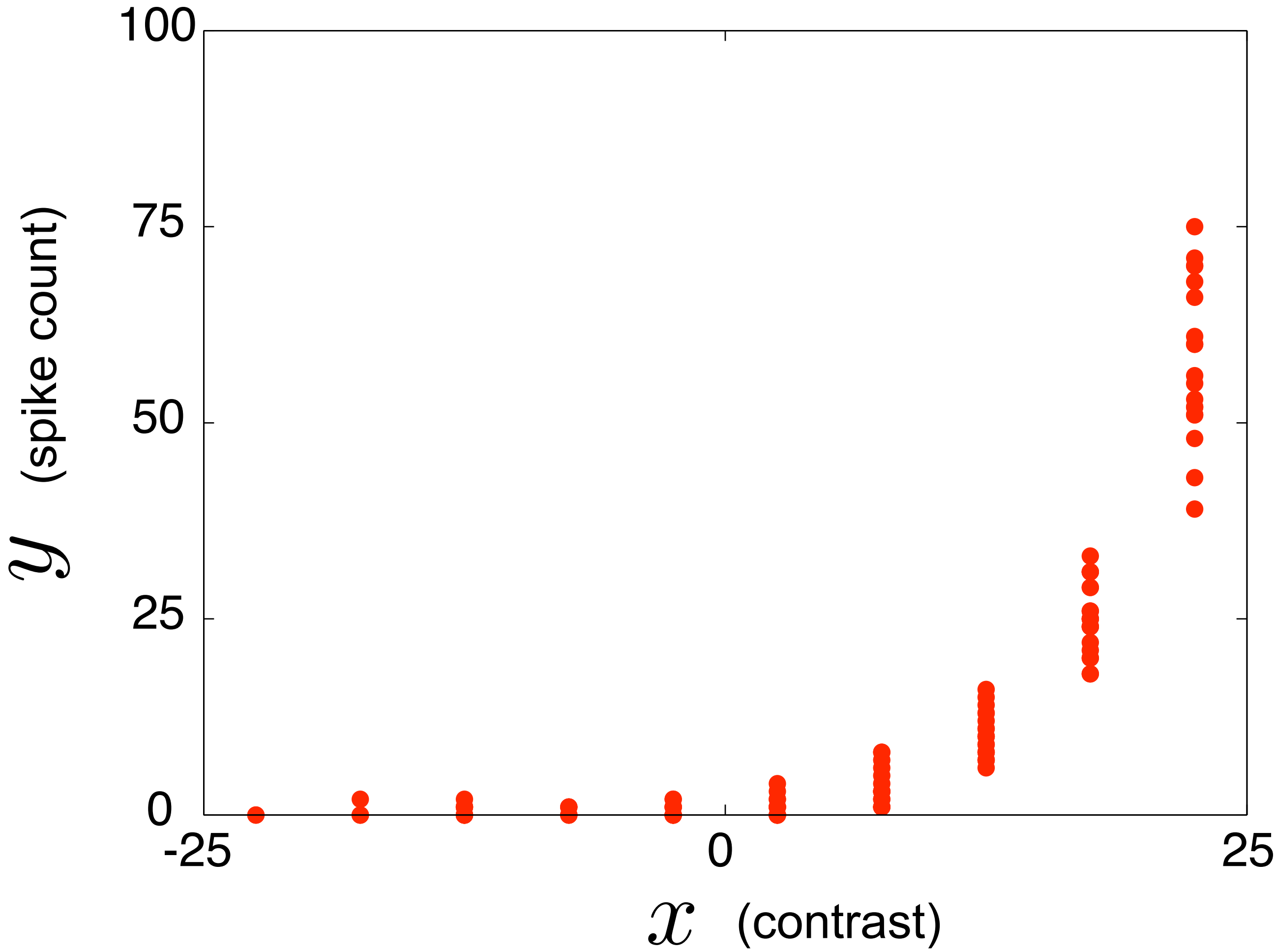
Generalized Linear Models & Logistic Regression

Jonathan Pillow

Mathematical Tools for Neuroscience (NEU 314)
Spring, 2016

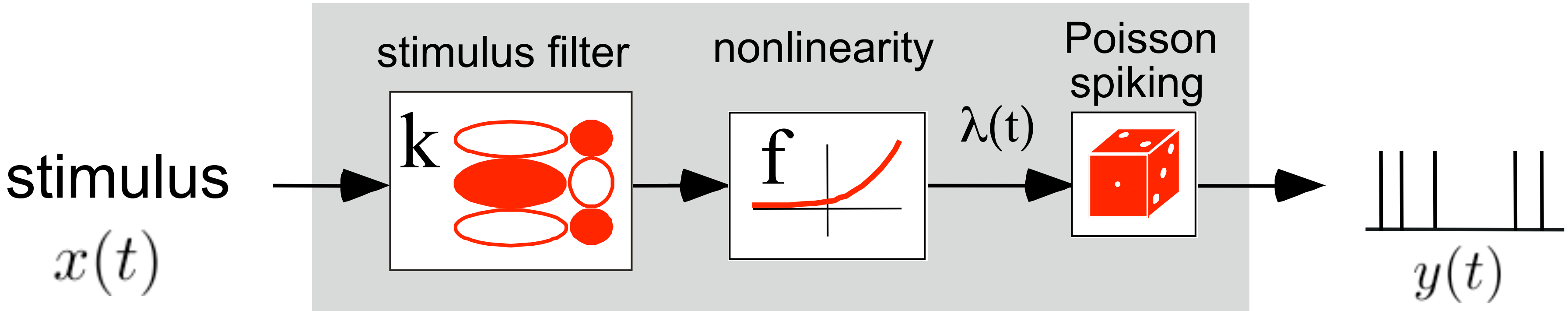
lecture 17

Example 3: unknown neuron



What model would you use to fit this neuron?

Linear-Nonlinear-Poisson model



conditional intensity ("spike rate")

$$\lambda(t) = f(k \cdot x(t))$$

Poisson spiking

$$y(t)|x(t) \sim \mathrm{Poiss}(\lambda(t))$$

- example of *generalized linear model* (GLM)

Aside on GLMs:

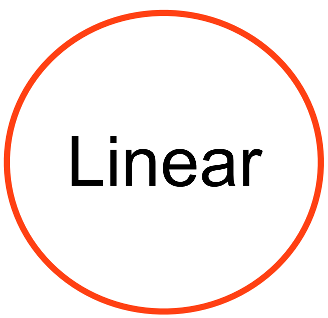
1. Be careful about terminology:

GLM ≠ GLM

General Linear Model

Generalized Linear Model

(Nelder 1972)



2003 interview with John Nelder...

Stephen Senn: I must confess to having some confusion when I was a young statistician between general linear models and generalized linear models. Do you regret the terminology?

