Estimation of a Mean Template from Spike-Train Data*

Wei Wu and Anuj Srivastava

***Abstract*— Computing the template, or the *mean*, of a set of spike trains is a novel and important task in neural coding. Due to the random nature of spike trains taken from experimental recordings, probabilistic and statistical methods have gained prominence in examining underlying firing patterns. However, these methods focus on modeling neural activity at each given time and therefore their results depend heavily on model assumptions. Taking a model-free and metric-based approach, we analyze the space of spike trains directly and reach algorithms for estimating statistical summaries, such as the *mean spike train*, of a given set. In our data-driven approach the mean is defined directly in a function space in which the spike trains are viewed as individual points. Here we develop an efficient and convergence-proven algorithm to compute the mean spike train in a general scenario. Experimental result from a neural recoding in primate motor cortex indicates that the estimated means successfully capture the typical patterns in spike trains. In addition, these mean spike trains provide an accurate and efficient performance in decoding motor behaviors.**

I. INTRODUCTION

Classical mathematical frameworks on spike trains have largely been based on parametric or semiparametric probability models, such as Poisson processes or other point processes [1], [5]. These methods are now widely used for reaching scientific conclusions from experimental results [2], [6]. However, they only focus on parametric representations at each given time and therefore can prove limited in data-driven problems in the space of spike trains directly. For example, if we look each spike train in a sample as one single point in an infinite dimensional spike train function space, in a non-parametric way, we may be interested in measuring the central tendency (or *mean*) of the sample and the associated variability. In this case, time-based stochastic models will not be suitable to use, and an overall measurement in the function space is desired for characterizing spike trains across the entire time domain.

In a recent paper [11], we have proposed a principled, data-driven framework to address this issue where we propose metric-based summary statistics in the function space occupied by spike trains. We introduced a parametrized family of metrics that takes into account different warpings in the time domain. These new metrics are essentially penalized L^p norms, involving appropriate functions of spike trains, with penalties associated with time-warping. The notion of means of spike trains was then defined based on the new metrics when $p = 2$ (corresponding to the "Euclidean distance"). This "Euclidean" property turns out to be essentially important to compute the templates, or means, in the spike train space. Once the templates are obtained, we have a more ambitious goal to develop various classical statistical inferences such as hypothesis tests, confidence intervals, functional PCA, and regressions, in the function space such as those presented in [7]. This new set of tools are expected to provide an alternative pathway to the classical methods for spike train analysis such as firing rate models and temporal models [8], [6].

However, the current result on estimating mean spike train is based on a strong assumption that the penalty coefficient λ (see Eqn. 1) is sufficiently small. This assumption limits the applicability of the estimate as certain amount of constraint on time warping is often needed to balance the matching and degree of warping. In this paper, we extend our investigation by proposing an efficient algorithm, call Matching-Centering-Pruning (MCP) algorithm, to compute the mean spike train of a set of spike trains with arbitrary number spikes for any penalty coefficient $\lambda > 0$. We demonstrate that the MCP algorithm converges over iteration. We then perform a distance-based classification using the estimated mean in an experimental data from primate motor cortex and obtain desirable result.

II. THEORY AND METHODS

A. Metrics between Spike Trains

In [11], we have defined a new family of metrics between two spike trains. The metrics correspond to the classical L^p norm for $1 \leq p < \infty$. In particular, when $p = 2$, the metric corresponds to the standard L^2 Euclidean distance. Here we briefly review the definition and algorithm for the metrics.

1) Definition Review: Assume $S(t)$ is a spike train with spike times $0 < t_1 < t_2 < \cdots < t_M < T$, where $[0,T]$ denotes the recording time domain. That is,

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

where $\delta(\cdot)$ is the Dirac delta function. We define the space of all spike-trains containing M spikes to be $\mathscr{S}_M$ and the set of all spike-trains to be $\mathscr{S} = \cup_{M=0}^{\infty} \mathscr{S}_M$. To simplify the notation, for any spike train $S \in \mathscr{S}$, we use $[S]$ to denote the set of all spike times in S (e. g. if $S(t) = \sum \delta(t - t_i)$, then $[S] = \{t_i\}$). For any finite set U, we use $|U|$ to denote the cardinality (i.e. number of elements) of that set.

Let Γ be the set of all time warping functions on the domain $[0,T]$, in which $\gamma : [0,T] \to [0,T]$ is a time-warping if, in

*This work was supported in part by the grants NSF IIS-0916154 to WW and NSF DMS-0915003 and ONR N00014-09-1-0664 to AS. We thank Prof. Nicholas Hatsopoulos for providing experimental data.

