(\epsilon S) = V(T)$ and write the system in terms of U , then let $\epsilon \rightarrow 0$ - we get a 2nd order differential equation in U which can be solved piecewise for $S < 0$ and $S > 0$.
- (f) Solve for θ (very complicated - see Peskin's notes). Do the same for a solution for T near 1 with change of variables $T = T_1 + \epsilon S$.

2.4 LIF

1. Assumptions:

- (a) There is only one channel (the leak channel) with fixed conductance \bar{g}_L so that the ionic current is just $-(V - E_L)$ (again constant leak reversal potential is assumed).
- (b) Action potentials are stereotyped in the sense that whenever $V(t) = V_{threshold}$, V is reset to V_{reset} and subthreshold dynamics continue as usual.

2. Resulting ODE:

$$\dot{V} = -\frac{(V - E_L)}{\tau_M} + \frac{1}{C_M}I(t) \quad (18)$$

- ### 3. Solutions for the membrane potential:
- The ODE above can be solved explicitly using standard nonhomogeneous linear ODE techniques: The homogeneous solution (ignoring $R_M I(t)$) for $\tilde{V}(t) = V(t) - E_L$ is:

$$\tilde{V}(t) = \tilde{V}_0 e^{-\frac{(t-t_0)}{\tau_M}} \quad (19)$$

(where $\tilde{V}(t_0) = \tilde{V}_0$). The general solution for $\tilde{V}(t)$ is:

$$\tilde{V}(t) = \tilde{V}_0 \exp\left(-\frac{(t-t_0)}{\tau_M}\right) + \frac{1}{C_M} \int_{t_0}^t e^{\frac{s-t}{\tau_M}} I(s) ds \quad (20)$$

For a constant applied current $I(t) \equiv I_0$ with initial condition $V_0 = 0$ this reduces to:

$$\tilde{V}(t) = R_M I_0 (1 - e^{-\frac{(t-t_0)}{\tau_M}}) \quad (21)$$

Notice that $V(t) \rightarrow R_M I_0 + E_L$ as $t \rightarrow \infty$, so if $R_M k + E_L < V_{threshold}$, no spikes will occur.

4. ISI time for and firing rate for constant stimulus:

For a constant stimulus $I(t) = k$, we can set the solution $V(t) = V_{th}$ and solve for t :

$$t_{ISI} = t_{fire} - t_0 = \tau_M \log\left(\frac{R_M k + E_L - V_{reset}}{R_M k + E_L - V_{th}}\right) \quad (22)$$

The inverse $r_{ISI} = t_{ISI}^{-1}$ represents the firing rate, which for sufficiently large k , can be approximated (using $\log(1+z) \approx z$ for small z) as:

$$r_{ISI} \approx \left[\frac{E_L - V_{threshold} + R_M k}{\tau_M (V_{threshold} - V_{reset})} \right]_+ \quad (23)$$

which shows that the firing rate is asymptotically linear in the driving current I_0 .

5. **Refractory periods and spike rate adaptation:** The following techniques can be employed to account for refractory/adaptation effects:

- (a) Make the spiking mechanism more complicated - add an explicit refractory period Δ_{abs} . In the constant current case this just increases t_{ISI} by Δ_{abs} , and thus makes the firing rate r_{ISI} saturate for large k . Alternatively we can increase the threshold itself just after a spike.
- (b) Make the voltage dynamics more complicated - add another current term with conductance $g_{sra}(t)$ to the model. The subthreshold dynamics of $g_{sra}(t)$ is simple exponential decay with time constant τ_{sra} . However, whenever there is a spike we increment this conductance by Δg_{sra} . The result will be that repeated spiking will increase g_{sra} and thus inhibit $V(t)$ from constantly crossing the threshold ('adaptation'). If Δg_{sra} is extremely large and has a rapid recovery, this results in a kind of 'refractory' effect.

6. **Extension to nonlinear models:** IF models can be extended to the NLIF models where the time constant and membrane resistance are themselves voltage-dependent. The resulting ODE is:

$$\tau_M \dot{V} = F(V) + G(V)I(t) \quad (24)$$

which can always be reduced to:

$$\tau_M \dot{\tilde{V}} = \gamma(\tilde{V}) + I(t) \quad (25)$$

via the change of variables $\tilde{V}(t) = \tau_M \int_0^{V(t)} G(x)^{-1} dx$

A simple example of this is the *quadratic IF model*:

$$\tau_M \dot{V} = a_0(V - E_L)(V - V_c) + R_M I(t) \quad (26)$$

where $a_0 > 0, V_c > V_{rest}$ are parameters. Notice that by observing the phase (V, \dot{V}) plane, one can see the relation to the type I models described earlier.