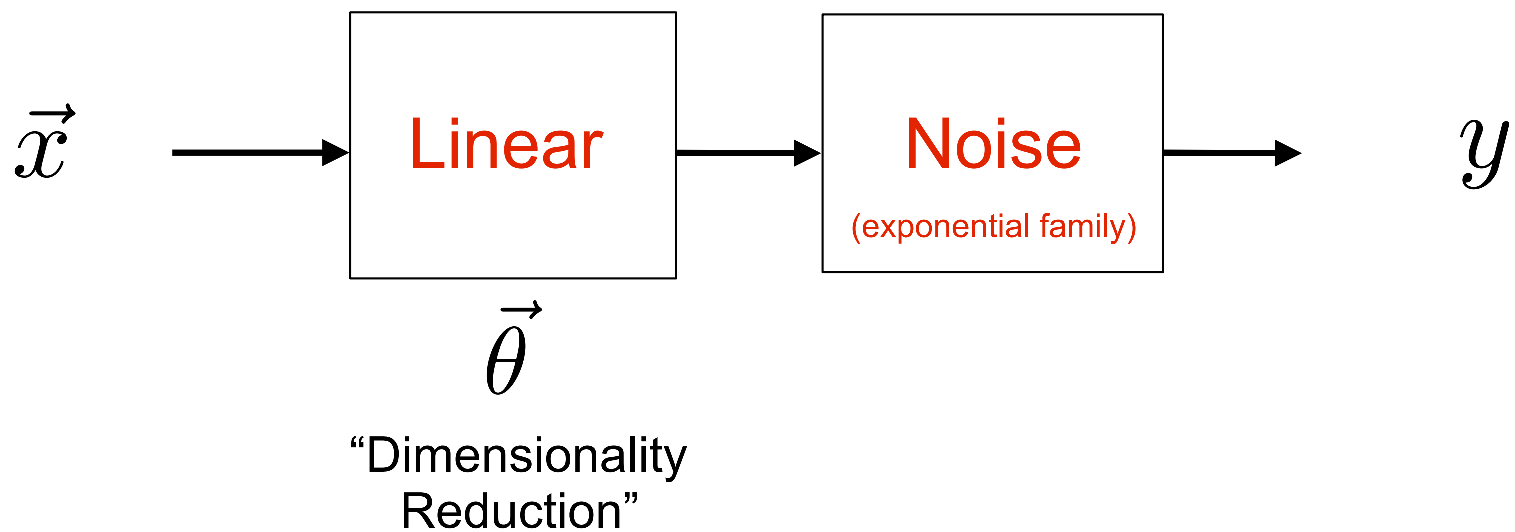
John Nelder: I think probably I do. I suspect we should have found some more fancy name for it that would have stuck and not been confused with the general linear model, although general and generalized are not quite the same. I can see why it might have been better to have thought of something else.

Senn, (2003). *Statistical Science*

Moral:

Be careful when naming your model!

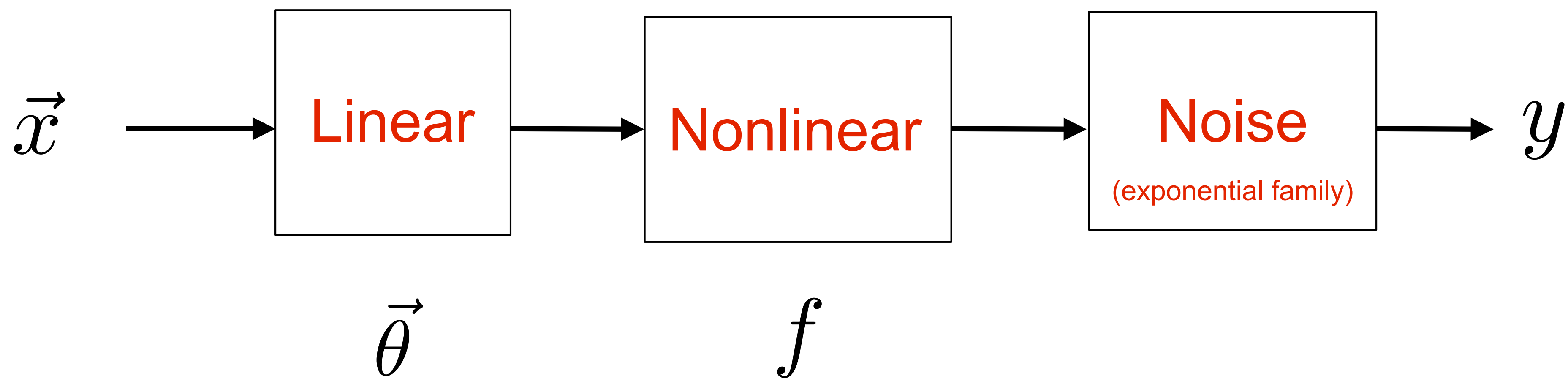
2. General Linear Model



Examples:

1. Gaussian $$y = \vec{\theta} \cdot \vec{x} + \sigma^2 \epsilon$$
2. Poisson $$y \sim \text{Poiss}(\vec{\theta} \cdot \vec{x})$$

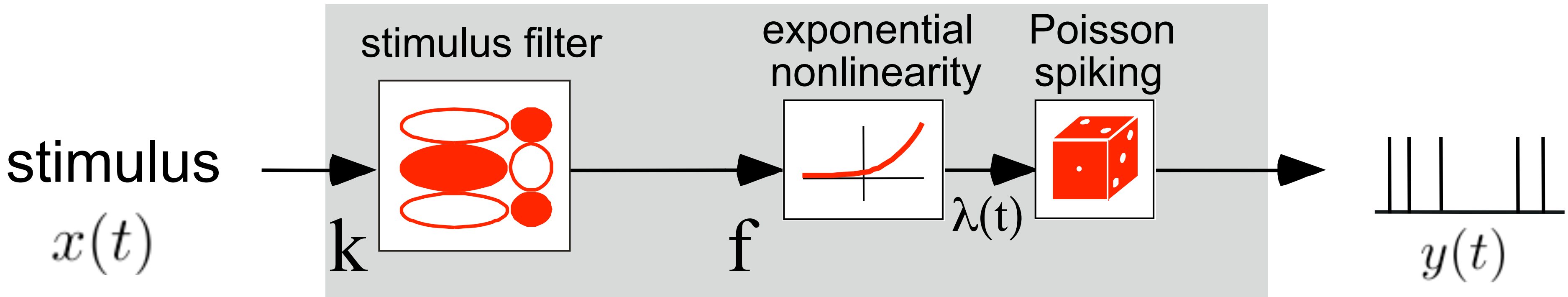
3. Generalized Linear Model



Examples:

1. Gaussian $\quad y = f(\vec{\theta} \cdot \vec{x}) + \sigma^2 \epsilon$
2. Poisson $\quad y \sim \mathrm{Poiss}(f(\vec{\theta} \cdot \vec{x}))$