W. Wu and A. Srivastava are both with the Department of Statistics, Florida State University, Tallahassee, FL 32306, USA. Emails: wwu@stat.fsu.edu, anuj@stat.fsu.edu

addition to being continuous and (piecewise) differentiable, it satisfies these three conditions:

$$\gamma(0) = 0, \quad \gamma(T) = T, \quad 0 < \dot{\gamma}(t) < \infty.$$

It is easy to verify that Γ is a *group* with the operation being the composition of functions.

Assume $f(t) = \sum_{i=1}^{M} \delta(t - t_i) \in \mathscr{S}_M$ and $g(t) = \sum_{j=1}^{N} \delta(t - s_j) \in \mathscr{S}_N$ are two spike trains in $[0,T]$. For the set of time-warping functions Γ, and constants $1 \leq p < \infty$ and $0 < \lambda < \infty$, we define a metric between f and g in the following form:

$$d_p[\lambda](f,g) = \inf_{\gamma \in \Gamma} \left(X([f],[g \circ \gamma]) + \lambda \int_0^T |1 - \dot{\gamma}(t)^{1/p}|^p dt \right)^{1/p}, \tag{1}$$

where $X(\cdot,\cdot)$ denotes the cardinality of the Exclusive OR (i.e. union minus intersection) of two sets. That is,

$$X([f],[g \circ \gamma]) = M + N - 2 \sum_{i=1}^{M} \sum_{j=1}^{N} \mathbf{1}_{\{\gamma(t_i) = s_j\}}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function.

The first term, $X([f],[g \circ \gamma])$, in Eqn. 1 is called *matching term* which measures the goodness-of-match between f and g in presence of warping. The second term, $\int_0^T |1 - \dot{\gamma}(t)^{1/p}|^p dt$, is called *penalty term* and it penalizes the amount of warping. The constant $\lambda (> 0)$ is the penalty coefficient. We emphasize that d_p is a proper metric; that is, it satisfies *non-negativity*, *symmetry*, and *the triangle-inequality*. Indeed, d_p shares a lot of similarities to the classical L^p norm in functional analysis. When $p = 2$, the distance d_p can be viewed as a penalized classical "Euclidean distance" between two spike trains. We note that the d_p metrics generalize the commonly-used Victor-Purpura metric [10] and van Rossum metric [9] when $p = 1$ and 2, respectively.

2) Optimal Time Warping between Two Spike Trains: To examine the optimal warping function for the infimum in Eqn. 1, we at first define the set of all piecewise linear warping functions from spike times $\{t_i\}$ (in f) to $\{s_j\}$ (in g). That is, for a desired γ, in addition to meeting the conditions for functions in Γ, it also salsifies:

1) It is continuous, increasing, and piecewise linear from $[0,T]$ to $[0,T]$;
2) The set of all interior nodes of the pieces (i.e. excluding 0 and T), $\{a_l\}$, is a subset of $\{t_i\}$;
3) The set of the mappings of all interior nodes, $\{\gamma(a_l)\}$, is a subset of $\{s_j\}$.

We denote this set by $\Gamma_{f,g}$, which is a (finite) subset of Γ. One can prove (detail is omitted due to space limitation) that there exists a warping function $\gamma^* \in \Gamma_{f,g}$, such that

$$d_p[\lambda](f,g) = \left(X([f],[g \circ \gamma^*]) + \lambda \int_0^T |1 - \dot{\gamma}^*(t)^{1/p}|^p dt \right)^{1/p}.$$

In [11], we proposed an efficient *dynamic programming (DP)* procedure to compute the $d_p[\lambda]$ distance between any two spike trains. This procedure is mathematically validated because the optimal warping function in Eqn. 1 has a piecewise linear form on spike times between two trains, which can be efficiently estimated by a DP process. Assume the numbers of spikes in f and g are M and N, respectively, the computational cost will be in the order of $O(MN)$.