2.5 Stochastic LIF

2.5.1 Langevin's equation (diffusive noise) \rightarrow Fokker-Planck equation

Consider the SDE:

$$dV_t = a(V, t)dt + \sigma(V, t)dB_t \quad (27)$$

The density $\rho(v, t)$ of $V(t)$ satisfies the Fokker-Planck equation:

$$\frac{d}{dt}\rho(v, t) = -\frac{d}{dv}[a(v, t)\rho(v, t)] + \frac{1}{2}\frac{d^2}{dv^2}[\sigma^2(v, t)\rho(v, t)] \quad (28)$$

The idea of the proof is based on 4 ingredients involving the transition density $\rho(v, t|V(s) = x)$:

1. It is apparent from the SDE that $\mathbb{E}[V(t + \Delta t) - V(t)|V(t) = v] = a(v, t)\Delta t + o(\Delta t)$ and $\mathbb{E}[(V(t + \Delta t) - V(t))^2|V(t) = v] = \sigma^2(v, t)\Delta t + o(\Delta t)$
2. Computing $\int R(x)\rho(V(t + \Delta t) = x|V(t) = v)dx$ by taylor-expanding R about v to order $o(\Delta t)$
3. Computing $\int R(x)\rho(V(t + \Delta t|V(s) = v)dx$ by taylor-expanding rho as a function of t about t to order $o(\Delta t)$.
4. The Kolmogorov-Chapman equation $\rho(V(t) = x|V(s) = v) = \int \rho(V(t) = x|V(s') = v')\rho(V_s' = v'|V(s) = v)dv'$ for $s < s' < t$.

2.5.2 Stein model \rightarrow diffusive noise model

Consider the following model (Stein model) for the voltage evolution:

$$\tau_M \dot{V} = -(V - E_L) + R_M I(t) + \tau_M \tau_M \sum_k \sum_{t_r^k > t_0} w_k \delta(t - t_r^k) \quad (29)$$

where the last term describes the stochastic effects of presynaptic/background neurons (k). The assumptions are:

1. $\forall k$, the k^{th} input is Poisson with rate $\nu_k(t)$.
2. a spike from input k induces a jump in the voltage by w_k followed by an exponential decay at rate τ_M . The w_k are small.

It turns out that in the 'diffusion' limit $w_k \rightarrow 0$, this model reduces to a Fokker-Planck equation:

$$\tau_M \frac{d\rho}{dt}(V', t) \approx -\frac{d}{dV}[-V' + \tau_M \sum_k \nu_k(t)w_k]\rho(V', t) + \frac{1}{2}[\tau_M \sum_k \nu_k(t)w_k^2] \frac{d^2\rho}{dV^2}(V, t) \quad (30)$$

This equation corresponds to the Langevin equation:

$$\tau_M dV = -(V - E_L)dt + R_M I(t)dt + \tau_M \sigma W_t \quad (31)$$

with $R_M I(t) = \tau_M \sum_k \nu_k(t)w_k$ and $\sigma^2(t) = \tau_M \sum_k \nu_k(t)w_k^2$

The result can be shown as follows:

1. Derive the 'transition' equation by conditioning on whether or not there was a spike $\in [t, t + \Delta t]$. Use $e^{-x} \approx 1 - x$:

$$\rho(V', t + \Delta t | V(t) = v) = (1 - \Delta t \sum_k \nu_k(t)) \delta(V' - v e^{-\Delta t / \tau_M}) + \Delta t \sum_k (\nu_k(t) \delta(V' - w_k - v e^{-\Delta t / \tau_M})) \quad (32)$$

2. Use the transition density in with the Chapman-Kolmogorov equation:

$$\rho(V', t + \Delta t) = \int \rho(V', t + \Delta t | V(t) = v) \rho(v, t) dv \quad (33)$$

and integrate to get:

$$\rho(V', t + \Delta t) = [1 - \delta t \sum_k \nu_k(t)] e^{\delta t / \tau_M} \rho(V' e^{\Delta t / \tau_M}, t) + [\Delta t \sum_k \nu_k(t)] e^{\Delta t / \tau_M} \rho(e^{\Delta t / \tau_M} (V' - w_k), t) \quad (34)$$