Linear-Nonlinear-Poisson



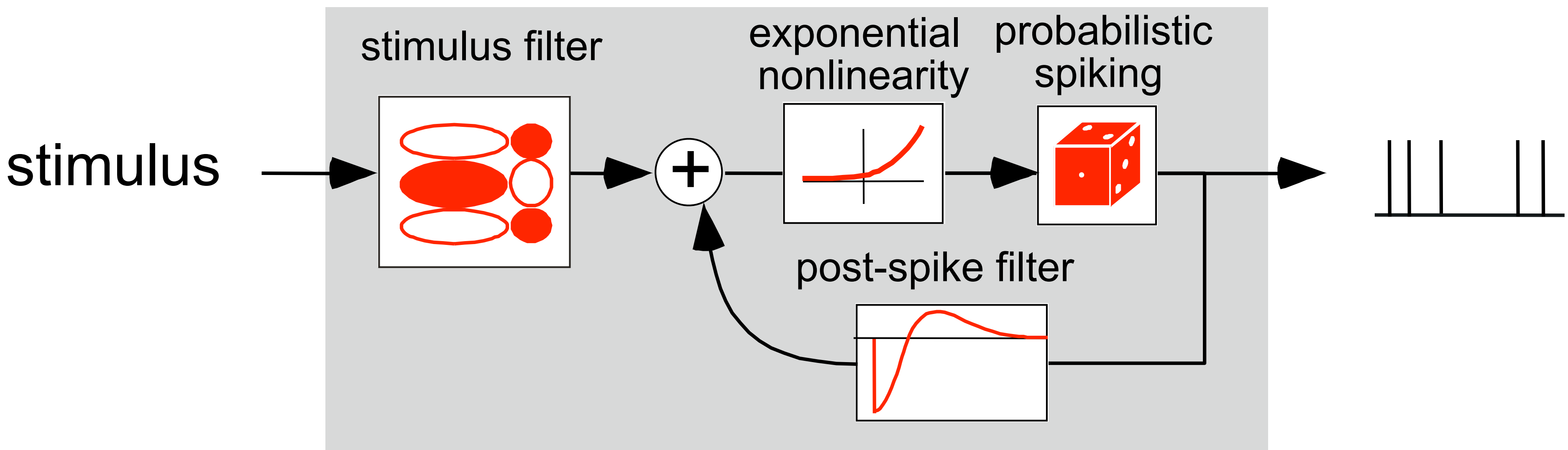
conditional intensity
("spike rate")

$$\lambda(t) = f(k \cdot x(t))$$

- output: Poisson process

(Truccolo et al 04)

GLM with spike-history dependence



conditional intensity
(spike rate)

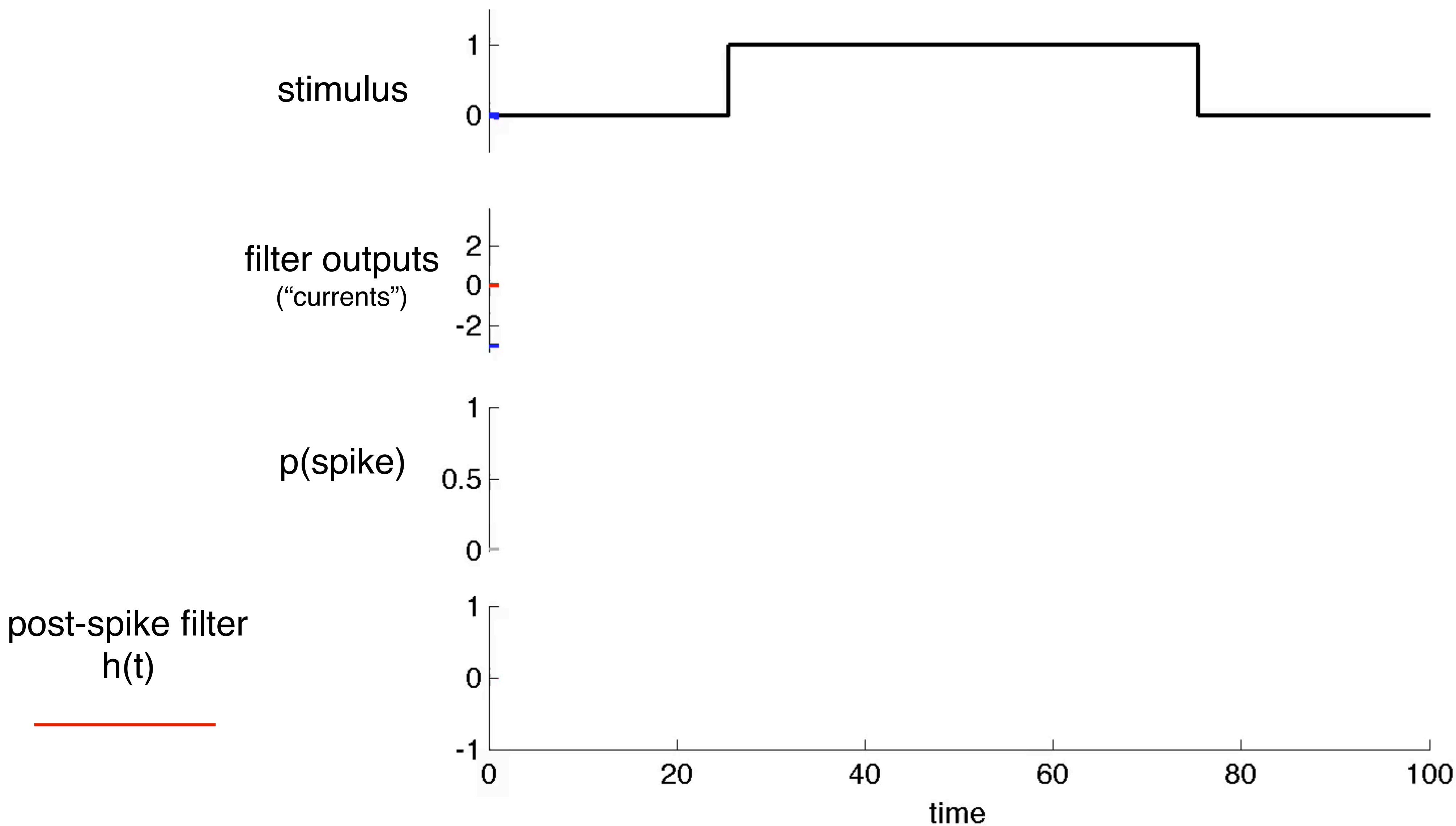
$$\lambda(t) = f(\vec{k} \cdot \vec{x}(t) + \vec{h} \cdot \vec{y}_{hst}(t))$$

