Neural Coding 2007 Abstract p120-123, Montevideo, Uruguay

Kernel Width Optimization in the Spike-rate Estimation

Hideaki Shimazaki
Department of Physics, Kyoto University
Kyoto 606-8502, Japan
shimazaki@ton.scphys.kyoto-u.ac.jp

Shigeru Shinomoto
Department of Physics, Kyoto University
Kyoto 606-8502, Japan
shinomoto@scphys.kyoto-u.ac.jp

ABSTRACT

A classical tool for estimating the neuronal spike rate is a peri-stimulus time histogram (PSTH) constructed from spike sequences aligned at the onset of a stimulus repeatedly applied to an animal. We have recently established a method for selecting the bin size, so that the PSTH best represents the unknown underlying rate [1]. The goodness of the fit we adopted as the optimization principle is minimizing the mean integrated squared error (MISE) between the underlying rate λ_t and the PSTH $\hat{\lambda}_t$,

$$\text{MISE} = \int_a^b E(\lambda_t - \hat{\lambda}_t)^2 dt, \tag{1}$$

where E refers to the expectation with respect to the spike generation process under a given time-dependent rate λ_t. The method allows us to minimize the MISE from spike count statistics alone, without knowing the underlying rate.

In this contribution, we consider a kernel rate estimator as $\hat{\lambda}_t$ and suggest a method to select the width of a kernel under the MISE criterion. Generally, the cross-validation method is applicable to the least squares minimization [2, 3]. Here, we estimate the MISE fully utilizing the Poissonian nature of spikes, as we have done in the PSTH optimization. The Poissonian assumption holds in the limit of large number of trials, because spikes repeatedly recorded from a single neuron under identical experimental conditions are in the majority mutually independent.

For a small number of spike sequences generated from modestly fluctuating underlying rate, the optimal kernel width may become comparable to the observation period. This phenomenon, similar to what we have observed in the PSTH optimization [4], indicates that more experimental trials are needed if one wishes to uncover the time-dependent rate. We also construct a method for estimating the number of additional experimental trials needed to analyze the data with the resolution we deem sufficient.

Selection of the Kernel Width We consider independently and identically obtained n spike sequences, which contain N spikes as a whole. Due to the general limit theorem for the sum of independent point processes, a superposition of the spike sequences can be regarded as being drawn from a time-dependent Poisson point process. We define the superimposed sequence as $x_t = n^{-1} \sum_{i=1}^{N} \delta(t - t_i)$, where t_i is the timing of the ith spike. $\delta(t)$ is the Dirac delta function. An estimator $\hat{\lambda}_t$ is constructed by smoothing the point process by a kernel $k_w(t)$ of the bandwidth w: $\hat{\lambda}_t = \int k_w(t-s) x_s \, ds$.

We wish to obtain a kernel function that minimizes the MISE (Eq. 1). The integrand of the MISE is decomposed into three parts: $\lambda_t^2 - 2\lambda_t E\hat{\lambda}_t + E\hat{\lambda}_t^2$. Since the first component does not depend on the choice of a kernel, we subtract it from the MISE and define a cost function as a

Table 1: A Method for Selecting a Bandwidth of a Kernel Function

(i)	Superimpose all the n spike sequences. Obtain a series of spike times $\{t_i\}_{i=1}^{N}$ in $[a,b]$. N is the total number of spikes.
(ii)	Compute the cost function of a kernel $k_w(t)$ as $$\hat{C}_n(w) = -\frac{4}{n^2}\sum_{i<j} k_w(t_i - t_j) + \frac{1}{n^2}\sum_{i,j}\psi_{w,a,b}(t_i - t_j),$$ where $\psi_{w,a,b}(t) \equiv \int_a^b k_w(s)\, k_w(s+t)\, ds$ is the correlation function $^{(*)}$.
(iii)	Repeat ii while changing w to search for w^* that minimizes $\hat{C}_n(w)$.

$^{(*)}$ For a Gaussian kernel $k_w(t) = \frac{1}{\sqrt{2\pi}w}\exp\left(-\frac{t^2}{2w^2}\right)$,
$\psi_{w,a,b}(t) = \frac{1}{\sqrt{\pi}4w}\exp\left(-\frac{t^2}{4w^2}\right)\left\{\mathrm{erf}\left(\frac{2b+t}{2w}\right) - \mathrm{erf}\left(\frac{2a+t}{2w}\right)\right\}$.

function of the kernel, or its bandwidth w,

$$C_n(w) \equiv \text{MISE} - \int_a^b \lambda_t^2\, dt = -2\int_a^b \lambda_t E\hat{\lambda}_t\, dt + \int_a^b E\hat{\lambda}_t^2\, dt. \tag{2}$$

From a general decomposition rule of a covariance of two random variables, we obtain the relation,

$$\begin{aligned}\int_a^b \lambda_t E\hat{\lambda}_t\, dt &= \int_a^b E[x_t\hat{\lambda}_t]\, dt - \int_a^b E(x_t - Ex_t)(\hat{\lambda}_t - E\hat{\lambda}_t)\, dt \\ &= E\int_a^b x_t\hat{\lambda}_t\, dt - \frac{k_w(0)}{n}E\int_a^b x_t\, dt. \end{aligned}\tag{3}$$

