# Towards Statistical Summaries of Spike Train Data

Wei Wu, Anuj Srivastava

*Department of Statistics, Florida State University*

---

**Abstract**

Statistical inference has an important role in analysis of neural spike trains. While current approaches are mostly model-based, and designed for capturing the temporal evolution of the underlying stochastic processes, we focus on a data-driven approach where statistics are defined and computed in function spaces where individual spike trains are viewed as points. The first contribution of this paper is to endow spike train space with a parametrized family of metrics that takes into account different time warpings and generalizes several currently used metrics. These metrics are essentially penalized $L^p$ norms, involving appropriate functions of spike trains, with penalties associated with time-warpings. The second contribution of this paper is to derive a notion of a mean spike train in the case when $p = 2$. We present an efficient recursive algorithm, termed Matching-Minimization algorithm, to compute the sample mean of a set of spike trains. The proposed metrics as well as the mean spike trains ideas are demonstrated using an experimental recording from the motor cortex.

*Keywords:* Spike train metrics, $L^p$ norms, statistical summaries, time warping, mean spike train.

---

## 1. Introduction

Time-dependent information is represented via sequences of stereotyped spike waveforms in the nervous system. A precise characterization and analysis of waveform sequences (or spike trains) is one of the central problems in neural coding. Due to the random nature of the spike trains, probabilistic and statistical methods have become important and have been investigated extensively to examine firing patterns (Rieke et al., 1997). In this regard, the parametric probability models, such as Poisson processes or general point processes, have been used predominantly (Tukey, 1977; Box et al., 1978; Kass and Ventura, 2001). These approaches have focused on modeling the biological processes underlying spike train data, using model-based parametric families that study the temporal aspects of the data. This modeling paradigm is currently a standard for deriving of statistical inferences from experimental results (Brown et al., 2002; Kass et al., 2005). However, such models have only a limited use in evaluating data-driven statistics on spaces where full spike trains are treated as individual data points. Given a sample of spike trains, one may naturally pose questions such as "What is the mean trend of the sample?" and "What is the variability of the sample?" Although these questions are very basic, the past models may not be able to answer them as they only characterize variability at each specific time. We propose a principled framework to address this issue, where we compute metric-based summary statistics in the spike train domain, and then suggest their use in developing statistical inferences.

The definition of summary statistics (e.g. means and variances) of spike trains is challenging in itself. To highlight the main issue, we take a simple example with two spike trains $A$ and $B$. Analogous to the classical descriptive statistics of real-valued variables, the Karcher mean of $A$ and $B$ can be defined as follows (Karcher, 1977):

$$C^* = \arg\min_C [d(A, C)^2 + d(B, C)^2] \tag{1}$$

where $d(\cdot, \cdot)$ is a distance between spike trains and the minimization is on the space of spike trains. The issue is to choose a distance measure that provides intuitive and tractable solutions to the minimization problem. Not all metrics lead to interesting solutions. For example, the commonly-used Victor-Purpura metrics (Victor and Purpura, 1996, 1997) would result in infinite number of solutions to this minimization. This is because these metrics are like Manhattan distances and lack the convexity needed to find the unique minimum (e.g. the Karcher mean of two points (0, 0) and (1, 1) under the Manhattan distance is any point $(x, y)$ that satisfies $0 \leq x \leq 1, 0 \leq y \leq 1$, and $x + y = 1$, whereas the mean under the classical Euclidean distance has a unique solution at (0.5, 0.5)). This example underscores the fact that nature of metric is important in defining statistical quantities. Many past metrics have been successful in some specific scenarios (Lim and Capranica, 1994; van Rossum, 2001; Aronov, 2003; Houghton and Sen, 2008; Schreiber et al., 2003; Kreuz et al., 2007; Quiroga et al., 2002; Hunter and Milton, 2003; Paiva et al., 2009a,b), mostly in quantifying dissimilarities between spike trains. However, their relevance in deriving statistical summaries remain unclear.