3. Expand to first-order Δt , then manipulate to get $\tau_M \frac{\rho(V', t + \delta t) - \rho(V', t)}{\Delta t}$ on the LHS, take $\Delta t \rightarrow 0$, and then expand the RHS about $V' = V$ up to second order w_k^2 (implicitly assumes that w_k are small) to get the Fokker-Planck equation:

$$\tau_M \frac{d\rho}{dt}(V', t) \approx -\frac{d}{dV}[-V' + \tau_M \sum_k \nu_k(t)w_k]\rho(V', t) + \frac{1}{2}[\tau_M \sum_k \nu_k(t)w_k^2] \frac{d^2\rho}{dV^2}(V, t) \quad (35)$$

2.5.3 Solutions for $V(t)$

To solve for $\rho(v, t)$ in light of the threshold condition we impose boundary condition $\rho(v_{th}, t) = 0 \forall t$, and also an initial condition $\rho(v, 0) = \delta(V_0)$. **Typically, we cannot analytically solve the FP with these boundary conditions. However, without a boundary condition the solution is:**

$$V(t) = \mathcal{N}(\hat{V}(t), \frac{\sigma^2}{2}(1 - e^{-2(t-t_0)/\tau_M})) \quad (36)$$

i.e. an expanding Gaussian 'cloud' around the noise-free trajectory $\hat{V}(t)$. We can show this by looking at the Langevin form and just treating the white noise term as a nonhomogeneous linear term to get the general solution:

$$V(t) = \frac{R}{\tau_M} \int_{t_0}^t e^{-\frac{(s-t)}{\tau_M}} I(s) ds + \int_{t_0}^t e^{-(s-t_0)/\tau_M} dB_s \quad (37)$$

We can compute the mean and covariance to be the above values by using the properties of Gaussian white noise. **The ISI's for stochastic LIF models cannot be computed analytically generally.**

2.5.4 Connection of LIF with stochastic point processes

1. **Escape rates:** Another way to add a stochastic aspect to the deterministic LIF model is to randomize the threshold. One formulation is to define an 'escape probability' $\rho(t|t_0) = f(V(t|t_0) - \mathcal{V})$ for a 'soft threshold' \mathcal{V} (and t_0 is time of last spike), which is the instantaneous probability that the voltage (with noise) has crossed \mathcal{V} given that the deterministic voltage is V and there was a spike at time t_0 . Thus $\rho(t|t_0)$ converts the voltage into a conditional intensity function for a point process (for example we can choose the escape rate to be step (this would give the usual threshold rule), exponential, piecewise linear, sigmoidal in its argument $V(t) - \mathcal{V}$).
2. **Approximating diffusive noise model:** From above we see that the 'free' solution to the FP equation is an expanding cloud around the noise-free trajectory with time-dependent variance $\sigma^2(t)$ that starts at 0 and rapidly approaches the limit $\sigma^2/2$). **The key is that we can approx. the probability of firing in $[t, t + \Delta t]$ by $\Delta t \exp\{-\frac{(\tilde{V}(t) - \mathcal{V})^2}{2\sigma^2(t)}\}$. where $\tilde{V}(t)$ is the noise-free trajectory. This motivates an escape model with with escape rate:**

$$\rho(t|t_0) = \frac{C_1}{\tau} \exp\left\{-\frac{(V_0(t) - \mathcal{V})^2}{\sigma^2}\right\} \quad (38)$$

(note we have assumed $\sigma^2(t)$ is effectively the asymptotic value. One problem with this is that it does not account for the fact that a small jump in voltage at time t_0 can create an 'atom' of probability of firing at time t_0 since the Gaussian distribution drifts across the threshold \underline{V} . To correct for this we include dependence of $\rho(t|t_0)$ on the (left-hand) *derivative* of $V(t)$ as well:

$$\rho(t|t_0) = \left[\frac{C_1}{\tau} + \frac{C_2}{\sigma} \dot{V}_{0+} \right] \exp\left\{ -\frac{(V_0(t) - \mathcal{V})^2}{\sigma^2} \right\} \quad (39)$$