$$= e^{\vec{k} \cdot \vec{x}(t)} \cdot e^{\vec{h} \cdot \vec{y}_{hst}(t)}$$

- output: product of stimulus and spike-history term

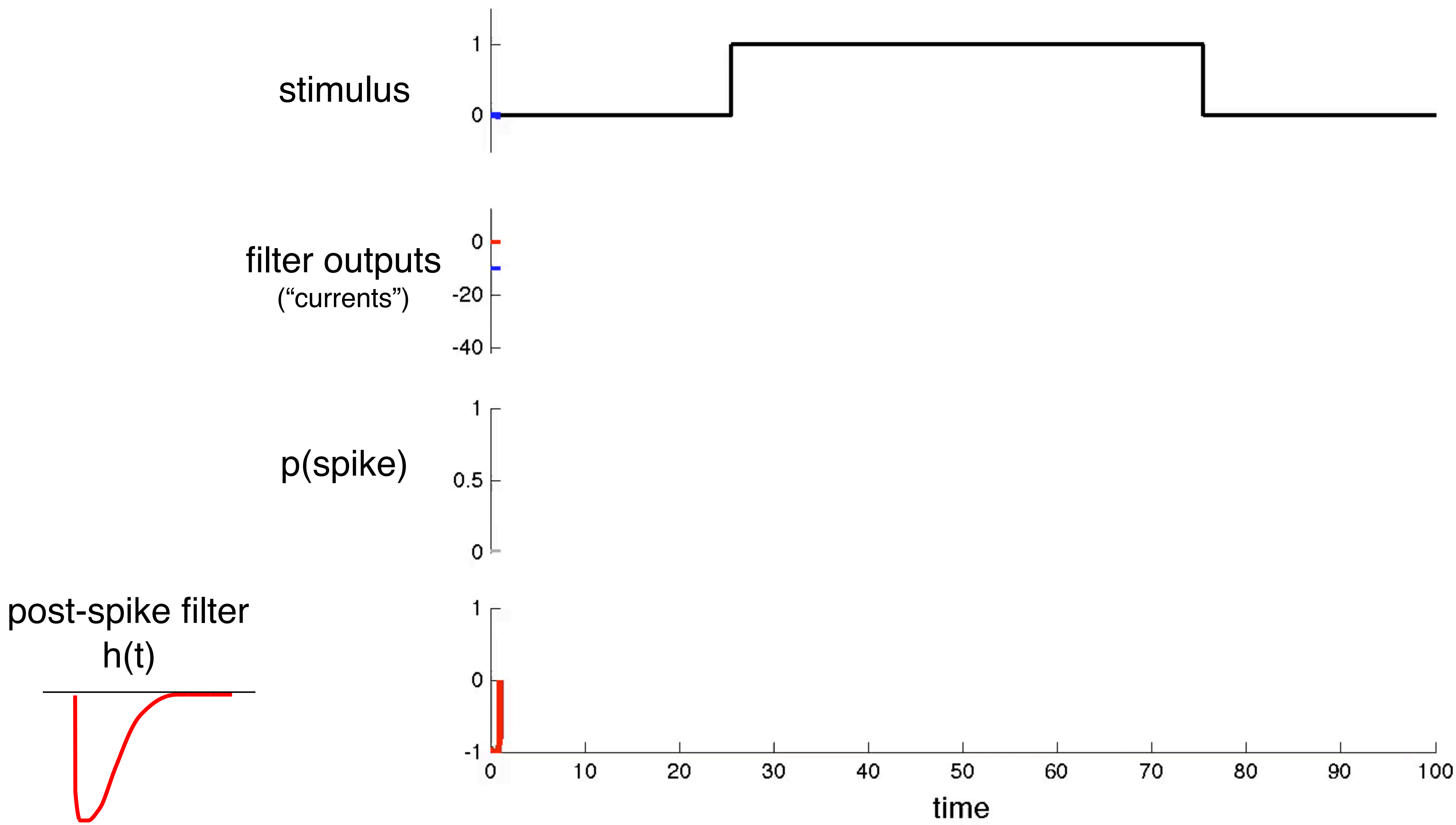
GLM dynamic behaviors

- irregular spiking (Poisson process)