B. Mean Spike Train in $\mathscr{S}$ under $\mathbf{d_2}$

For spike trains $S_1, S_2, \cdots, S_N \in \mathscr{S}$ where the corresponding number of spikes are $\{n_1, n_2, \cdots, n_N\}$ (arbitrary non-negative integers), we have defined their sample *mean* using the classical Karcher mean under the "Euclidean" d_2 distance [4] as follows:

$$S^* = \arg\min_{S \in \mathscr{S}} \sum_{i=1}^{N} d_2(S_i, S)^2. \tag{2}$$

When the mean spike train S^* is known, the associated (scalar) sample variance, σ^2, can be defined in the following form,

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} d_2(S_i, S^*)^2. \tag{3}$$

The computation of σ^2 is straightforward, and the main challenge is to compute S^* for arbitrary value of λ.

1) Estimation of Mean Spike Train: For the mean spike train S, there are two unknowns: the number of spikes n and the placements of these spikes in $[0,T]$. In [11], we showed that when $\lambda < 1/(2NT)$, the sum of squared distance is dominated by the sum on the matching terms. Therefore, we can at first identify n, which turns out to be the median of $\{n_1, n_2, \cdots, n_N\}$. Then we proposed a Matching-Minimization (MM) algorithm to find the placements of the n spikes. However, for general value of $\lambda > 0$, neither the matching term nor the penalty term will be dominant, and therefore we cannot identify the number of spikes in the mean before estimating their placements.

We here propose a general algorithm to estimate the mean spike train. We initialize the number of spikes, n, in the mean spike train to be the maximum of $\{n_1, n_2, \cdots, n_N\}$, and then reduce the value of n during the iteration. The iterative process is analogous to the MM-algorithm except that we add a *pruning step* to remove redundant spikes in the mean during the iteration. We refer to this new method as Matching-Centering-Pruning (MCP) algorithm, where the *centering step* is equivalent to the *minimization step* in the MM-algorithm. This MCP-algorithm is described as follows:

1) Let $n = \max\{n_1, n_2, \cdots, n_N\}$. (Randomly) set initial times for the n spikes in [0, T] to form an initial S.
2) **Matching Step:** Using the dynamic programming procedure to find the optimal time matching from S_i to S, $i = 1, \cdots, N$. That is,

$$\gamma_i = \arg\min_{\gamma \in \Gamma} (X([S_i \circ \gamma],[S]) + \lambda ||1 - \sqrt{\dot{\gamma}}||^2)$$

3) **Centering Step:**
   a) Compute the extrinsic mean of $\{\gamma_1, \cdots, \gamma_N\}$, denoted by $\bar{\gamma}$.
   b) Apply $\bar{\gamma}^{-1}$ on S and $S_i \circ \gamma_i$. That is, the mean spike train and aligned trains are updated as

$$S^{(*)} = S \circ \bar{\gamma}^{-1}, \; (S_i \circ \gamma_i)^{(*)} = S_i \circ \gamma_i \circ \bar{\gamma}^{-1}, \; i = 1, \cdots, N.$$

4) **Pruning Step:**
   a) For each spike $s_k \in S^{(*)}$, count the number of times s_k appear in $\{[(S_i \circ \gamma_i)^{(*)}]\}_{i=1}^{N}$. That is,
   $$h_k = |\{1 \leq i \leq N | s_k \in [(S_i \circ \gamma_i)^{(*)}]\}|.$$
   b) Remove the spikes from $[S^{(*)}]$ which appear at most $N/2$ times in $\{[(S_i \circ \gamma_i)^{(*)}]\}$. Denote the new set of spikes as $[\widetilde{S^{(*)}}]$. That is,
   $$[\widetilde{S^{(*)}}] = \{s_k \in [S^{(*)}] | h_k > N/2\}.$$
   c) Let $[\widetilde{S^{(*)}}^{-}]$ be $[\widetilde{S^{(*)}}]$ except that one spike with minimal number of appearance is removed (randomly take one if there are multiple spikes at the minimum). That is, let $\tilde{k} \in \arg\min_{1 \leq k \leq n} \{h_k | h_k > N/2\}$. Then
   $$[\widetilde{S^{(*)}}^{-}] = \{s_k \in [\widetilde{S^{(*)}}] | k \neq \tilde{k}\}.$$
   d) If $\sum_{i=1}^{N} d_2(S_i, \widetilde{S^{(*)}}^{-})^2 \leq \sum_{i=1}^{N} d_2(S_i, \widetilde{S^{(*)}})^2$,
   then update the mean $S = \widetilde{S^{(*)}}^{-}$, and the number of spikes $n = |S|$.
   Otherwise,
   update the mean $S = \widetilde{S^{(*)}}$, and the number of spikes $n = |S|$.
5) If the sum of squared distance stabilizes over steps 2 to 4, 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 jth iteration as $S^{(j)}$. One can show (detail is omitted due to space limitation) that the sum of squared distances (SSD), $\sum_{i=1}^{N} d_2(S_i, S^{(j)})^2$, decreases iteratively. As 0 is a natural lower bound, $\sum_{i=1}^{N} d_2(S_i, S^{(j)})^2$ will always converge when j gets large. In practical applications, we find the process only takes a few iterations to reach a reasonable convergence of the SSD and the mean spike train.