This implies that new distance metrics, especially the ones with Euclidean properties, are needed for computing means. In this paper, we develop a comprehensive framework involving a parametrized family of metrics that: (1) incorporates the commonly-used metrics such as the Victor-Purpura metrics and van Rossum metric (van Rossum, 2001), and (2) allows us to compute intuitive solutions for means of spike trains. Depending on the values of the parameter, the new metrics can have properties similar to the Manhattan distance or the Euclidean distance. We will show that the resulting mean spike train can reasonably summarize the firing pattern of a set of spike trains and can be used to provide an efficient spike train classification.

## 2. A Family of Metrics

Let $s(t)$ be a spike train with spike times $0 < t_1 < t_2 < \cdots < t_M < T$, where $[0, T]$ denotes the recording time domain. Then, $s(t)$ can be written in the following form,

$$s(t) = \sum_{i=1}^{M} \delta(t - t_i), \quad t \in [0, T] \,,$$

where $\delta(\cdot)$ is the Dirac delta function. We smooth the spike trains using a bump kernel function

$$K_\sigma(t) = \begin{cases} e^{-1/(1-(t/\sigma))^2}/Z & \text{if } |t| < \sigma \\ 0 & \text{otherwise} \end{cases}$$

which is infinitely differentiable with support $(-\sigma, \sigma)$, $\sigma > 0$ and $Z$ is a normalizing constant such that $\int K_\sigma(t)dt = 1$.

Therefore, the smoothed spike train is

$$f(t) = s(t) * K_\sigma(t) = \int s(\tau) K_\sigma(t - \tau) d\tau = \sum_{i=1}^{M} K_\sigma(t - t_i).$$

For a fixed $M$, we define the space of all smoothed spike-trains containing $M$ spikes to be $\mathcal{S}_M(\sigma)$, and the set of all smoothed spike-trains to be $\mathcal{S}(\sigma) = \cup_{M=0}^{\infty} \mathcal{S}_M(\sigma)$.

Define a family of metric distances in the smoothed domain as follows: for $f, g \in \mathcal{S}(\sigma)$,

$$D_p[\sigma, \lambda](f, g) = \inf_{\gamma \in \Gamma} \left( \int_0^T [|f(t)^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p}|^p + \lambda |1 - \dot{\gamma}(t)^{1/p}|^p] dt \right)^{1/p}, \quad (2)$$

where $1 \leq p < \infty$ and $0 < \lambda < \infty$, and $\Gamma$ is the set of all time warping (or diffeomorphism) $\gamma$'s from $[0, T]$ to $[0, T]$ which satisfies $\gamma(0) = 0, \gamma(T) = T$, and $0 < \dot{\gamma}(t) < \infty$. The first term on the right side of Eqn. 2 is called the *matching* term since it measures the *goodness-of-match* between $f$ and $g$ in presence of warping. The second one is called the *penalty* term which penalizes the amount of warping. It can be shown that $D_p$ is a proper distance. That is, it satisfies non-negativity, symmetry, and triangle inequality. Indeed, $D_p$ shares a lot of similarities to the classical $L^p$ norm in functional analysis. The distance in Eqn. 2 can be obtained via efficient dynamic programming techniques (see Supplementary Material).

In particular, the metric distance $D_2$ (when $p = 2$) is similar to the $L^2$ norm and its distance can be viewed as a penalized "Euclidean distance". When there is no time warping (i.e. $\gamma(t) = t$), the distance reduces to the van Rossum metric (van Rossum, 2001) ($L^2$ distance between $\sqrt{f}$ and $\sqrt{g}$). By penalizing the amount of warping, $D_2$ is able to capture the temporal spacing between spikes and describe the dissimilarity as an (increasing) function of such spacing.