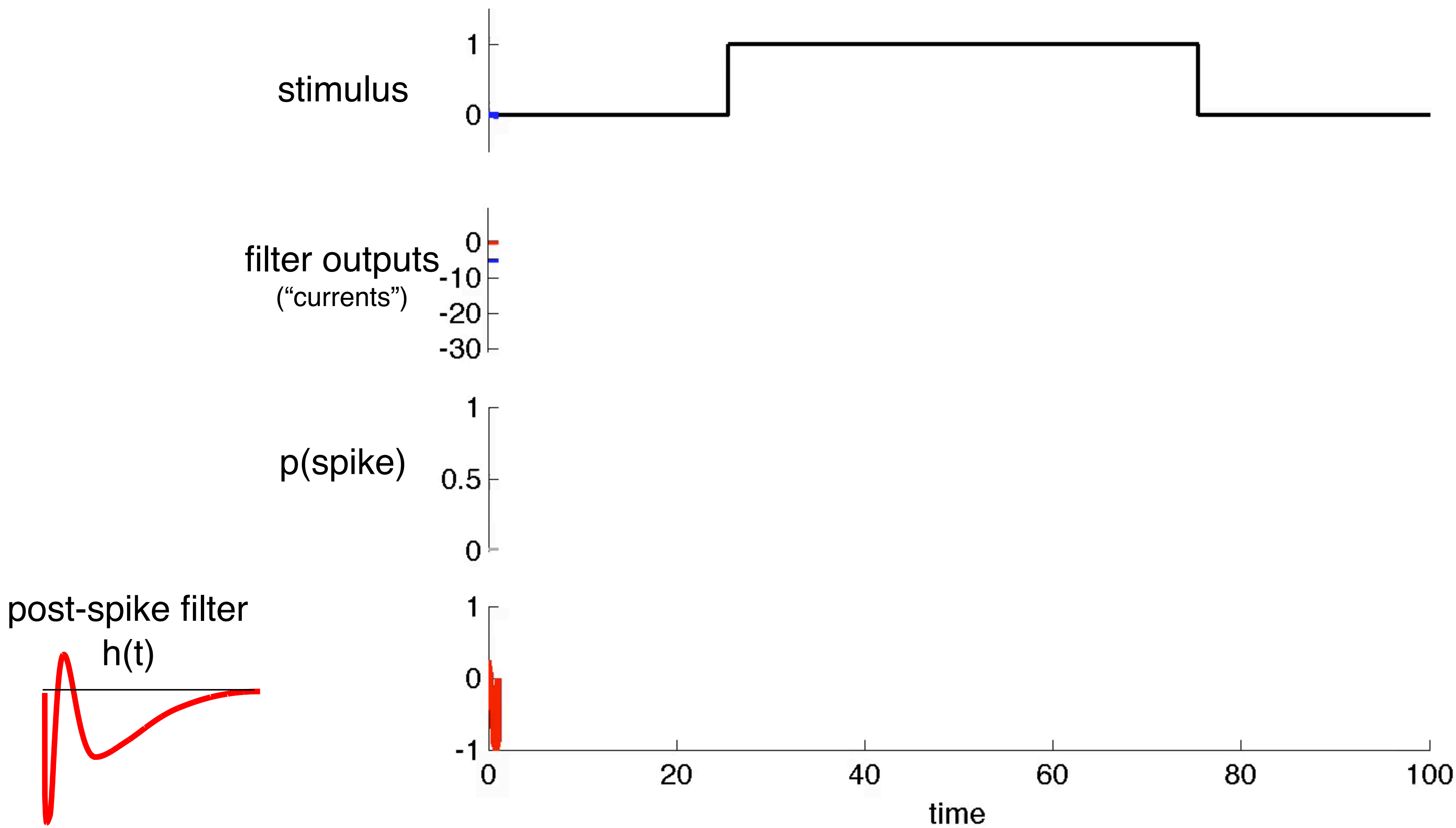
GLM dynamic behaviors

- regular spiking



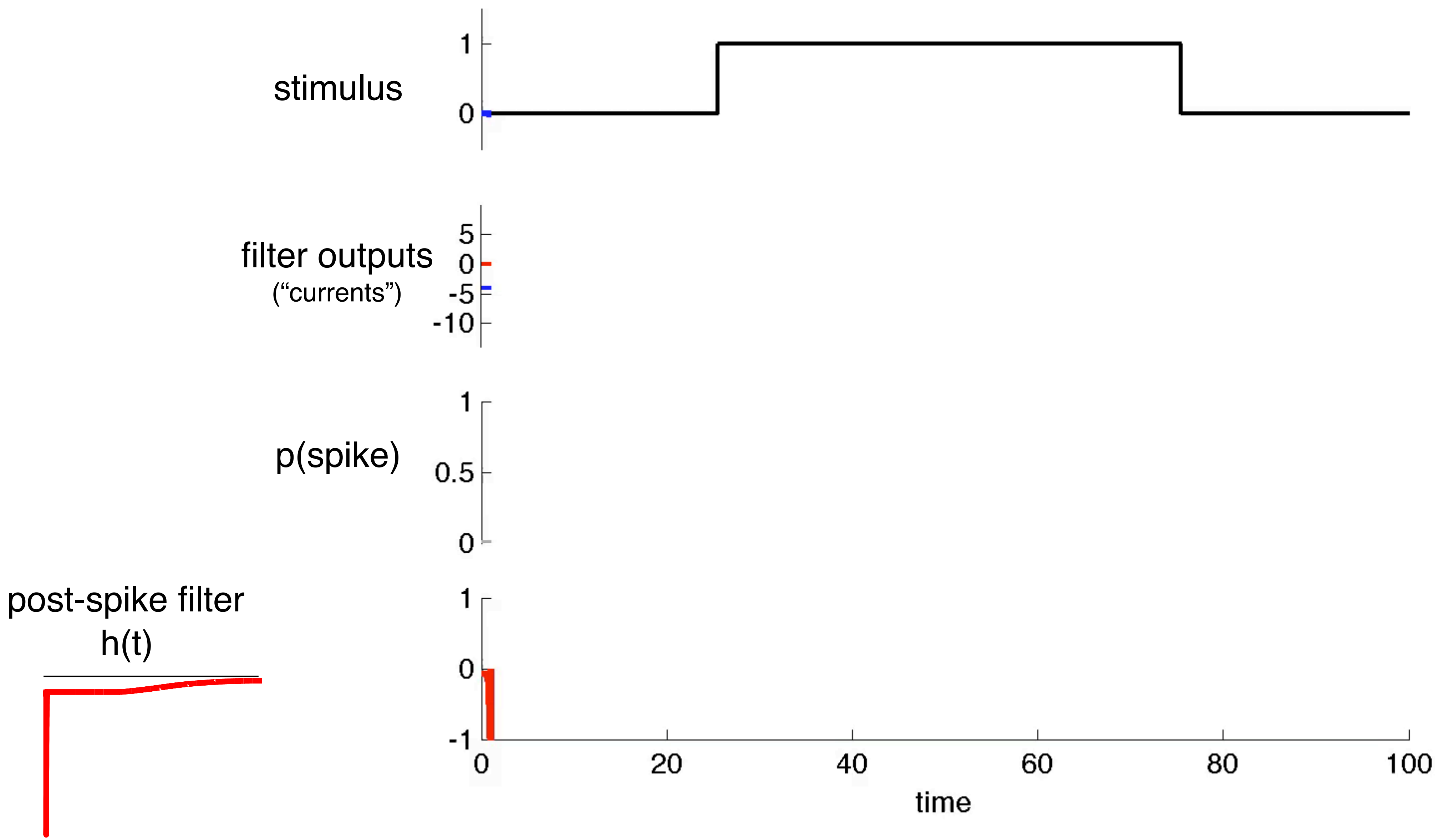
GLM dynamic behaviors

- bursting

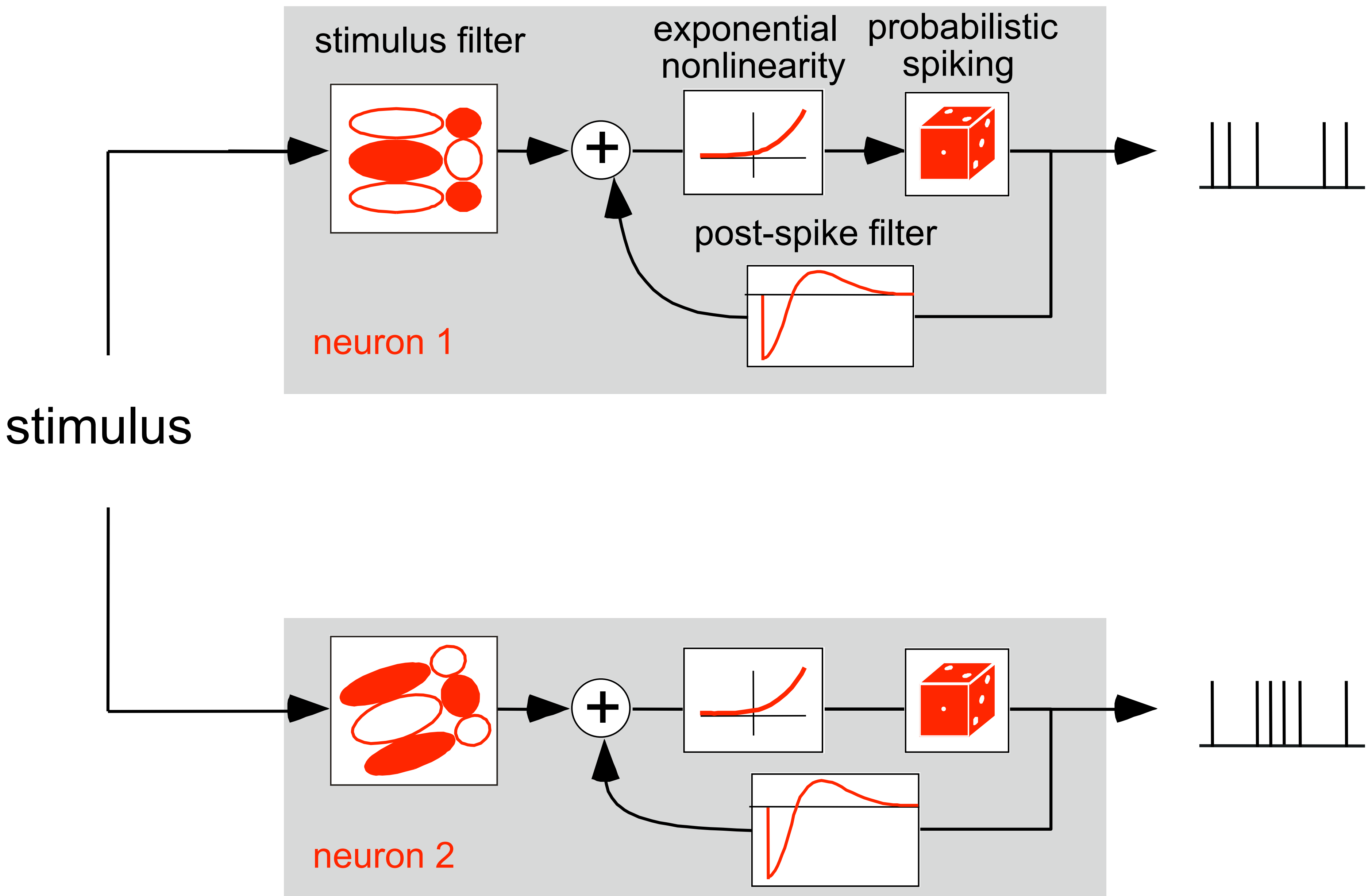


GLM dynamic behaviors

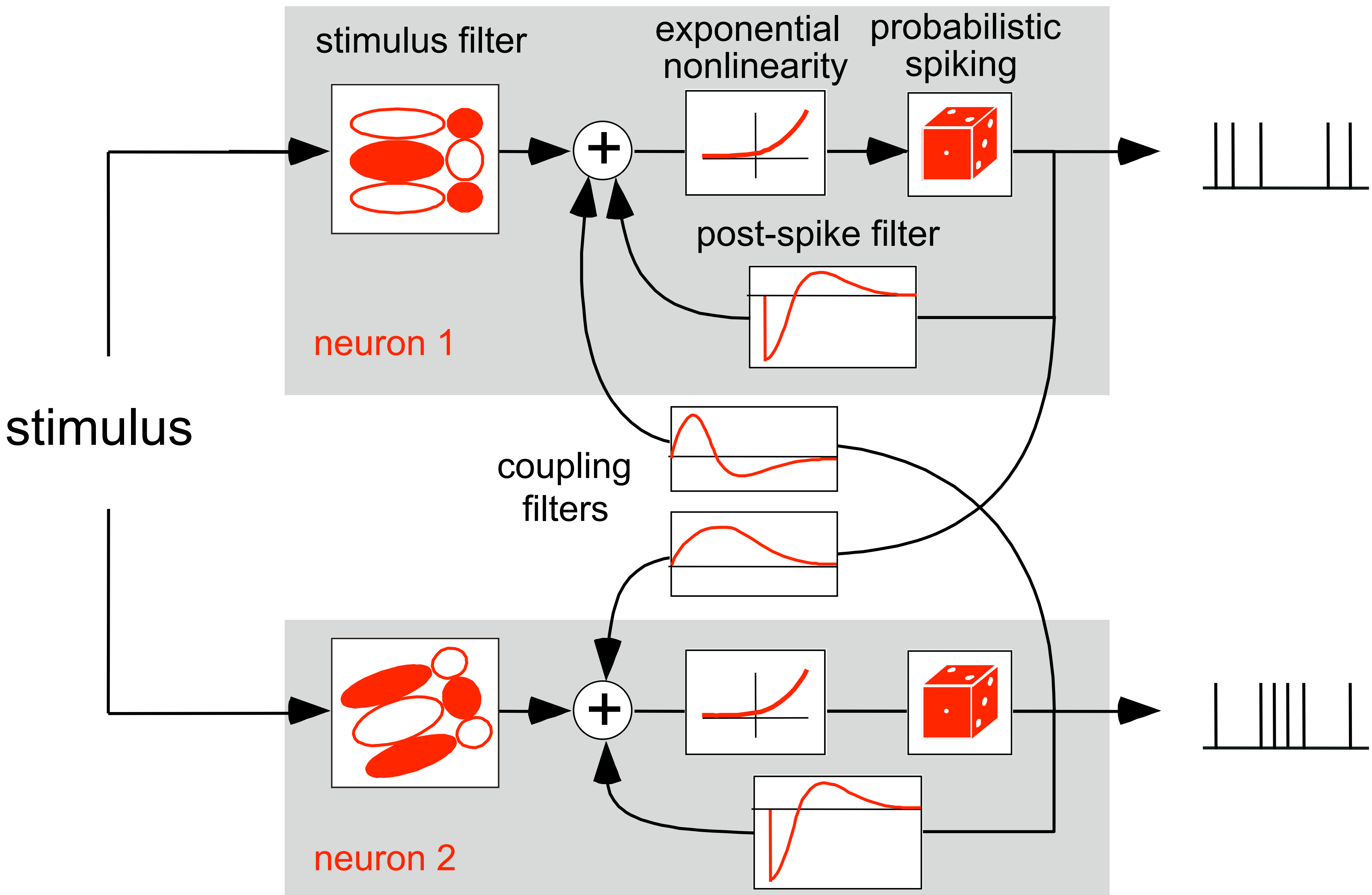
- adaptation



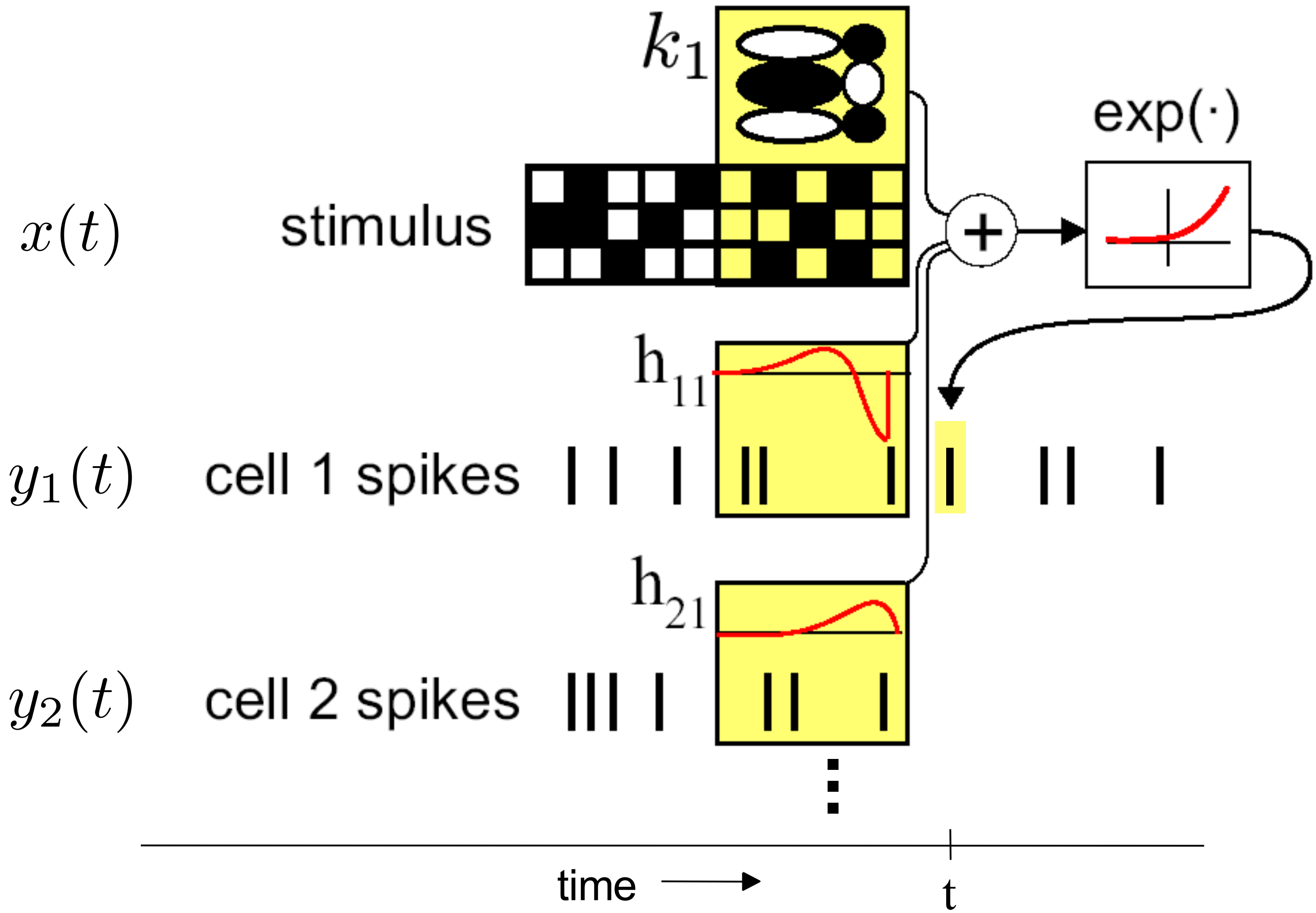
multi-neuron GLM



multi-neuron GLM



GLM equivalent diagram:

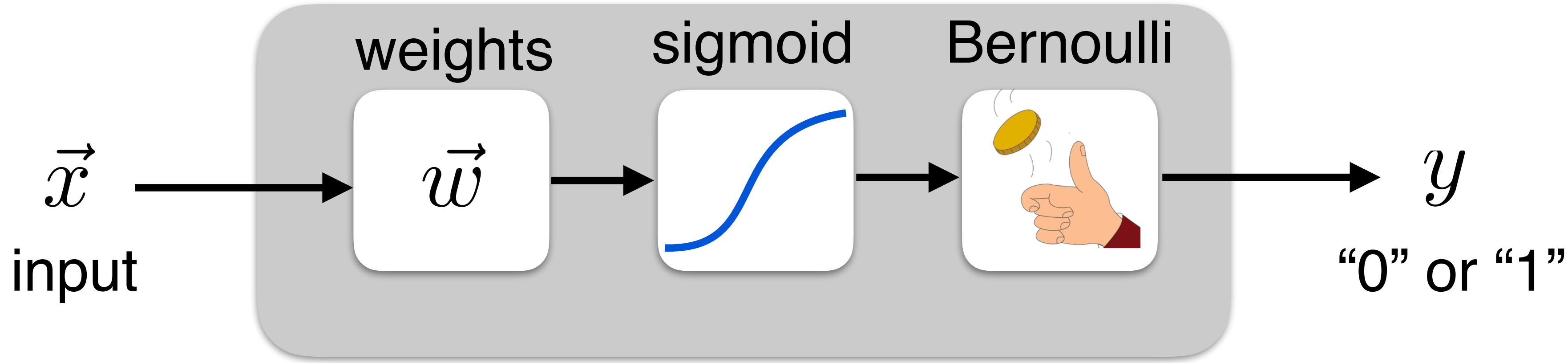


conditional intensity
(spike rate)

$$\lambda_i(t) = \exp(k_i \cdot x(t) \; + \; \sum_j h_{ij} \cdot y(t))$$

Logistic Regression

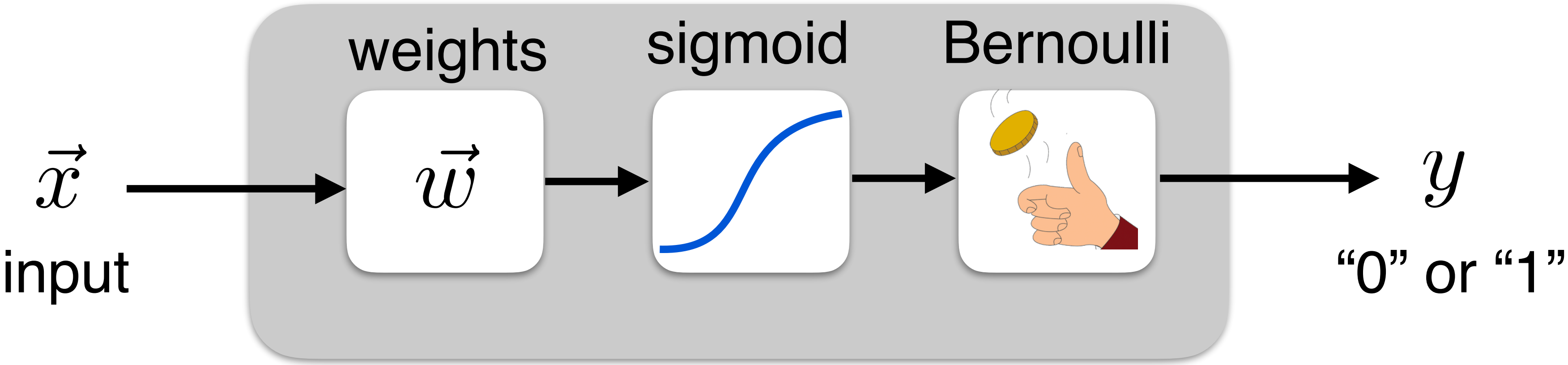
GLM for binary classification



1. Linear weights $$z = \vec{w} \cdot \vec{x}$$
2. sigmoid (“logistic”) function $$f(z) = \frac{1}{1 + e^{-z}}$$