2.6 Linear-nonlinear-Poisson (LNP)

1. Assumptions:

- (a) Spikes are generated according to a nonhomogeneous poisson process depending entirely on the stimulus and model paramters. This automatically implies the rate code and independent spike and independent neuron assumption. Also, it follows that spike train statistics are the same as those derived above for PP's (Fano factor, Coeff of variation, ISI, etc.).
- (b) The firing rate depends *instantaneously* on the projection of the stimulus onto 1 or more axes corresponding to 'receptive fields', followed by a static nonlinearity. The receptive field is the analog of the first-order kernel of a Wiener-Volterra series expansion of the response rate. However, while the Wiener-Volterra approach typically applies simple polynomial-like nonlinearities to high-dimensional input to estimate the firing rate, the LNP approach reduces the dimension to 1 and applies arbitrarily complex point nonlinearities.

- 2. **Resulting equation:** The spike count $N(t)$ is given by a Poisson process with rate:

$$\lambda(t) = R(\langle k_1, s(t) \rangle, \dots, \langle k_d, s(t) \rangle) \quad (40)$$

where each $k_j \in \mathbb{R}^n$ and $R : \mathbb{R}^d \rightarrow \mathbb{R}$ is some (usually nonlinear) function.

2.7 Generalized linear models

1. Assumptions:

- (a) Spikes are generated according to point process in which the rate (conditional intensity function) depends on linear projections of the stimulus, spike history, and spikes from neighboring neurons. Note that this allows for the possibility of temporal coding, and dependent spike/neuronal covariance structure (in particular, the output is NOT necessarily Poisson).
- (b) The same cascade approach is applied as in the LNP- the linear combination of the 3 covariates is passed through a point nonlinearity to get the instantaneous rate. The nonlinearity is in fact the link function when this model is interpreted as a GLM where the dependent variable $Y_t = \Delta N_t$, whose mean is given by the CIF $\lambda(t|H_t)$.

2. Resulting equation:

$$\lambda(t|H_t) = R(b + \sum_{j=1}^d \langle k_j, s(t) \rangle + \sum_k \sum_{t_j^k \leq t} h_k(t - t_j^k)) \quad (41)$$

where k indexes neighboring neurons (including itself) and t_j^k indexes the j 'th spike of neuron k .

Remark: We can achieve the same model by taking the limit of a (conditional) Bernoulli process with probability of success at time t being $\lambda(t|H_t)(\Delta t)$ and link function $\log(x/(1-x))$ via the approximation $\log(\frac{\lambda(t)\Delta t}{1-\lambda(t)\Delta t}) \approx \log(\lambda(t)\Delta t)$ for small Δt

Remark: This model is related to the 'escape-rate' approximation to the stochastic LIF model. In the stochastic IF solution, if we assume that the variance of the free distribution around the noise-free trajectory is constantly $\sigma/2$, we can use the Gaussian 'cloud' interpretation to define an 'escape rate' i.e. the instantaneous probability of firing (crossing the threshold) when $V(t)$ is at a certain value. **Thus the firing rate of the GLM is like the state-variable of the stochastic LIF with an escape rate.**

3 Learning model parameters

3.0.1 Estimators for the linear filter k

Consider static stimuli $s \in \mathbb{R}^n$ with prior distribution $P(s)$. Let $X \in \mathbb{R}_{m \times n}$ data matrix where each row $(x^{(j)})^T$ is one stimulus and let $r \in \mathbb{R}^n$ be the

vector of responses (say, number of spikes). Consider the following estimators:

1. Linear regression (LR):

- (a) Motivating idea: try to approximate the response (minimizing mean squared error) by projecting stimulus onto a vector v .
- (b) Definition: $k_{LR} = \arg \min_{v \in \mathbb{R}^n} \|Xv - r\|_2^2 = (X^T X)^{-1} X^T r$ (this is the optimal Volterra-Weiner 1st-order kernel)

2. Spike-triggered average (STA):

- (a) Motivating idea: average the stimulus weighting them by the response they induced
- (b) Definition: $k_{STA} = \frac{1}{n} X^T r$

3. Fisher linear discriminant (FLD):

- (a) Motivating idea: take the inner product that maximizes the difference between the average spike-triggered stimulus and average nonspike-triggered stimulus, divided by their variance

