

Estimating summary statistics in the spike-train space

Wei Wu · Anuj Srivastava

Received: 6 February 2012 / Revised: 19 September 2012 / Accepted: 20 September 2012
© Springer Science+Business Media New York 2012

Abstract Estimating sample averages and sample variability is important in analyzing neural spike trains data in computational neuroscience. Current approaches have focused on advancing the use of parametric or semiparametric probability models of the underlying stochastic process, where the probabilistic distribution is characterized at each time point with basic statistics such as mean and variance. To directly capture and analyze the average and variability in the observation space of the spike trains, we focus on a data-driven approach where statistics are defined and computed in a function space in which the spike trains are viewed as individual points. Based on the definition of a “Euclidean” metric, a recent paper introduced the notion of the mean of a set of spike trains and developed an efficient algorithm to compute it under some restrictive conditions. Here we extend this study by: (1) developing a novel algorithm for mean computation that is quite general, and (2) introducing a notion of covariance of a set of spike trains. Specifically, we estimate the covariance matrix using the geometry of the warping functions that map the mean spike train to each of the spike trains in the dataset. Results from simulations as well as a neural recording in primate motor cortex indicate that the proposed mean and covariance successfully capture

the observed variability in spike trains. In addition, a “Gaussian-type” probability model (defined using the estimated mean and covariance) reasonably characterizes the distribution of the spike trains and achieves a desirable performance in the classification of the spike trains.

Keywords Estimation · Summary statistics · Spike-train space · Data-driven · Time warping · Euclidean metric

1 Introduction

Statistical modeling and inference in spike train data has been one of the central problems in computational neuroscience (Rieke et al. 1997; Dayan and Abbott 2001). Classical mathematical frameworks have largely been based on parametric or semiparametric probability models, such as Poisson processes or other point processes (Tukey 1977; Box et al. 1978; Kass and Ventura 2001). These methods are now widely used for reaching scientific conclusions from experimental results (Brown et al. 2002; Kass et al. 2005). However, these methods only focus on parametric representations at each given time and therefore can prove limited in situations where model assumptions are not met. An alternative is to use a data-driven approach where one views each observed spike train as a single point in an infinite-dimensional function space, in a non-parametric way, and generate summary statistics using the geometry of that space. Specifically, we may be interested in measuring the central tendency (or *mean*) of the sample and the associated variability around the mean. The computations of these summaries then rely

Action Editor: Mark van Rossum

W. Wu (✉) · A. Srivastava
Department of Statistics, Florida State University,
Tallahassee, FL 32306, USA
e-mail: wwu@stat.fsu.edu

A. Srivastava
e-mail: anuj@stat.fsu.edu

solely on the choice of metric used in the spike-train space, and avoid any model assumptions.

In a recent paper (Wu and Srivastava 2011), 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 have introduced a parametrized family of metrics (see Eq. (1)) that explicitly takes into account the time warpings needed to align spikes. These new metrics are essentially penalized L^p norms, involving appropriate functions of spike trains, with penalties associated with time warping for matching. 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 very important to compute the summary statistics (e.g. mean and variance) in the spike train space. We emphasize that the mean spike train, in principle, can be defined for any of the previous spike train distances. For example, the metrics for discrete state, discrete time models (Lim and Capranica 1994; MacLeod et al. 1998; Rieke et al. 1997), the metrics for discrete state, continuous time models (Victor and Purpura 1996, 1997; Aronov et al. 2003; Aronov 2003; Aronov and Victor 2004; Victor et al. 2007; Dubbs et al. 2010), the metrics for continuous state, continuous time models (van Rossum 2001; Houghton and Sen 2008; Houghton 2009), and a number of others (Schreiber et al. 2003; Kreuz et al. 2007; Quiroga et al. 2002; Hunter and Milton 2003; Paiva et al. 2009a, b). However, these authors have seldom sought summary statistics or ensuing statistical models under these metrics. In fact, these distances may not always provide interesting intuition or tractable solutions for establishing summaries of spike trains. Since our distances can be viewed as penalized L^p metrics, albeit not on the smoothed spike trains but some functions derived from them, the average value corresponds to more conventional statistics, such as the Euclidean mean. Based on the closed-form results for optimal time warping (Theorems 1 and 2 in Wu and Srivastava 2011), we are now able to derive an efficient and theoretically-proven convergent algorithm for computing mean of a set of spike trains with arbitrary numbers of spikes.

Our investigation is motivated in part by the recent progress in computational information geometry and shape analysis (Kass and Vos 1997; Klassen et al. 2004; Michor and Mumford 2007; Younes et al. 2008). The new metric framework that we proposed incorporates time warping between spike trains that can be used to estimate summary statistics in the spike train function space. Once the summary statistics 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 Ramsay and Silverman (2005) and Valderrama (2007). 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 (Rieke et al. 1997; Kass et al. 2005). However, this paper only represents a preliminary stage in this research and solves two outstanding problems for summarizing observed spike trains: (1) The current estimation of mean spike train is based on a strong assumption that the penalty coefficient λ (see Eq. (1)) is sufficiently small. This assumption limits the applicability of the estimate in practice. (2) So far we have only investigated the mean spike train (first order-statistic) and a scalar measure of total variance that lacks the detailed co-variability over the time domain. In order to develop a “Gaussian-type” probability model on the spike train space, a second-order statistic, namely covariance, would be needed. Once a model is developed, one can simulate the spiking process as well as conduct statistical inferences on the data.

In this paper, we extend our investigation on statistical inference in the spike train space and make the following contributions: (1) Based on the chosen spike train metric (see Eq. (1)), we prove that there exists an optimal time warping between two spike trains and it has a continuous, piecewise-linear form. This proof validates the dynamic programming algorithm that is used for estimating the metric (Wu and Srivastava 2011); (2) We propose 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 (i.e. the sum of squared distances decreases in the process). This algorithm actually generalizes the Matching-Minimization algorithm in Wu and Srivastava (2011); (3) Based on the optimal time warpings from the mean spike train, we propose a framework to estimate covariance of the set of spike trains. This estimation adopts the standard theory of covariance estimation to the manifold of interest here; (4) We further propose a classical Gaussian model using the estimated mean and covariance, and perform likelihood-based classification on a simulation as well as a real recording from primate motor cortex.