In general, when the penalty coefficient λ gets large, the optimal time warping between spike trains will choose to have few matchings between spikes to lower the warping cost. Some of the spikes in the mean will be removed during the iteration to minimize the SSD. In the extreme case when λ is sufficiently large, any time warping would be discouraged and the mean spike train will be an empty train (i.e. a train with no spikes). This result indicates that in order to get a meaningful estimate, λ should not take a very large value. We have proposed an empirical value range of λ to balance the contributions between matching term and penalty term [11]. In practical use, one may use a cross-validation technique to decide its optimal value.

III. Experimental Results

We have developed an efficient and convergent approach, the MCP-algorithm, to estimate the mean spike train of a set of spike trains for arbitrary penalty coefficient λ. In this section, we will apply this new tool to perform some statistical analysis on a real experimental recording. This recording was previously used and clearly described when we first developed the metrics d_p between spike trains [11]. Briefly, a microelectrode arrays was implanted in the arm area of primary motor cortex (MI) in a juvenile macaque monkey (Macaca mulatta). Signals were filtered, amplified and recorded digitally and single units were manually extracted. The subject was trained to perform a closed Squared-Path (SP) task by moving a cursor to targets via contralateral arm movements in the horizontal plane. The example movement in each path is shown in Fig. 1 (extracted from [11]). Each sequence of 5 targets defined a path, and there were four different paths in the SP task (depending on the starting point). In this experiment, we recorded 60 trials for each path, and the total number of trials was 240. The time length of each trial varies from 5 to 6 seconds. As the new metrics are defined on a fixed time interval, we normalize the kinematics and spiking activity in each trial to 5 seconds.

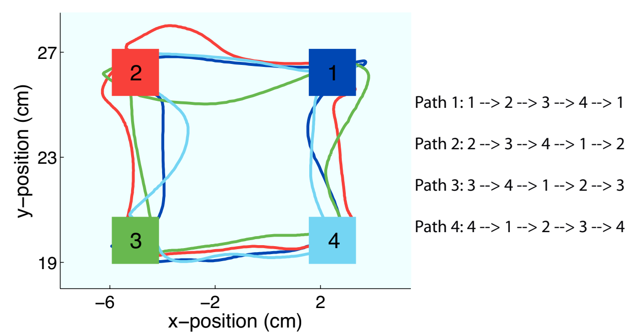

Fig. 1. Four trajectories of hand movement in the SP task. The four colors (blue, red, green, and cyan) indicate the trajectories started at corners 1, 2, 3, and 4, respectively, where the corners are also shown in the correspond colors. The four paths are also described with text legend (on the right side).

Here we will apply the mean spike trains to perform decoding, or classification, on the movement behaviors (four different paths). We at first selectively choose spiking activity from one neuron which show significant tuning property over four different paths. For illustration, 10 spike trains from each movement path as well as the reaching times at the five corners are shown in Fig. 2A. Let E denote the average number of spikes in all four paths in the entire dataset, and we find $E = 32.5$. Let $\lambda_0 = (E+E)/(2T) = 6.5$ be the penalty coefficient that makes equal maximal contribution from matching term and penalty term (see [11]). Here we choose three different values of λ at $(0.1, 1, 10)\lambda_0$, or (0.65, 6.5, 65), to allow different degrees of time warpings.

For the 60 trials in each path, we randomly choose 30 of them as the training data and the other 30 as the test data. For the 30 training spike trains, we apply the MCP-algorithm to compute their mean for $\lambda = 0.65, 6.5,$ and 65. The result is shown in Fig. 2B. We see that the number of spikes in the mean is a decreasing function with respect to the value of λ. Nevertheless, all these means appropriately represent the firing patterns in the respective paths. That is, more spikes in the means where the intensity is high in the original data, and few spikes where the intensity is low.