- (b) Definition: $k_{FLD} = \arg \max_{v \in \mathbb{R}^n} J(v) = \frac{[\langle v, \hat{x}_{spike} \rangle - \langle v, \hat{x}_{nospike} \rangle]^2}{\hat{p}_{spike} \langle v, \hat{\Sigma}_{spike} v \rangle + \hat{p}_{nospike} \langle v, \hat{\Sigma}_{nospike} v \rangle} =$
 $\arg \max_{v \in \mathbb{R}^n} \frac{\langle k, S_B k \rangle}{\langle k, S_W k \rangle} = S_W^{-1} [\hat{x}_{spike} - \hat{x}_{nospike}]$ where $\{\hat{x}_*, \hat{\Sigma}_*, \hat{p}_*\}$ are the empirically estimated spike/non-spike {avg,cov,prob} respectively.

S_B is the 'between-class' covariance $(\hat{x}_{spike} - \hat{x}_{nospike})(\hat{x}_{spike} - \hat{x}_{nospike})^T$
 S_W is the 'intra-class' covariance $S_W = \hat{p}_{spike} \hat{\Sigma}_{spike} + \hat{p}_{nospike} \hat{\Sigma}_{nospike}$

4. Maximum likelihood estimate (MLE):

- (a) Motivating idea: take the filter v which maximizes the probability of the data given $k = v$
- (b) Definition: $k_{MLE} = \arg \max_{v \in \mathbb{R}^n} \sum_{j=1}^n \log P(r_j | x^{(j)}) = \sum_{j=1}^n \log R(\langle v, x^{(j)} \rangle)$

3.0.2 Useful facts

1. **If the mean of X and r are 0 ($X^T e = \langle r, e \rangle = 0$), then $k_{LR} \propto k_{FLD} \propto (X^T X)^{-1} k_{STA}$.**

The second follows directly from definitions. The first follows because $X^T r \propto (\hat{x}_{spike} - \hat{x}_{nospike})$ and $X^T X \propto S_W + aS_B$ for some scalar a , we have that $k_{LS} \propto k_{FLD}$.

2. **If $P(s)$ is elliptically symmetric ($P(s) = f(\|As\|_2) \forall s \in \mathbb{R}^n$ for some nonsingular symmetric $A_{n \times n}$), then $k_{LR}, A^2 k_{STA}, k_{FLD}$ are all unbiased and consistent (expected values are proportional to the true k) estimators.** If there are multiple filters, then these estimators all lie in the span of these filters.

From the 0 mean condition, we know that all 3 estimators are proportional, so it suffices to show for k_{STA} . Consistency follows from unbiasedness and SLLN. Unbiasedness follows from a symmetry argument: split \mathbb{R}^n into two spaces via reflection about the $(A^{-1}k)$ -axis $s \mapsto s^*$ s.t.:

- (a) $\|s\|_2 = \|s^*\|_2$
- (b) $\langle A^{-1}k, s \rangle = \langle A^{-1}k, s^* \rangle$
- (c) $s + s^* \propto A^{-1}k$

$$\begin{aligned} \text{Therefore: } \mathbb{E}[k_{STA}] &= \mathbb{E}\left[\frac{Q_{rs}}{\tilde{r}}\right] = \int_{\mathbb{R}^n} s \frac{TR(\langle k, s \rangle)}{\tilde{r}} f(\|As\|_2) ds = \\ A^{-1} \int_{\mathbb{R}^n} s \frac{TR(\langle A^{-1}k, s \rangle)}{\tilde{r}} f(\|s\|_2) |A| ds &\propto A^{-2}k \text{ by symmetry.} \end{aligned}$$

The idea is that A 'rotates out' the correlations in $P(s)$, and so one can estimate A^{-2} by $X^T X$. The same symmetry argument holds in the multiple-filter case - just reflect wrt the plane spanned by the filters.

Remark: For example, any gaussian distribution is elliptically symmetric. However, if the nonlinearity is such that it does not induce a difference in mean responses of spike-triggered and non-spike-triggered ensembles, then the proportionality constant is 0, so these estimators are useless. Thus this 'asymmetry' condition on the nonlinearity is very important in practice.