The rest of this manuscript is organized as follows. Section 2 provides a detailed theory and methods for estimating mean spike train of a set of trains for any $\lambda > 0$. We also introduce a warping-based covariance estimation for the set of spike trains. Each mean and

covariance estimation is illustrated by one simulation example. Section 3 examines the performance of the summary statistics in a real experimental recording from motor cortex in a monkey subject. Section 4 discusses related studies and presents future work. The appendix provides the mathematical details of the existence of optimal time warping and proof of the convergence of the MCP algorithm for mean spike trains.

2 Theory and methods

2.1 Metrics between spike trains

In Wu and Srivastava (2011), we have defined a new family of metrics to compare any two spike trains. The metrics correspond to the classical L^p norm for $1 \leq p < \infty$, albeit with a slight change in representations. 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.

2.1.1 Definition review

Assume $S(t)$ is a spike train with spike times $0 < t_1 < t_2 < \dots < t_M < T$, 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 \mathcal{S}_M and the set of all spike-trains to be $\mathcal{S} = \cup_{M=0}^{\infty} \mathcal{S}_M$. To simplify the notation, for any spike train $S \in \mathcal{S}$, 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, where a time warping is defined as an orientation-preserving diffeomorphism of the domain $[0, T]$. That is, $\gamma : [0, T] \rightarrow [0, T]$ is a time-warping if, in addition to being continuous and (piecewise) differentiable, it satisfies these three conditions:

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

It is easy to verify that Γ is a *group* with the operation being the composition of functions. Assume $f(t) = \sum_{i=1}^M \delta(t - t_i) \in \mathcal{S}_M$ and $g(t) = \sum_{j=1}^N \delta(t - s_j) \in \mathcal{S}_N$ are two spike trains in $[0, T]$. For the set of time-warping

functions Γ , and scalar constants $1 \leq p < \infty$ and $0 < \lambda < \infty$, we define a metric between f and g in the following form:

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

where $X(\cdot, \cdot)$ denotes the cardinality of the Exclusive OR (i.e. union minus intersection) of two sets. That is, $X(\cdot, \cdot)$ measures the number of unmatched spike times between $[f]$ and $[g \circ \gamma]$ and can be computed as

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

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function. We emphasize that the d_p metrics generalize the commonly-used Victor-Purpura metric (Victor and Purpura 1996) when $p = 1$. In case the spike trains are smoothed with some kernel functions, the metric d_p has a corresponding general form, D_p , which measures the elastic distance between two smoothed spike train functions (Wu and Srivastava 2011). By measuring the separation between spike kernels in the time domain and penalizing the amount of warping, we have shown that D_2 metric (when $p = 2$) also generalizes the van Rossum metric (van Rossum 2001).

The integration in Eq. (1) includes two terms. The first term, $X([f], [g \circ \gamma])$, is called *matching term* which measures the goodness-of-match between f and g in presence of warping. The second term, $\int_0^T |1 - \dot{\gamma}(t)|^{1/p} dt$, 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 between $f(t)$ and $g(t)$ has the following form:

$$d_2[\lambda](f, g) = \inf_{\gamma \in \Gamma} \left(X([f], [g \circ \gamma]) + \lambda \left\| 1 - \sqrt{\dot{\gamma}} \right\|^2 \right)^{1/2}, \quad (2)$$

where $\|\cdot\|$ denote the standard L^2 norm of a function on $[0, T]$. Due to its similarity to the L^2 norm, this distance corresponds to the classical Euclidean distance in the function space.

2.1.2 Optimal time warping between two spike trains

Note that since Γ is an infinite set, there may not exist a warping function in Γ such that the infimum in Eq. (1) can be reached. However, to clearly understand the optimal warping from one spike train to the other, it is desired that such transformation exists and can be computed with some efficient procedure. Here we show that the optimal warping function *does exist* and it has a specific form that can be effectively estimated. This result indicates that the exact lower bound “inf” (infimum) in Eqs. (1) and (2) can actually be changed to the achievable “min” (minimum).

To demonstrate this, 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 satisfies the following three conditions:

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), $\{a_i\}$, is a subset of $\{t_i\}$;
3. The set of the mappings of all interior nodes, $\{\gamma(a_i)\}$, is a subset of $\{s_j\}$.

We denote this set by $\Gamma_{f,g}$, which is a subset of Γ . Note that as the number of the spikes in f and g are finite, $\Gamma_{f,g}$ is a finite set. In Appendix A (Theorem 1), we prove that there exists a warping function $\gamma^* \in \Gamma_{f,g}$, such that

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

In Wu and Srivastava (2011), 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 here because by Theorem 1 we only need to have a finite search (by DP) for optimal matching on spike times between two trains. Assume the numbers of spikes in f and g are M and N , respectively, the computational cost will be in the order of $O(MN)$. This efficiency is the same as that for the Victor–Purpura metric D^{interval} ; d_p can be computed for any value of $p \geq 1$ while $p = 1$ in D^{interval} .

2.2 Mean of a set of warping functions

In the remainder of this paper, we focus on the special case when $p = 2$. This d_2 distance corresponds to a

Euclidean distance and therefore can be used to define conventional summary statistics in the spike train space (Wu and Srivastava 2011). Our goal here is to estimate the sample mean and the associated covariance for a set of spike trains $S_1, \dots, S_N \in \mathcal{S}$. Towards that goal, we first investigate the problem of defining and computing the sample mean of a set of warping functions $\gamma_i \in \Gamma, i = 1, \dots, N$. In particular, we assume that γ_i is the optimal warping from S_i to a candidate mean spike train S under d_2 . That is,

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

Based on Theorem 1 in Appendix A, each optimal warping function γ_i must be piecewise linear from $[0, T]$ to $[0, T]$, and the nodes of the pieces must be at the spike times (and 0, T).