Similarly, the $D_1$ distance (when $p = 1$) corresponds to a penalized version of the "Manhattan distance". When $\sigma$ is sufficiently small, the proposed $D_1$ metric can also be characterized by the interval-spike intervals (ISIs) in the spike trains. For example, we can let $\sigma < 0.5ms$ as successive spikes should be at least $1ms$ apart in practical data. In this case, $D_1$ provides results equivalent to a Victor-Purpura metric $D^{interval}$. That is, inserting or deleting a spike will have cost (or distance) 1, and shifting an ISI with time length $\Delta t$ will have cost $\lambda \Delta t$. Indeed, an optimal mapping between two spike trains can be looked as a path from one to the other via insertion, deletion and shifting, and vice versa. Therefore, when $\sigma < 0.5ms$, $D_1[\sigma, \lambda] = D^{interval}[\lambda]$.

## 3. Mean Computation

We define the Karcher mean of a set of spike trains $S_1, \cdots, S_N \in \mathcal{S}(\sigma)$ in $[0, T]$ using the "Euclidean distance" $D_2$ as follows:

$$S^* = \arg\min_{S \in \mathcal{S}(\sigma)} \sum_{k=1}^{N} D_2(S_k, S)^2 \tag{3}$$

Our investigation here is based on the assumption that both $\sigma$ and $\lambda$ are sufficiently small. When $\sigma < 0.5ms$, the spike trains can be equivalently represented by the associated ISIs. We automatically insert two spikes at times 0 and $T$ for each train. Hence, a spike train with $M$ spikes results in an ISI vector with $M + 1$ components. When $\lambda < 1/(2T)$, the penalty on time warping is relatively minor (the distance is dominated by the matching term), and the ISIs can be freely expanded or compressed (see Supplementary Materials for the selection of the upper bound $1/(2T)$).

### 3.1. Same number of spikes in each train

At first, we consider a limited scenario where the number of spikes is same in each $S_k$. This case is simpler since for spike trains $S_1, \cdots, S_N \in \mathcal{S}_M(\sigma)$, their mean will also be in $\mathcal{S}_M(\sigma)$. For $\sigma < 0.5ms$, each train $S_k$ can be equivalently represented by an ISI vector $(s_{k,1}, \cdots, s_{k,M+1})$ with $\sum_{i=1}^{M+1} s_{k,i} = T - Ml_\sigma$, where $l_\sigma$ denotes the width of each spike. The mean spike train $S$ can also be denoted using an ISI vector $(c_1, \cdots, c_{M+1})$. In this notation, Eqn. 3 is a constrained optimization problem whose solution takes the following unique, closed form:

$$c_j^* = \frac{(\sum_{k=1}^{N} \sqrt{s_{k,j}})^2}{\sum_{j=1}^{M+1} (\sum_{k=1}^{N} \sqrt{s_{k,j}})^2} (T - Ml_\sigma), \quad j = 1, \cdots, M + 1. \tag{4}$$

### 3.2. Arbitrary number of spikes in each train

Now we consider the mean for a set of spike trains $S_1, \cdots, S_N \in \mathcal{S}(\sigma)$ where the numbers of spikes are $n_1, \cdots, n_N$, respectively. There are two unknowns in the mean: 1) the number of spikes $n$ in $[0, T]$; and 2) the placements of these $n$ spikes in $[0, T]$. Based on the assumption that $\lambda$ and $\sigma$ are sufficiently small (i.e. $\sigma < 0.5ms$ and $\lambda < 1/(2T)$), one can show that for $k = 1, \cdots, N$,

$$D_2(S_k, S)^2 = |n_k - n| + \lambda \cdot (\text{optimal ISI warping distance}). \tag{5}$$

Here the distance is dominated by the first term on the right side of Eqn. 5. Therefore, the optimal $n$ is the one that can minimize $\sum_{k=1}^{N} |n_k - n|$, which is actually the median of $(n_1, \cdots, n_N)$.

Once $n$ is identified, we can then focus on the placements of these spikes. We propose an approach, termed Matching-Minimization (MM) algorithm, to solve for these placements in the following recursive fashion:

1. Set initial times for the $n$ spikes in [0, T] to form $S$.
2. **Matching Step:** Find the optimal time matching between $S$ and $S_k$, $k = 1, \cdots, N$. There are two cases:

   Case 1: $n_k \geq n$,

   (a) Identify $n$ spikes in $S_k$ that are optimally matched to the $n$ spikes in $S$.
   (b) Compute the $n + 1$ ISIs in $S_k$ using these $n$ spikes.