3. **If $P(s)$ is elliptically symmetric, and the nonlinearity $R(\cdot)$ is convex/log-concave, then $\mathbb{E}[k_{MLE}] \propto k$.**

The proof is very similar to that for the k_{STA} . The convexity of R

ensures that there is a unique maximum (asymptotically) of the log-likelihood function:

$$L(v) \propto \sum_{j:r_j=1} \log R(\langle v, x^{(j)} \rangle) + \sum_{j:r_j=0} \log(1 - R(\langle v, x^{(j)} \rangle))$$

We know that the ML estimator asymptotically maximizes $\mathbb{E}_k[L(v)]$. Writing out this integral again gives us a symmetrical integral which is proportional to k . So the unique maximizer, again by convexity, must lie on the axis spanned by k . As before, the proportionality constant could be 0 if the nonlinearity does not satisfy the asymmetry condition.

Remark: The k_{MLE} is like a 'nonlinearly-weighted' version of $A^2 k_{STA}$: Setting $\nabla_v L(v) = 0$ gives:

$$\sum_{j=1}^n x^{(j)} \frac{R'}{R}(\langle k_{MLE}, x^{(j)} \rangle) r_j \propto \int xp(x) R'(\langle k_{MLE}, x \rangle) dx$$

The RHS converges to something proportional to $A^2 k_{MLE}$ by the same symmetry argument, and the LHS is reweighted type of STA with weights $\frac{R'}{R}(\langle k_{MLE}, \cdot \rangle)$.

3.0.3 Comparison of estimators

Est.	Bias/consistency	Variance	Tractability
LR	P(s) symm.	?	EAsy
STA	P(s) symm.	?	Easy
FLD	P(s) symm	?	Easy
MLE	Unbiased (asymptotically) if P(s) symm and R log-concave	Optimal if R is the correct nonlinearity	Tractable for R log-concave

3.0.4 The STC estimator

1. Definition: $\hat{C} = eig(\frac{1}{N-1} \sum_{n=1}^N [s(t_n) - \hat{A}][s(t_n) - \hat{A}]^T)$, where the operator *eig* extracts eigenvectors corresponding to 'outlying' eigenvalues. Note that the matrix inside *eig*, called the STC matrix, is an estimate of the 2nd-order term in the Wiener-Volterra expansion. If the stimuli have spherical covariance, then \hat{C} represents *the change in variance between the spike-triggered and general ensemble*.
2. **If the prior distribution on the stimuli is Gaussian, then \hat{C} is a consistent estimator for $\{k_1, \dots, k_d\}$.**

Proof: WLOG assume $P(s) = N(0, I_{d \times d})$. The key is that the conditional distribution (and therefore the conditional variance) in the orthogonal complement of $K = \text{span}\{k_j\}$ does not depend on the response:

$$\forall s \in \mathbb{R}^n, s = s_1 + s_2 \text{ where } s_1 \in K, s_2 \in K^\perp \Rightarrow P(s_2|n, s_1) = \frac{P(s_1, s_2)P(n|s_1)}{P(s_1, n)} = \frac{P(s_1)P(s_2)P(n|s_1)}{P(s_1, n)} = P(s_2) \text{ where } P(s_1, s_2) = P(s_1)P(s_2)$$

follows from the the multivariate Gaussian (with identity covariance) assumption on $P(s)$. It follows that axes along which the variance of the spike-triggered ensemble is very different from 1 represent directions in K . Thus we look at the eigenvectors of the covariance matrix corresponding to eigenvalues significantly different from 1.

Remark: It can also be shown that if $P(s)$ is not symmetric Gaussian, then $\exists f$ s.t. the STC estimator is biased. Also, if the nonlinearity does not induce a change in the variance of the spike-triggered ensemble versus the general ensemble, the STC will simply yield 0 and is useless.

4 Decoding/Discrimination

Decoding task: reconstruct the stimulus $s(t)$ given the response $\rho(t)$?

Discrimination task: given a set of a stimulus/response, how *sensitive* is the model to certain features??

Motivation

1. What aspects of the stimulus are encoded/thrown away? What are the invariances of the neural code?
2. How good is a particular encoding model in telling us about the stimulus?

4.1 Some decoders

1. **Regression solution:** $s_{est}(t - \tau_0) = \int K_1(\tau)(\rho(t - \tau) - \tilde{r})d\tau + \int d\tau_1 d\tau_2 K_2(t - \tau_1, t - \tau_2)\rho(t - \tau_1)\rho(t - \tau_2) + \dots$ The idea is that we try to predict the value of a stimulus at time $t - \tau_0$ by convolving filters with spikes occurring up till time t . τ_0 is called the 'prediction delay' and represents how far into the future we can use spikes to estimate a

stimulus at a given time.