We denote the spike times in S be $\{s_1, \dots, s_A\}$, where $A = |[S]|$ is the number of spikes in S . Here we count time 0 and time T as spikes and denote them by s_0 and s_{A+1} , respectively. Then, there are $A + 1$ associated Inter-Spike Intervals (ISIs), denoted by $(\delta_1, \dots, \delta_{A+1})$. Therefore, $\sum_{k=1}^{A+1} \delta_k = T$ and $s_k = \sum_{j=1}^k \delta_j, k = 0, 1, \dots, A + 1$. We also denote the corresponding (in the optimal warping) intervals in S_i by $(\delta_{i,1}, \dots, \delta_{i,A+1})$. That is, $\delta_{i,k} = \gamma_i(s_k) - \gamma_i(s_{k-1}), k = 1, \dots, A + 1, i = 1, \dots, N$. Then γ_i can be denoted in the following form:

$$\gamma_i(t) = \frac{\delta_{i,k}}{\delta_k} (t - s_{k-1}) + \sum_{j=1}^{k-1} \delta_{i,j}, \text{ if } s_{k-1} \leq t \leq s_k. \quad (3)$$

One example warping function is illustrated in Fig. 1. One can easily verify that in addition to being continuous and piecewise linear, γ_i in Eq. (3) satisfies the

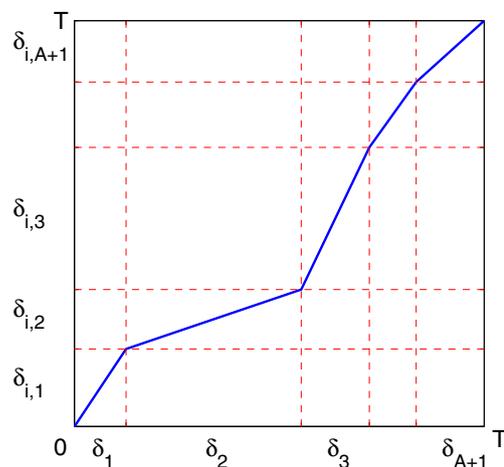


Fig. 1 Optimal warping function from S_i to S

following three conditions: $\gamma_i(0) = 0, \gamma_i(T) = T$, and the slope in each piece $\delta_{i,k}/\delta_k > 0$.

Before we define the mean of $\{\gamma_i\}$, we need to have a quantity to measure the distance between two warping functions. Note that $\{\gamma_i\}$ are constrained piecewise linear functions, and therefore their variability cannot be properly measured via conventional functional analysis in the L^2 space. Here we adopt a transformation method, called square-root velocity function (SRVF), to map the function to the hyper-sphere with radius \sqrt{T} , a subset in the L^2 space. The transformation is given as follows: for any warping function γ_i , define $\psi_i = \sqrt{\dot{\gamma}_i}$. This mapping is one-to-one because when ψ_i is known, we can compute γ_i as $\gamma_i(t) = \int_0^t \psi_i(t)^2 dt$. As γ_i is piecewise linear, ψ_i is piecewise constant. Using Eq. (3), ψ_i is given in the following form:

$$\psi_i(t) = \sqrt{\delta_{i,k}/\delta_k}, \text{ if } s_{k-1} < t < s_k. \tag{4}$$

The L^2 norm of ψ_i is:

$$\begin{aligned} \|\psi_i\| &= \left(\int_0^T \psi_i(t)^2 dt \right)^{1/2} = \left(\int_0^T \dot{\gamma}_i(t) dt \right)^{1/2} \\ &= (\gamma_i(T) - \gamma_i(0))^{1/2} = \sqrt{T}. \end{aligned}$$

Therefore, the SRVF representation simplifies the sophisticated geometry of warping functions to a hyper-sphere. Recall that the distance between any two points on the sphere can have two different, yet related forms. One is the intrinsic distance which is the shortest arc length of a great circle connecting them on the sphere. The other is the extrinsic distance under the L^2 metric. In Eq. (2), we have used the extrinsic distance between the constant 1 (SRVF of the identity mapping $\gamma_{id}(t) = t$) and the SRVF of a warping function. To be consistent, we define the (extrinsic) mean, $\bar{\gamma} \in \Gamma$, of $\{\gamma_i\}$ in the following form:

$$\bar{\gamma} = \arg \min_{\gamma \in \Gamma} \sum_{i=1}^N \left\| \sqrt{\dot{\gamma}} - \sqrt{\dot{\gamma}_i} \right\|^2 \tag{5}$$

Therefore, we have

$$\begin{aligned} \bar{\gamma} &= \arg \min_{\gamma \in \Gamma} \sum_{i=1}^N \int_0^T (\sqrt{\dot{\gamma}} - \sqrt{\dot{\gamma}_i})^2 dt \\ &= \arg \max_{\gamma \in \Gamma} \sum_{k=1}^{A+1} \int_{s_{k-1}}^{s_k} \sqrt{\dot{\gamma}} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}/\delta_k} \right) dt \end{aligned}$$

Using the Cauchy–Schwarz inequality,

$$\begin{aligned} &\sum_{k=1}^{A+1} \int_{s_{k-1}}^{s_k} \sqrt{\dot{\gamma}} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}/\delta_k} \right) dt \\ &\leq \sum_{k=1}^{A+1} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}} \right) \left(\int_{s_{k-1}}^{s_k} \dot{\gamma} dt \right)^{1/2}. \end{aligned}$$