To obtain the last equality, we used the assumption that the spikes are independent each other (an assumption of a Poisson point process). Hence from sample sequences, the cost function is estimated as

$$\begin{aligned}\hat{C}_n(w) &= \frac{2k_w(0)}{n}\int_a^b x_t\, dt - 2\int_a^b x_t\hat{\lambda}_t\, dt + \int_a^b \hat{\lambda}_t^2\, dt \\ &= \frac{2k_w(0)}{n^2}N - \frac{2}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{N} k_w(t_i - t_j) + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\psi_{w,a,b}(t_i - t_j), \end{aligned}\tag{4}$$

where $\psi_{w,a,b}(t)$ is given by

$$\psi_{w,a,b}(t) \equiv \int_a^b k_w(s)\, k_w(s+t)\, ds. \tag{5}$$

We summarize the method for selecting the width of a symmetric kernel function as Table 1. Figure 1 displays application of our method to the spike data of an MT neuron responding to a random dot stimulus [5]. In this example, we changed the number of spike sequences used to estimate the time-dependent rate.

Extrapolation of the cost function Here we provide the method to estimate how many additional experimental trials are needed to make an optimized rate estimate with the required resolution. Suppose we have n spike sequences. The superposition of n spike sequences is denoted as $x_t^{(n)}$. The kernel estimator is denoted as $\hat{\lambda}_t^{(n)}$.

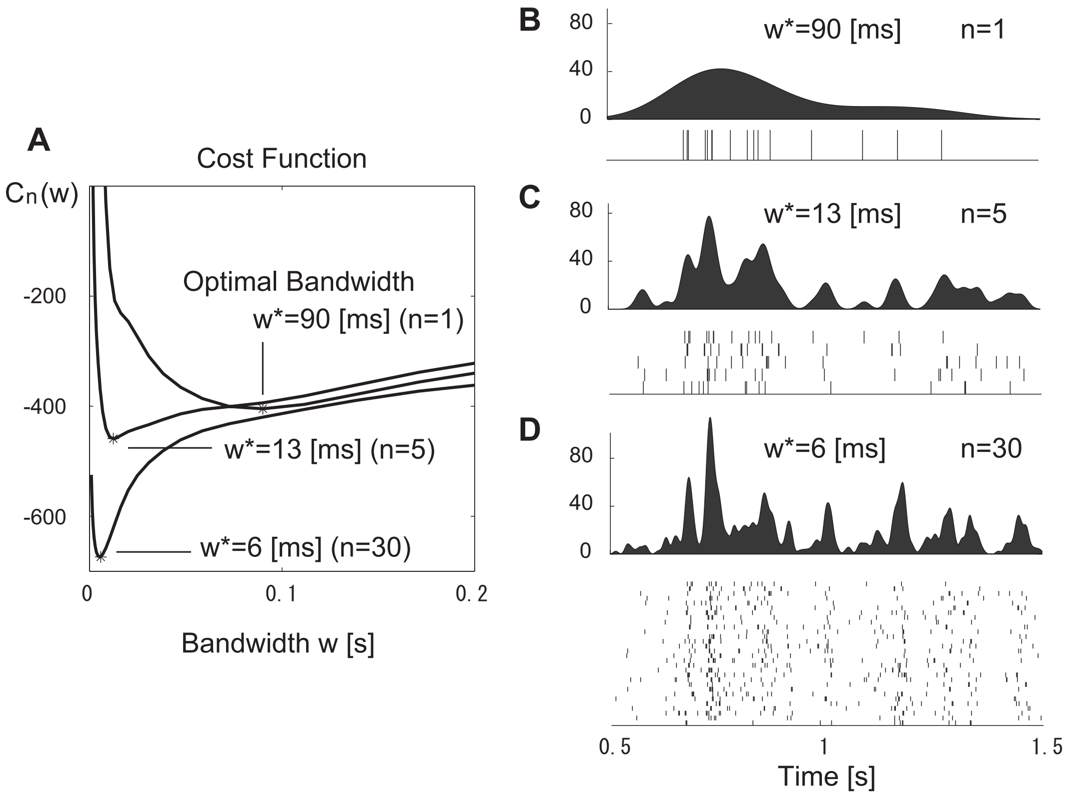

Figure 1: Optimization of a kernel estimate for spike-rate of an MT neuron (j024 in nsa2004.1 [5]) with a Gaussian kernel $(\exp\left(-t^2/2w^2\right)/\sqrt{2\pi w})$, according to the method in Table 1. A: Cost functions of $n = 1$, 5, and 30 spike sequences. B, C, and D: (Top) Optimized time-dependent rates. (Bottom) Spike sequences.