Optimal linear kernel $K^*(\tau)$ satisfies: $\int Q_{\rho\rho}(\tau - \tau')K^*(\tau')d\tau' = Q_{rs}(\tau - \tau_0)$ Thus if the spike train is uncorrelated ($Q_{\rho\rho}(\tau) = \tilde{r}\delta(\tau)$) then we have $K^*(\tau) = \frac{1}{\tilde{r}}Q_{rs}(\tau - \tau_0) = C(\tau_0 - \tau)$, i.e. the optimal linear decoder is the STA, which gives the intuitive rule: Every time there is a spike, add the STA timeshifted by τ_0 , then add up to get the estimated stimulus $s_{est}(t)$.

2. **Bayes-optimal solution:** $\mathbb{E}[s|\rho(t)] \propto \int_{\mathbb{R}^n} xP(\rho(t)|x)P(x)dx$

More generally, we want to minimize the expected value of a loss function $L(s, \hat{s})$ given the response $\rho(t)$. When L is mean-squared error, \hat{s} is the conditional mean of $P(x|\rho)$ (above). If L is absolute error, we get the conditional median of $P(x|\rho)$. Generally, this is extremely hard to compute since this is an n -dimensional integral in stimulus space. Under the 'Laplace approximation' (i.e. the posterior can be approximated by a Gaussian - usually valid when it is unimodal, continuous, and smooth to some extent), however, the MAP estimate gives a much more tractable estimator.

3. **MAP solution:** $\hat{s}_{MAP} = \arg \max_{x \in \mathbb{R}^n} P(x|\rho(t)) = \arg \max_{x \in \mathbb{R}^n} [\log P(\rho(t)|x) + \log P(x)]$ This is easy to compute because it is just an optimization problem in \mathbb{R}^n . Note that if $P(s)$ is independent of s (e.g., uniform prior), the MAP estimate is equivalent to an ML estimate. In the case of the GLM with log-concave prior $P(s)$, the problem is even easier since the log posterior probability is log-concave and so there is only 1 maximum. However, the MAP is not accurate if the 'Laplace approximation' for the posterior distribution does not hold (i.e. it is not the case that $p(x|\rho(t)) \approx N(\hat{x}_{MAP}, C)$ where C is the inverse of the Hessian at \hat{x}_{MAP}).

4.2 Some discriminability measures

1. **ROC and likelihood ratio:** Suppose we sample observations, each coming from 1 of 2 possible distributions P_+, P_- . Given a single observation X , what test can we use to discern whether it came from P_+ or P_- . Consider the class of threshold tests 'choose P_+ if $(X \geq t)$.' We can plot the power ($P_+(X \geq t)$) vs. the size ($P_-(X \geq t)$) for various thresholds t to get the ROC curve. By the Neyman-Pearson Lemma,

we know the optimal test (in terms of size/power domination) is the threshold rule acting on the likelihood ratio $R(X) = P_+(X)/P_-(X)$. If $R(\cdot)$ is a monotonic increasing function, then threshold tests on the observation ($X \geq t$) are equivalent and therefore also optimal. A few interesting facts:

- (a) In the 2AFC domain (we have to pair 2 stimuli X_1, X_2 with 2 distributions P_+, P_-), we can use X_2 as a threshold for X_1 . Then $P(\text{correct}) = \int P_+(X_1 \geq x_2) dP_-(x_2)$. Using the expression for the size $\alpha(t) = \int_t^\infty P_-(x) dx$, we can do a change of variables to get $P(\text{correct}) = \int_0^1 \beta d\alpha = (\text{Area under ROC curve})$.
- (b) Writing the power $\beta(t) = \int_t^\infty P_+(x) dx$ we can show that $\frac{d\beta}{d\alpha}(x) = R(x)$, i.e. the likelihood ratio is the slope of the ROC curve

2. **Discrimination in Bayesian setting:** Given the posterior on stimuli given the response $P(x|\rho(t))$, one measure of sensitivity of the model to a stimulus direction u is to look at the leading terms of the Taylor expansion of the posterior at a fixed stimuli (say \hat{x}_{MAP}) (gradient, Hessian).