The equality holds if and only if $\sqrt{\dot{\gamma}}$ is constant on $[s_{k-1}, s_k], k = 1, \dots, A + 1$. Note that γ has constraint $\int_0^T \dot{\gamma} dt = T$. Using the Cauchy inequality again,

$$\begin{aligned} &\sum_{k=1}^{A+1} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}} \right) \left(\int_{s_{k-1}}^{s_k} \dot{\gamma} dt \right)^{1/2} \\ &\leq \left(\sum_{k=1}^{A+1} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}} \right)^2 \right)^{1/2} (T)^{1/2}. \end{aligned}$$

The equality holds if and only if

$$\sqrt{\delta_k \dot{\gamma}(t)} = \frac{\sum_{i=1}^N \sqrt{\delta_{i,k}}}{\sqrt{\sum_{k=1}^{A+1} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}} \right)^2 / T}}, \text{ if } s_{k-1} < t < s_k.$$

In this case the function in Eq. (5) is minimized. Therefore, the mean time warping $\bar{\gamma}$ is also a piecewise linear function, and its SRVF has the following closed-form,

$$\dot{\bar{\gamma}}(t) = \frac{\left(\sum_{i=1}^N \sqrt{\delta_{i,k}/\delta_k} \right)^2}{\sum_{k=1}^{A+1} \left(\sum_{i=1}^N \sqrt{\delta_{i,k}} \right)^2} T, \text{ if } s_{k-1} < t < s_k. \tag{6}$$

2.3 Mean spike train in \mathcal{S} under d_2

Based on d_2 , we have introduced the notion of sample mean and sample variance of a set of spike trains. However, to utilize some efficient, analytical expressions, we have assumed that the penalty coefficient λ is sufficiently small (Wu and Srivastava 2011). This assumption may limit the application of the methods when certain penalty is expected for practical purposes. In this paper, we will generalize the computation of the mean spike train to any value of $\lambda \in (0, \infty)$.

For spike trains $S_1, S_2, \dots, S_N \in \mathcal{S}$ where the corresponding numbers of spikes are $\{n_1, n_2, \dots, n_N\}$ (arbitrary non-negative integers), we have defined their sample mean using the classical Karcher mean under the d_2 distance (Karcher 1977) as follows:

$$S^* = \arg \min_{S \in \mathcal{S}} \sum_{i=1}^N d_2[\lambda](S_i, S)^2. \tag{7}$$

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

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N d_2[\lambda](S_i, S^*)^2. \tag{8}$$

The computation of the variance is straightforward, and the main challenge is at computing the mean spike train for arbitrary value of λ . Here we propose an efficient, convergent algorithm to address this problem.

2.3.1 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 Wu and Srivastava (2011), 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 $\{n_1, n_2, \dots, n_N\}$. Then we proposed a Matching-Minimization (MM) algorithm which is a numerical procedure to find the placements of the n spikes by iteratively decreasing the sum on the penalty terms for warping.

For general value of $\lambda > 0$, neither the matching term nor the penalty term will be universally dominant, and therefore we cannot identify the number of spikes in the mean before estimating their placements. A key challenge is that we need to update the number of spike trains in the mean during the iteration. In this paper, we propose a general algorithm to estimate the mean spike train. We initialize the number of spikes in the mean spike train equal to the maximum of

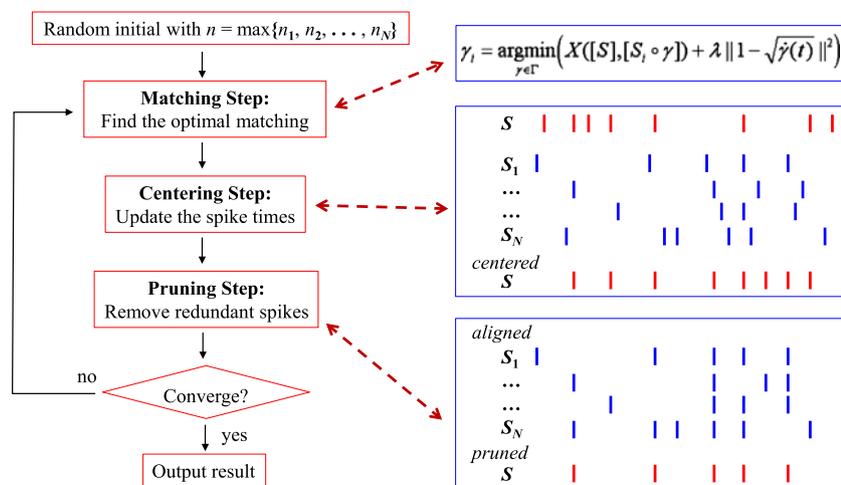
$\{n_1, n_2, \dots, n_N\}$, and then reduce this number during the iteration. The iterative process is analogous to that in the MM-algorithm except that we add a *pruning step* to remove redundant spikes in the mean during the iteration. We refer to this new method as Matching-Centering-Pruning (MCP) algorithm, where the *centering step* is equivalent to the *minimization step* in the MM-algorithm.

The basic idea of the MCP-algorithm is illustrated using a diagram in Fig. 2. We at first let n be the maximum of $\{n_1, n_2, \dots, n_N\}$ and randomly assign the placements of these n spikes in the mean. In the *Matching step* of each iteration, we compute the optimal time warping γ_i from each train in the set to the current mean. This is done using the dynamic programming algorithm. In the *Centering step*, we compute the mean of $\{\gamma_i\}$ (using the solution in Eq. (6)), and then apply its inverse to the mean spike train as well as the aligned spike trains. In this case, if we recompute the optimal time warping from each original spike train to the transformed, or centered, mean spike train, the mean of all warping functions will be identity. This is why we call this step as a *Centering step*. In the *Pruning step*, we remove all spikes in the mean train that match at most $N/2$ spikes with the aligned spike trains. We also test if any spike that matches more than $N/2$ spikes can be removed here. This is to avoid the MCP process being trapped in certain local minima. Finally, we check the convergence on the sum of squared distances. If it converges, we stop the iteration.

Here we formally state the MCP-algorithm in the following recursive form:

1. Let $n = \max\{n_1, n_2, \dots, n_N\}$. (Randomly) set initial times for the n spikes in $[0, T]$ to form an initial S .