3. Bernoulli (coin flip) $$P(y = 1) = f(z)$$ $$P(y = 0) = 1 - f(z)$$

Logistic Regression

GLM for binary classification



$$z = \vec{w} \cdot \vec{x}$$

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$P(y = 1) = f(z)$$
$$P(y = 0) = 1 - f(z)$$

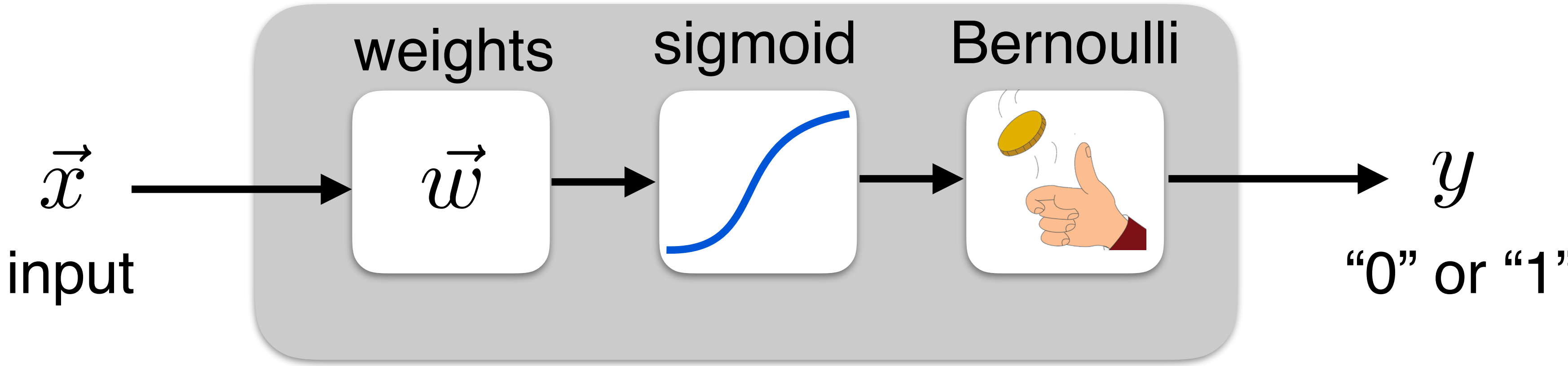
compact expression:

$$P(y|\vec{x}, \vec{w}) = \frac{e^{y(\vec{w}\cdot\vec{x})}}{1 + e^{\vec{w}\cdot\vec{x}}}$$