   Case 2: $n_k < n$,

   (a) Identify the $n_k$ spikes in $S$ that are optimally matched to the $n_k$ spikes in $S_k$.
   (b) Based on the $n_k$ matched pairs of spikes (between $S_k$ and $S$), compute the spike time in $S_k$ as a function of spike time in $S$ via linear interpolation. Insert $n - n_k$ (imaginary) spikes in $S_k$ using the interpolated values computed on the other $n - n_k$ spike times in $S$.
   (c) Compute the $n+1$ ISIs in $S_k$ using the $n$ spikes ($n_k$ real and $n-n_k$ imaginary).
3. **Minimization Step:** Compute the optimal $(n + 1)$ ISIs in $S$ based on Eqn. 4. These ISIs minimize the sum of squared distance for the given matching. Then update the mean spike train with these ISIs.
4. If spike times in the mean stabilize over steps 2 and 3, then the mean spike train is the current estimate and stop the process. Otherwise, go back to step 2.

Denote the estimated mean in the $i$th iteration as $S^{(i)}$. It can be shown that the sum of squared distance $\sum_{k=1}^{N} D_2^2(S_k, S^{(i)})$ monotonically decreases with each iteration. As 0 is a natural lower bound, the recursive procedure will always converge. In practical applications, we find the process only takes a few iterations to reach a reasonable convergence.

## 4. Experimental Illustrations

Here we demonstrate the use of the new metrics and the mean spike train in a real experimental recording in motor cortex (Chi et al., 2007). Spiking activity was recorded using a microelectrode array in the arm area of primary motor cortex in a Macaque monkey. The monkey was trained to perform a Squared-Path (SP) task by moving a cursor to targets via contralateral arm movements in the horizontal plane. To illustrate the new metrics in measuring distances between spike trains, we selectively choose spiking activity from 3 neurons over 10 movement trials in the recording. The first neuron fires more frequently when the hand moves to the left, the second one fires relatively evenly over the movement, and the third one fires more when the hand moves to the right. The trajectories of the 10 trials as well as the spike trains in the three neurons are shown in Fig. 1A-D.

[Figure 1 is about here]

With 30 spike trains over three neurons, there are $30 \times 29/2 = 435$ distances. As a summary, we found that the average distances within neurons are 12.2(±3.5) and 3.4(±0.5) for $D_1$ and $D_2$, respectively, and between neurons are 18.9(±5.3) and 4.5(±0.7), respectively. This indicates that the metrics appropriately address the dissimilarity between spike trains. To visually look at the clustering based on the pairwise distances, we use a multidimensional scaling technique (Seber, 2004) to show the 2-dimensional points associated with the two largest eigenvalues (see Fig. 1E, F). The plots clearly indicate that the spike trains can be well clustered via either the $D_1$ or $D_2$ metric. For example, using a leave-one-out cross-validation, the accuracy is 93.3% and 96.7% for $D_1$ and $D_2$, respectively.

Based on the proposed MM-algorithm, we compute the mean spike train for each neuron (Fig. 2A). Comparing with the original spike trains, the mean spike trains appropriately represent the firing patterns in the corresponding neurons. These mean spike trains can be used for classification problems. For that purpose, we randomly choose another 10-trial movement data from the same three neurons. For the 30 trains (10 trains from each neuron) in this test data, we label each spike train based on its distance to each of the three means. In this case, the computational cost is $3 \times 30 = 90$ distances. The result on this distance-to-mean classification is shown in Fig. 2B. We see that except spike trains 12 and 18, each spike train has the shortest distance to the mean from the same neuron (over the three neurons). That is, the spike trains can be classified with 93.3% accuracy.