Fig. 2 Illustration of the MCP algorithm



2. **Matching Step:** Using the dynamic programming procedure to find the optimal time matching from S_i to S , $i = 1, \dots, N$. That is,

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

3. **Centering Step:**

- (a) Compute the extrinsic mean of $\{\gamma_1, \dots, \gamma_N\}$ (using Eq. (6)), 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

$$S^{(*)} = S \circ \bar{\gamma}^{-1},$$

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

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,

$$h_k = \left| \left\{ 1 \leq i \leq N \mid s_k \in [(S_i \circ \gamma_i)^{(*)}] \right\} \right|.$$

- (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 $[S^{(*)}]$. That is,

$$[\widetilde{S}^{(*)}] = \{s_k \in [S^{(*)}] \mid h_k > N/2\}.$$

- (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 $\tilde{k} \in \arg \min_{1 \leq k \leq n} \{h_k \mid h_k > N/2\}$. Then

$$[\widetilde{S}^{(*)-}] = \{s_k \in [\widetilde{S}^{(*)}] \mid k \neq \tilde{k}\}.$$

- (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 distances stabilizes over steps 2–4, 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 j th iteration as $S^{(j)}$. One can show 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 (to a local minimum) when j gets large. The detail of the proof is shown in Appendix B. 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. Note that this MCP algorithm only takes linear computational order on the number of spike trains in the set and is a very efficient algorithm. However, there may be multiple local minima in Eq. (7), and more work will be needed to search for the global minimum in the future.

Note that in the Centering Step, we compute the mean warping of N warping functions, and then apply the inverse of the mean warping to update the mean spike train. This procedure is actually equivalent to the *Minimization Step* in the MM-algorithm. In particular, the update on the mean warping (Eq. (6)) is the same as the closed form formula for updating mean spike train in the MM-algorithm (Eq. (11) in Wu and Srivastava 2011). In practical use of the MCP-algorithm, we find that when λ is sufficiently small, the number of spikes in the iterative procedure will decrease until it reaches the median number in the set. This result is consistent to the estimation in the MM-algorithm where the number of spikes in the mean is fixed as the median number in the set before the estimation of the spike placements.

In general, when the penalty coefficient λ gets large, the optimal time warping 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 as that will result in larger distance than simply the Exclusive OR operation. In this case, the mean spike train in this case will be an empty train (i.e. a train with no spikes). This result indicates that in order to get a meaningful estimate of the mean spike train, the penalty coefficient 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 (Wu and Srivastava 2011). In practical use, one may use a cross-validation technique to decide the optimal value of λ .

2.3.2 Illustration of the mean estimation

To test the performance of the MCP-algorithm, we here illustrate the mean estimation using spike trains generated from a homogeneous Poisson process. Let total time $T = 1$ (s), Poisson rate $\rho = 10$ (spikes/s) and we randomly generate 30 spike trains from this process. The individual spike trains are shown in Fig. 3(a). The numbers of spikes in these trains vary from 5 to 16, with the median value of 10. Therefore, n , the number of spikes in the mean, is initialized to be 16. We let λ be three different values at 20, 400, and 1200, which varies the warping penalty from small to large. To properly compare the results with these three values of λ , we