We wish to compute the cost function for m spike sequences. It is easily proved that the cost function for m spike sequences is written as

$$C_m(w) = \left(\frac{1}{m} - \frac{1}{n}\right) \int_a^b dt \int k_w\,(t-s)^2\, Ex_s^{(n)}\, ds + C_n(w), \tag{6}$$

where $C_n(w)$ is given by Eq. 2. To obtain Eq. 6, we used a relation, $\int E(\hat{\lambda}_t^{(m)} - E\hat{\lambda}_t^{(m)})^2\, dt = m^{-1} \int dt \int k_w\,(t-s)^2\, Ex_s^{(m)}\, ds$, and replaced $Ex_s^{(m)}$ with $Ex_s^{(n)}$. Hence the cost function for m spike sequences is estimated from a sample of n spike sequences as

$$\hat{C}_m\,[w\,|\,n] = \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n} \sum_{i=1}^{N} \int_{a-t_i}^{b-t_i} k_w\,(t)^2\; dt + \hat{C}_n(w). \tag{7}$$

The extrapolation method is summarized as Table 2. Figure 2 demonstrates the application of the extrapolation method to the spike data [5]. In Figure 2-b, the optimal widths were estimated with the extrapolated cost functions computed from the first two spike sequences (Solid line). They are close to the optimal kernel widths obtained by the method in Table 1 (Dots). With this plot, experimentalists can estimate how many additional experimental trials must be performed in order to achieve the resolution they deem sufficient.

Keywords: Optimization, Kernel Bandwidth, Rate Estimation, Spike Sequences

Table 2: A Method for Extrapolating the Cost Function

(A)	Compute the cost function $\hat{C}_n(w)$ as in Table 1.
(B)	Construct the extrapolated cost function: $$\hat{C}_m(w\,\|\,n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{1}{n}\sum_{i=1}^{N}\psi_{w,a-t_i,b-t_i}(0) + \hat{C}_n(w).$$
(C)	Repeat B while changing w to search for w^* that minimizes $\hat{C}_m(w\,\|\,n)$.

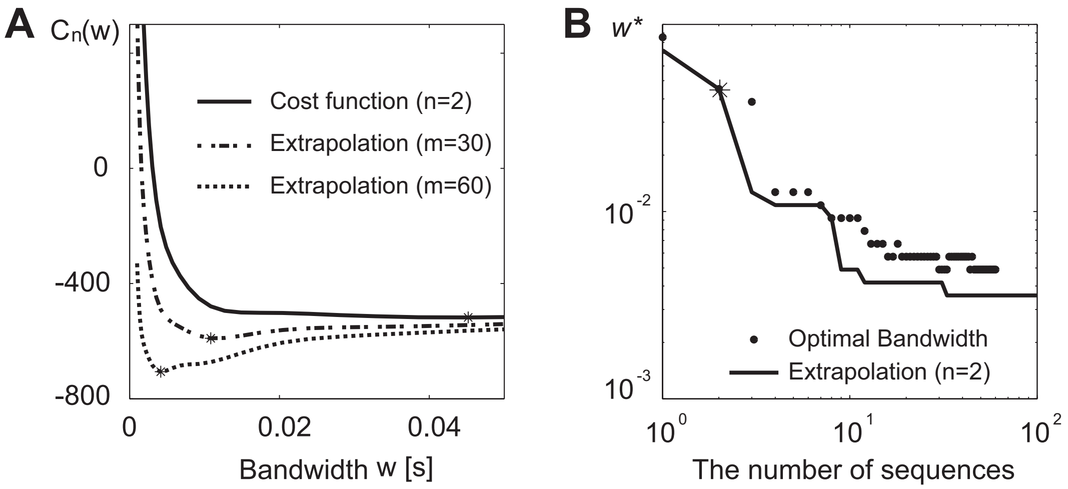

Figure 2: Application of the extrapolation method to spike data of an MT neuron (the data as in Fig. 1 [5]). A: A cost function of the first two spike sequences (Solid). Extrapolated cost functions for $m = 30$ and 60 computed from the first two spike sequences (Dashed and Dotted). B: Estimated optimal kernel widths of $n = 1$ through 60 spike sequences, according to the method in Table 1 (Dots). A solid line is the estimated optimal widths computed from the first 2 spike sequences, according to the method in Table 2.

References

[1] Shimazaki, H. and Shinomoto, S. (2007) A method for selecting the bin size of a time histogram, *Neural Computation*, **19**: 1503-1527.

[2] Rudemo, M. (1982) Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics*, **9**: 65-78.

[3] Bowman, A. W. (1984) An alternative method of cross-validation for the smoothing of density estimates, *Biometrika*, **71**: 353-360.

[4] Koyama, S. and Shinomoto, S. (2004) Histogram bin width selection for time-dependent Poisson processes, *Journal of Physics A-Mathematical and General*, **37**: 7255-7265.

[5] Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (2004). Responses of single neurons in macaque MT/V5 as a function of motion coherence in stochastic dot stimuli. *Neural Signal Archive*, nsa2004.1. http://www.neuralsignal.org.