As a comparison, we need to compute $30 \times 30 = 900$ $D_2$ distances in the classical pairwise-distance classification (label each spike based on the shortest averaged distance in each neuron). The result on this classification is shown in Fig. 2C. We see that the average distances are very similar to the distances to the mean in each neuron. Moreover, the accuracy is at 96.7%, a slight improvement over the distance-to-mean classification.

[Figure 2 is about here]

## 5. Discussion and Future Work

We have developed a parameterized family of metrics that takes into account different time warping of spike trains. These metrics successfully measure dissimilarity between observed spike trains. Furthermore, using the "Euclidean distance" $D_2$, we define the mean of a set of spike trains and propose an iterative numerical procedure, the MM-algorithm, for the mean computation. There are many uses of this statistic, two of which are demonstrated here: (1) the computation of a template or a representative spike train for a neuron under a fixed stimulus/behavior, and (2) the efficient classification of spike trains by comparing them to the mean templates for different neurons.

An important future direction is to extend this framework to include the second moment, i.e. the sample (scalar) variance for a set of spike train, which can be obtained by the minimum value obtained in Eqn. 3. Together, the first two moments can help us develop Gaussian-type models to capture variability in spike train data, and can lead to more useful statistical procedures such as confidence intervals and hypothesis tests.

**Acknowledgments**

This research is supported in part by the grants NSF IIS-0916154 to WW and NSF DMS-0915003, ONR N00014-09-1-0664, and AFOSR FA9550-06-1-0324 to AS. We thank Prof. Nicholas Hatsopoulos for providing experimental data. We also thank Prof. Eric Klassen for helpful discussions.

D. Aronov. Fast algorithm for the metric-space analysis of simultaneous responses of multiple single neurons. *Journal of Neuroscience Methods*, 124:175–179, 2003.

G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York: Wiley, 1978.

E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank. The time-rescaling theorem and its applicationto neural spike train data analysis. *Neural Computation*, 14:325–346, 2002.

Z. Chi, W. Wu, Z. Haga, N. Hatsopoulos, and D. Margoliash. Template-based spike pattern identification with linear convolution and dynamic time warping. *Journal of Neurophysiology*, 97:1221–1235, 2007.

C. Houghton and K. Sen. A new multineuron spike train metric. *Neural Computation*, 20:1495–1511, 2008.

J. D. Hunter and J. G. Milton. Amplitude and frequency dependence of spike timing: Implications for dynamic regulation. *Journal of Neurophysiology*, 90:387–394, 2003.

H. Karcher. Riemann center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

R. E. Kass and V. Ventura. A spike-train probability model. *Neural Computation*, 13: 1713–1720, 2001.

R. E. Kass, V. Ventura, and E. N. Brown. Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94:8–25, 2005.

T. Kreuz, J. S. Haas, A. Morelli, H.D.I. Abarbanel, and A. Politi. Measuring spike train synchrony. *Journal of Neuroscience Methods*, 165:151–161, 2007.

D. Lim and R. R. Capranica. Measurement of temporal regularity of spike train responses in auditory nerve fibers of the green treefrog. *Journal of Neurosceince Methods*, 52:203–213, 1994.

A. R. C. Paiva, I. Park, and J. C. Principe. A comparison of binless spike train measures. *Neural Computing and Applications*, 2009a. DOI 10.1007/s00521-009-0307-6.

A. R. C. Paiva, I. Park, and J. C. Principe. A reproducing kernel hilbert space framework for spike train signal processing. *Neural Computation*, 21:424–449, 2009b.

R. Quian Quiroga, T. Kreuz, and P. Grassberger. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Physical Review E*, 66, 041904, 2002.

F. Rieke, D. Warland, R. R. Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code.* MIT Press, 1997.

S. Schreiber, J.M. Fellousb, D. Whitmerc, P. Tiesingaa, and T.J. Sejnowskib. A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54:925–931, 2003.