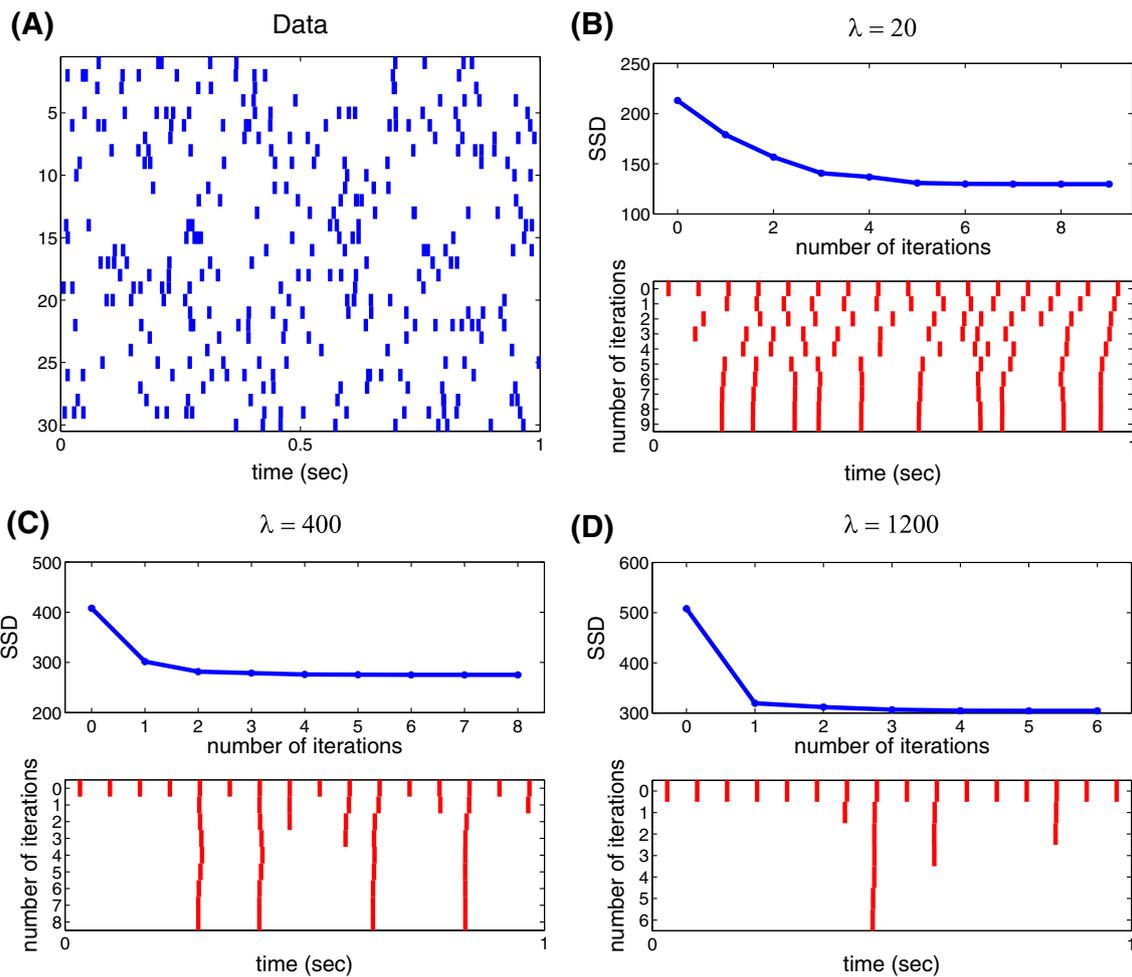


Fig. 3 (a) 30 spike trains generated from a homogeneous Poisson process. Each *blue vertical line* indicates a spike. (b) Estimation result when $\lambda = 20$. *Upper panel:* The sum of squared distances (SSD) over all iterations with the uniform initial. *Lower panel:* The estimated mean spike train over all iterations. The initial is the spike train on the top row (0th), and the final estimate

is the spike train on the bottom row (9th). We see that the SSD decreases over the iterations and the mean spike trains converges rapidly. We also see that the mean is approximately evenly distributed which reasonably captures the homogeneous nature of the process. (c, d) Estimation results when $\lambda = 400$ and 1200, respectively

adopt uniformly distributed 16 spikes in $[0, T]$ as the initial for the mean in each case.

The result by the MCP-algorithm for $\lambda = 20$ is shown in Fig. 3(b). The upper panel shows the evolution of the SSD in Eq. (7) versus the iteration index. We see that it only takes a few iterations for the SSD to decreasingly converge to a minimum. The mean spike train during the estimation process is shown in the lower panel. The first row (0th) is the initial, and the last row (9th) is the final estimate. Significant changes are observed in the first few (1–5) iterations, and then the process stabilizes. Note that the spikes in the mean train are approximately evenly distributed (although some variability is observed due to the finite random samples). Such distribution well captures the homogeneous nature of

the underlying process. This result is consistent to what we obtained using the MM-algorithm in Wu and Srivastava (2011). We also note that the number of spikes in this mean spike train is 10, which equals the median of the numbers of spikes in the set. This result indicates that when λ is small, the MCP-algorithm generates consistent estimate to the MM-algorithm.

The result for $\lambda = 400$ is shown in Fig. 3(c). With a larger penalty here, the optimal time warping between spike trains chooses to have fewer matchings between spikes to lower the warping cost. Some of the spikes in the mean are removed during the iteration to decrease the SSD. This explains why there are only four spikes in the estimated mean. However, we observe that the spikes are still approximately evenly distributed over

the time domain. In this case, the convergent SSD is about 280, which is greater than the SSD (at about 130) when $\lambda = 20$. The result for $\lambda = 1200$ is shown in Fig. 3(d). With such a large penalty, the convergent SSD is about 300. In this extreme case, there is only one spike left in the mean, and it remains near the center of the time domain. Note that when λ is even larger, we expect no spikes will appear in the mean.

2.4 Spike train covariance in \mathcal{S} under d_2

In this section, we extend the investigation on the first order statistic, mean, of a set of spike trains to the second order statistic, namely the covariance, that naturally depends on the mean statistic. Basically, given the mean spike train, we compute the variation of each train from the mean, and define and estimate the covariance using these variations (Srivastava and Jermyn 2009). The challenge in this process comes from the fact that there are two terms in each measure of variation: a matching term and a penalty term. The matching term is computed using Boolean operations on finite sets and it is very difficult to identify a time-based covariance structure with this term. Thus, in this paper, we limit the covariance description to only the penalty term where the co-variability in the set of spike trains is represented by the set of optimal time warping functions. In this sense, the definition and use of covariance is based only on a partial information here and denotes a limitation of this definition.

2.4.1 Definition of covariance in the tangent space

For a set of spike trains $S_1, \dots, S_N \in \mathcal{S}$, we assume their mean spike train $S \in \mathcal{S}$ is already obtained (e.g. using a MCP-algorithm). We denote the spike times in S by $\{s_1, \dots, s_A\}$, where $A = |[S]|$. We count time 0 and time T as spikes and denote them by s_0 and s_{A+1} , respectively. Then, there are $A + 1$ associated ISIs $\delta_1, \dots, \delta_{A+1}$. We also denote the corresponding (in the optimal warping) intervals in S_i by $(\delta_{i,1}, \dots, \delta_{i,A+1})$. Then the optimal warping (from S_i to S) γ_i has the piecewise linear form shown in Eq. (3). Focusing on the time warping between spike times, the dissimilarity between S and each S_i can be measured by the warping function γ_i . The two trains are similar when γ_i is near the identity warping function $\gamma_{id}(t) = t, t \in [0, T]$, and are dissimilar otherwise.

The next question is: How can we measure the variability in these warping functions? Note that the warping functions are constrained piecewise linear functions. To effectively measure their variability, we

adopt a *normalized version* of the SRVF transformation (see Section 2.2) to map the function to the unit hyper-sphere \mathbf{S}^∞ , a nonlinear manifold in the L^2 space. The transformation is given as follows: for each warping function γ_i , define $\psi_i = \sqrt{\dot{\gamma}_i/T}$. As γ_i is piecewise linear, ψ_i is piecewise constant. Using Eq. (3), ψ_i is given in the following form.

$$\psi_i(t) = \sqrt{\frac{\delta_{i,k}}{\delta_k T}}, \text{ if } s_{k-1} < t < s_k. \tag{9}$$

The L^2 norm of ψ_i is:

$$\begin{aligned} \|\psi_i\| &= \left(\int_0^T \psi_i(t)^2 dt \right)^{1/2} = \left(\int_0^T \dot{\gamma}_i(t)/T dt \right)^{1/2} \\ &= \left(\frac{\gamma_i(T) - \gamma_i(0)}{T} \right)^{1/2} = 1. \end{aligned}$$

Therefore, the normalized SRVF representation simplifies the geometry of warping functions to the unit sphere. Recall that the distance between any two points on the unit sphere, under the L^2 metric, is simply the length of the shortest arc of a great circle connecting them on the sphere. For the identity warping γ_{id} , its normalized SRVF is $\psi_{id} = \sqrt{\dot{\gamma}_{id}/T} = 1/\sqrt{T}, t \in [0, T]$.