(note when y = 1, this is equal to exp(wx)/(1+exp(wx)), which is equal to 1/(1+exp(-wx))

Logistic Regression

GLM for binary classification



compact expression:

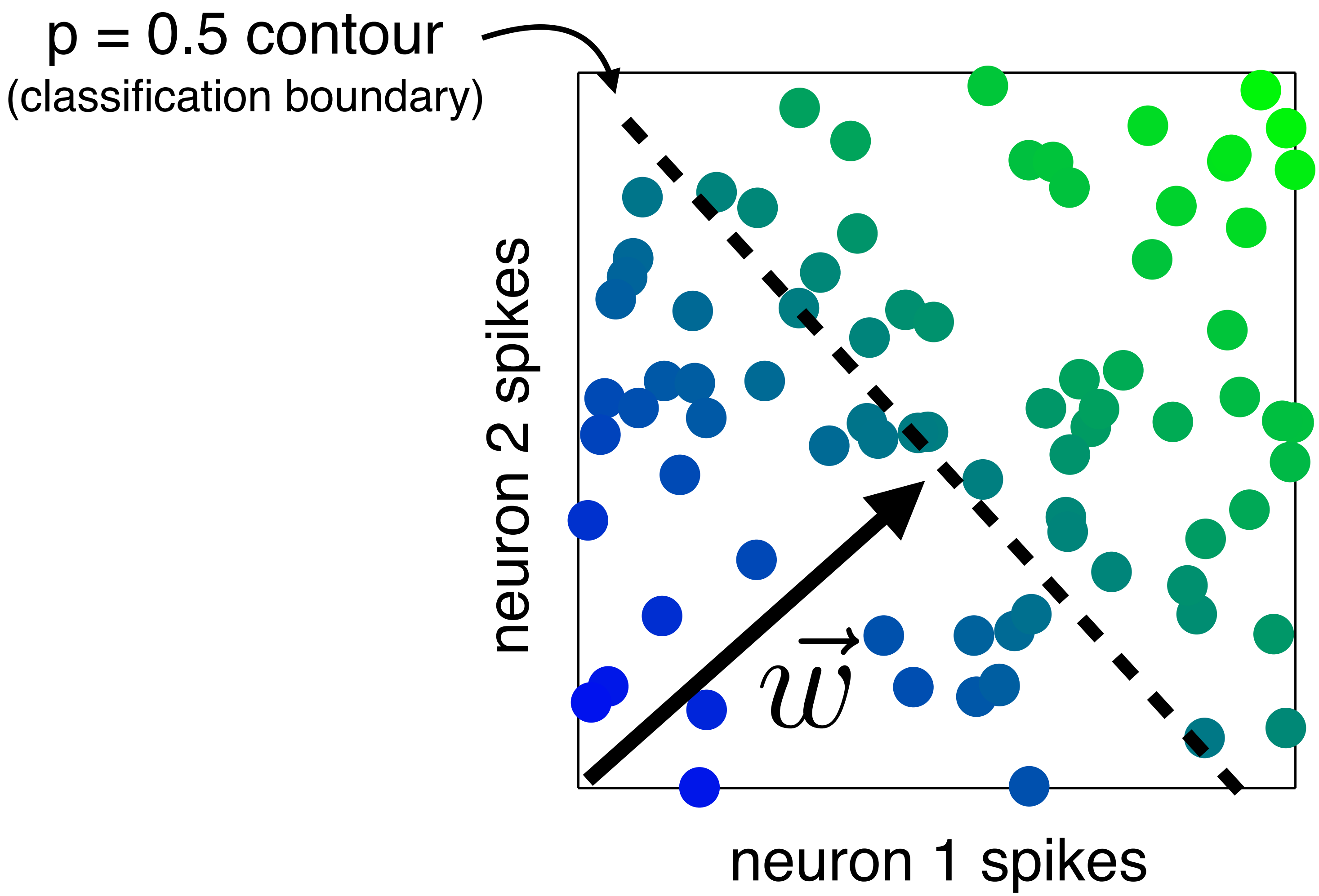
$$P(y|\vec{x}, \vec{w}) = \frac{e^{y(\vec{w}\cdot\vec{x})}}{1 + e^{\vec{w}\cdot\vec{x}}}$$

fit w by maximizing log-likelihood:

$$\log P(Y|X, w) = \sum_i \left[y_i x_i^\top w - \log(1 + \exp(x_i^\top w) \right]$$

Logistic Regression

geometric view



Bayesian Estimation

three basic ingredients:

1. Likelihood $p(m|\theta)$
2. Prior $p(\theta)$

} jointly determine the posterior $p(\theta|m)$

3. Loss function $L(\hat{\theta}, \theta)$ } "cost" of making an estimate $\hat{\theta}$ if the true value is θ

- fully specifies how to generate an estimate from the data

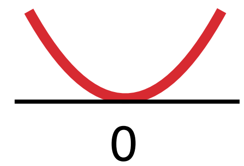
Bayesian estimator is defined as:

$$\hat{\theta}(m) = \arg\min_{\hat{\theta}} \boxed{\int L(\hat{\theta}, \theta) p(\theta|m) d\theta}$$

"Bayes' risk"

Typical Loss functions and Bayesian estimators

1. $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ squared error loss



need to find $\hat{\theta}$ minimizing the expected loss: $\int (\hat{\theta} - \theta)^2 p(\theta|m) d\theta$

Differentiate with respect to $\hat{\theta}$ and set to zero:

$$\int 2(\hat{\theta} - \theta) p(\theta|m) d\theta = 0$$

$$\int \hat{\theta}\, p(\theta|m) d\theta = \int \theta\, p(\theta|m) d\theta$$

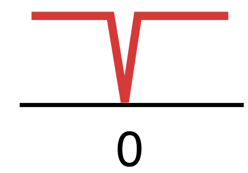
$$\hat{\theta} = \int \theta\, p(\theta|m) d\theta$$ “posterior mean”

also known as Bayes’ Least Squares (BLS) estimator

Typical Loss functions and Bayesian estimators

2. $$L(\hat{\theta}, \theta) = 1 - \delta(\hat{\theta} - \theta)$$

"zero-one" loss
(1 unless $\hat{\theta} = \theta$)



expected loss: $$\int (1 - \delta(\hat{\theta} - \theta)) p(\theta|m) d\theta$$

$$= 1 - p(\hat{\theta}|m)$$

which is minimized by: $$\hat{\theta} = \arg\max_{\theta} p(\theta|m)$$

- posterior maximum (or "mode").
- known as maximum a posteriori (MAP) estimate.