G. A. F. Seber. *Multivariate Observations*. Wiley, 2004.

J. W. Tukey. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.

M. C. W. van Rossum. A novel spike distance. *Neural Computation*, 13:751–763, 2001.

J. D. Victor and K. P. Purpura. Nature and precision of temporal coding in visual cortex:a metric-space analysis. *Journal of Neurophysiology*, 76:1310–1326, 1996.

J. D. Victor and K. P. Purpura. Metric-space analysis of spike trains: theory, algorithms and application. *Network*, 8:127–164, 1997.

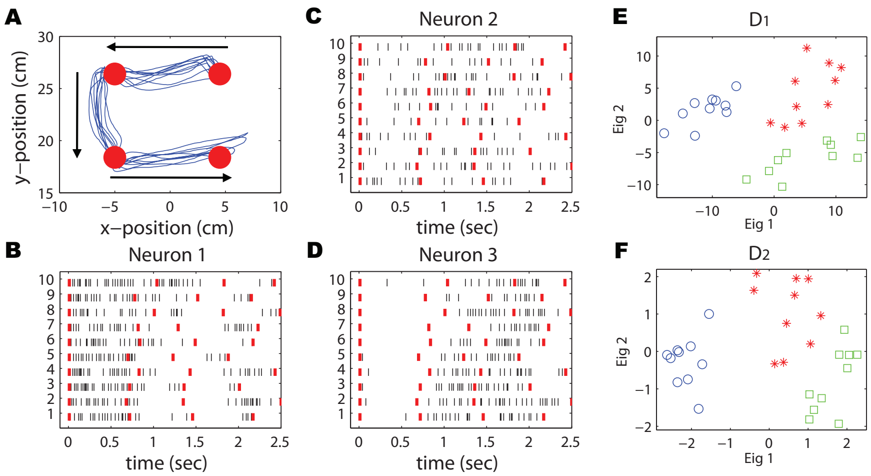


Figure 1: **A.** Trajectories (red lines) of hand movement in 10 successful trials of the SP task. In each of the trials, the monkey's hand passed the four corners of a square in counterclockwise order to reach a target (blue dot) that jumped from one corner to another, starting from the upper right corner. **B.** A raster plot of spike trains in the first neuron over the 10 movement trials. Each thin vertical line indicates the time of a spike, and one row is for one trial. The four thick, red vertical lines in each row indicate the times when the four corners were reached, respectively. **C. D.** Same as **B** except for the second and third neuron, respectively. **E.** The 2-dimensional points associated with the two largest eigenvalues based on metric $D_1$. Blue circles indicate spike trains from neuron 1, red stars are for neuron 2, and green squares are for neuron 3. **F.** Same as **E** except for metric $D_2$.

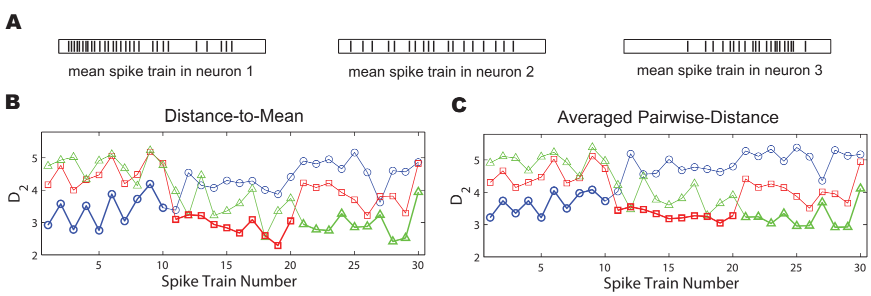


Figure 2: **A.** The means of the 10 trains in the 3 neurons. Each thin vertical line indicates the time of a spike. **B.** The distances from 30 test spike trains to the 3 means. Spike trains 1 to 10 are from neuron 1, 11 to 20 are from neuron 2, and 21 to 30 are from neuron 3. The blue line with circles shows the distances to the mean in neuron 1, the red line with squares to the mean in neuron 2, and the green line with triangles to the mean in neuron 3. **C.** Same as **B** except for the average distances from each of 30 test spike trains to the 10 trains of each neuron in the training data.