\mathbf{S}^∞ being a sphere is a nonlinear manifold and, therefore, it is difficult to directly measure the variability of points lying in that space. We use a commonly-used approximation by estimating the variation in the tangent space (Srivastava and Jermyn 2009; Dryden et al. 2009; Pennec 2006) at the point ψ_{id} , denoted by \mathcal{T}_{id} , since that space is a vector space and one can define a covariance matrix in that space. Our choice of mapping these points to the tangent space at ψ_{id} is the most natural and intrinsic (Riemannian) way of studying statistics on spheres. Furthermore, this representation ensures that the measurement of variation under this approach agrees with the metric d_2 on the spike train space. That is, we project each ψ_i on \mathbf{S}^∞ to one point in \mathcal{T}_{id} by preserving the arc distance. In particular, the origin of this tangent space is the projection of ψ_{id} on \mathbf{S}^∞ . The projection of $\psi_i(t)$, denoted by $v_i(t)$, is given by Kurtek et al. (2011):

$$v_i(t) = \frac{\theta_i}{\sin(\theta_i)} (\psi_i(t) - \cos(\theta_i)\psi_{id}(t)),$$

where $\theta_i = \cos^{-1}(\int_0^T \psi_{id}(t)\psi_i(t)dt)$ is the joint angle between $\psi_{id}(t)$ and $\psi_i(t)$. One can easily verify that $\|v_i\| = \theta_i$. That is, the arc length from ψ_{id} to ψ_i is preserved as the L^2 distance from the origin to v_i on \mathcal{T}_{id} . This projection is illustrated in Fig. 4.

assumed normally distributed with mean 0 covariance $\tilde{\Sigma}$. The likelihood is given as

$$\begin{aligned}
 p(S^*) &= N(\tilde{v}^*; 0, \tilde{\Sigma}) \\
 &= \frac{1}{(2\pi)^{(A+1)/2} \det(\tilde{\Sigma})^{1/2}} \\
 &\quad \times \exp\left(-\frac{1}{2}(\tilde{v}^*)^T \tilde{\Sigma}^{-1}(\tilde{v}^*)\right). \tag{12}
 \end{aligned}$$

We consider a special case when the covariance matrix is simply the identity matrix I_{A+1} . That is, we ignore the covariance and only use mean spike train for the likelihood computation. In this case, we have

$$\begin{aligned}
 p(S^*) &= N(\tilde{v}^*; 0, I_{A+1}) \propto \exp\left(-\frac{1}{2}(\tilde{v}^*)^T \tilde{v}^*\right) \\
 &= \exp\left(-\frac{1}{2} \sum_{k=1}^{A+1} (v_k^*)^2\right).
 \end{aligned}$$

Using the expression $v_k^* = \frac{1}{\sqrt{T}} \frac{\theta^*}{\sin(\theta^*)} (\sqrt{\delta_k^*} - \cos(\theta^*)\sqrt{\delta_k})$, this probability can be computed as:

$$\begin{aligned}
 p(S^*) &\propto \exp\left(-\frac{1}{2} \left(\frac{\theta^*}{\sin(\theta^*)}\right)^2\right) \\
 &\quad \times \left(1 + (\cos(\theta^*))^2 - 2 \cos(\theta^*) \frac{1}{T} \sum_{k=1}^{A+1} \sqrt{\delta_k^* \delta_k}\right) \\
 &= \exp\left(-\frac{1}{2}(\theta^*)^2\right). \tag{13}
 \end{aligned}$$

This indicates that when only the mean is used, the likelihood of a spike train in the model can be equivalently estimated by the intrinsic distance between the train and the mean (therefore maximizing likelihood will be equivalent to minimizing intrinsic distance).

2.4.3 Illustration of the covariance and probability model

To illustrate the covariance in spike train data, we randomly generate spike trains following two different Poisson processes. The total time $T = 2$ (s), and the Poisson rate (in the units of spikes/s) varies over time with two different functions

$$\begin{aligned}
 \text{Class 1: } \rho_1(t) &= 16 \cdot \mathcal{N}(t; 0.8, 0.15), \\
 \text{Class 2: } \rho_2(t) &= 16 \cdot \mathcal{N}(t; 1.0, 0.3), \quad 0 \leq t \leq T
 \end{aligned}$$

where $\mathcal{N}(t; \mu, \sigma)$ indicates the normal probability density at t with mean μ and variance σ^2 . These rate functions are shown in the upper panel of Fig. 5. We see that the two density functions have different forms; The first one is concentrated at around 0.8 s, whereas the second one is relatively more dispersed with center