The classical pairwise-distance classification [10] has been

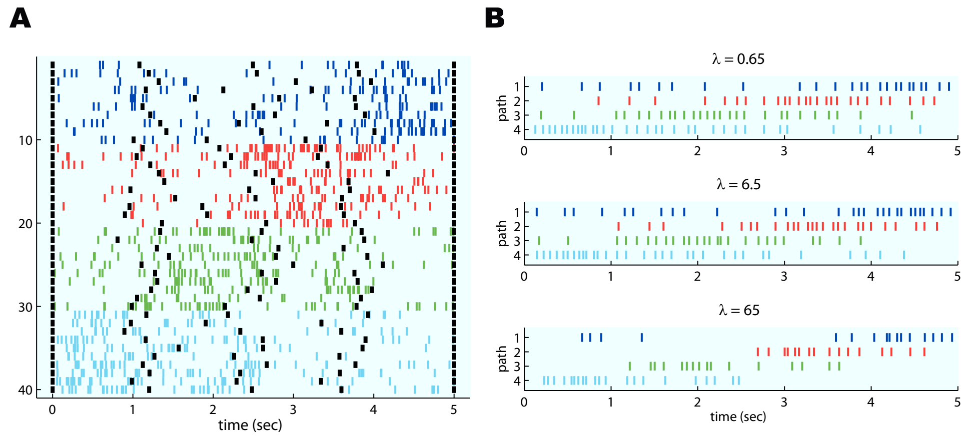

Fig. 2. **A.** 10 example spike trains generated from each path. The four colors (blue, red, green, and cyan) indicate spiking activity during hand movement in paths 1, 2, 3, and 4, respectively. Each thin vertical line indicates the time of a spike, and one row is for one trial. **B.** The mean spike trains of the four paths in the training data for $\lambda = 0.65, 6.5$, and 65, respectively. Each thin vertical line indicates the time of a spike, and one row is for one path. The colors are the same as that in **A**.

examined with the d_2 metric where we choose $\lambda = 10(E+E)/(2T) = 65$ (see [11]). Assuming there are N spike trains in the training and testing data, respectively ($N = 30 \times 4 = 120$ in the given SP task), we need to to compute N^2 pairs of distances (between training and testing trains). To simplify the computation, a mean-based method can be used where we classify each test train by the shortest distance to the four means in the training set. With means pre-estimated from training data by the MCP algorithm, the computation in classification will only have the order $O(N)$.

For the 120 test trains (30 trains from each path), we label each train by the shortest distance to the estimated four means. The result with all three values of λ is shown in Table I. In contrast to the pairwise classification, the mean-distance method has comparable accuracies. In particular, when $\lambda = 65$, this mean-distance classification reaches high accuracy at $87.5\%(105/120)$, which is about the same accuracy as the classification using pairwise-distances (90%), but significantly outperforms in terms of efficiency (linear order vs. quadratic order).

TABLE I

COMPARISON ON CLASSIFICATION PERFORMANCE.

Method	Comp Cost	$\lambda = 0.65$	$\lambda = 6.5$	$\lambda = 65$
Pairwise Distance	$O(N^2)$	30.0%	70.8%	90.0%
Mean-Distance	**O(N)**	**28.3%**	**68.3%**	**87.5%**

IV. Discussion

Based on the novel framework introduced in [11], the study in this paper significantly extends our investigation on statistical inferences in the spike train space: 1) We prove that the optimal time warping between any time spike trains exists and it has a continuous, piecewise-linear form. This result provides a solid foundation for the computational algorithm on the optimal warping function. 2) We propose an efficient, convergence-proven algorithm (the MCP-algorithm) to compute the mean spike train for any penalty coefficient $\lambda > 0$. This algorithm generalizes the MM-algorithm, which is based on a strong assumption that λ is sufficiently small.

Our long-term goal in this study is to build a new set of statistical inference methods such as hypothesis tests, confidence intervals, PCA, and regression in the spike train function space. We emphasize that the new framework is data-driven and we focus on building a mathematical representation for the neural spike train space and generating a set of inference tools in the space. This is different from many computational investigations on neural coding which are often motivated by real neurophysiological mechanisms [3]. The new set of tools in this project are expected to provide an alternative pathway for spike train analysis to the classical methods such as firing rate models and temporal models [2], [6].

References

[1] 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.

[2] 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.

[3] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2001.

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

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

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

[7] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis, Second Edition*. Springer Series in Statistics, 2005.

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

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

[10] 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.

[11] W. Wu and A. Srivastava. An information-geometric framework for statistical inferences in the neural spike train space. *Journal of Computational Neuroscience*, 31:725–748, 2011.