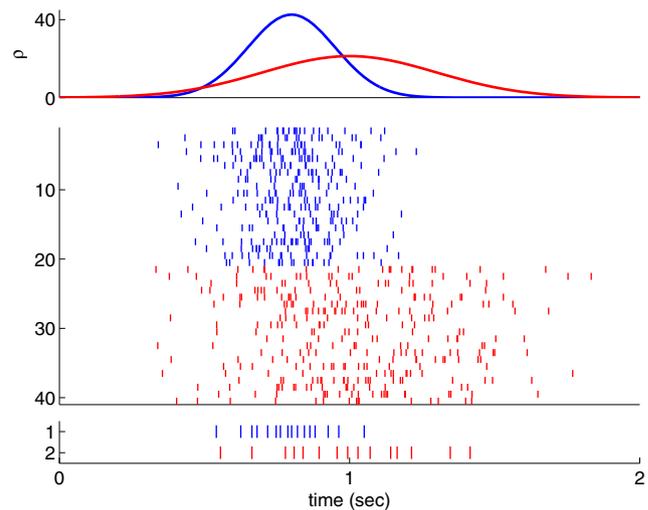


Fig. 5 *Upper panel:* Two different Poisson rate functions $\{\rho_i(t)\}$ on the time domain $[0, 2]$, colored by *blue* and *red*, respectively. *Middle panel:* 20 example spike trains generated from each class $\rho_i(t)$, $i = 1, 2$. Each *thin vertical line* indicates the time of a spike, and one row is for one trial. The colors are the same as that in the upper panel. *Lower panel:* The mean of the 30 spike trains in each class when $\lambda = 0.5$. The colors are the same as that in the upper panel

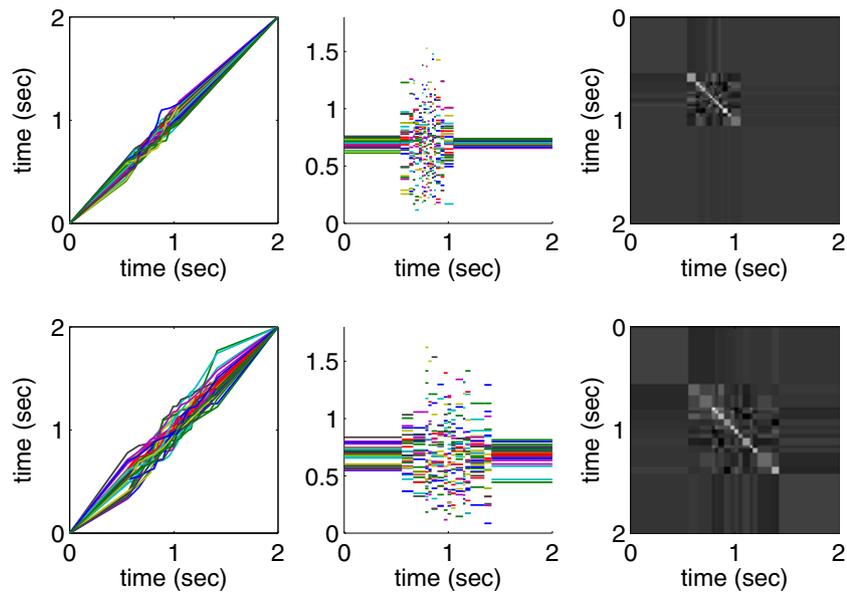
around 1 s. 30 spike trains are randomly generated with each density function. We denote these trains by S_{ij} , $i = 1, 2$, $j = 1, \dots, 30$. For illustration, 20 spike trains from each class are shown in the middle panel of Fig. 5. We know the spike train covariance is based on the mean spike train. To have a detailed characterization on the covariability, we take a small value of $\lambda = 1/T = 0.5$ and this penalty allows most time warpings on spike times (but not for very drastic ones). We compute the mean spike train in each class and the result is shown in the lower panel of Fig. 5. These mean spike trains appropriately represent the firing patterns in the corresponding paths. That is, the frequency of the spikes is approximately proportional to the value of the density function at each time.

We denote the two mean spike trains as S_1 and S_2 . Then for each class, we compute the optimal time warping from each train to their mean. That is,

$$\gamma_{i,j} = \arg \min_{\gamma \in \Gamma} \left(X([S_{ij} \circ \gamma], [S_i]) + \lambda \left\| 1 - \sqrt{\gamma} \right\|^2 \right)$$

These (piecewise linear) warping functions are shown in the first column of Fig. 6. We observe that the warping functions are more variable in Class 2 than in Class 1. This is a natural result as the spikes in Class 2 are more dispersed. We also show the (piecewise constant) SRVFs of these warping functions in the second column of Fig. 6. It is apparent that the variability in

Fig. 6 Row 1, column 1: The optimal warping functions from each train to their mean in Class 1. Row 1, column 2: The associated SRVFs with are piece-wise constant. Row 1, column 3: The gray-scale image of the covariance matrix. In the covariance image, *light colors* indicate higher variabilities and *dark* indicate lower. Row 2: Same as Row 1 except for Class 2



Class 1 is centered at around 0.8 s, and in Class 2 at around 1 s. These patterns appropriately follow the two Poisson density functions ρ_1 and ρ_2 . The covariance matrices are then computed using Eq. (10), and the result is shown in the third column of Fig. 6. These covariances not only describe the variability at each time point, but also measure the co-variability between any two time points.

Based on a simple Gaussian assumption, we can compute the log-likelihood of each spike train using Eq. (12) on the estimated mean and covariance. The result is shown in Fig. 7(a). It is found that all 30 spike trains in Class 1 have higher likelihood in the model of Class 1 than in the model of Class 2. Analogously, 29 out of 30 spike trains in Class 2 have higher likelihood in the model of Class 2 than in the model of Class 1. That is, if we perform a likelihood-based classification on the 60 spike trains, the accuracy will be 98 % (59/60). Note that this classification procedure is not cross-validated as the means and covariances are estimated

using the same 60 trains. To test the classification on an independent dataset, we randomly generate ten more spike trains from each class. The likelihood of these 20 test trains are shown in Fig. 7(b). It is found that 19 out of 20 test trains have higher likelihood in the same class than in the other. That is, the accuracy of a likelihood-based classification is 95 % (19/20), which is comparable to that in the training data. All these results indicate that a simple Gaussian model may be used to differentiate the distributions of two classes of spike trains.

2.4.4 Illustration of the covariance in classification

Spike train covariance describes the co-variability between ISIs, and this second-order statistic is expected to improve classification performance when the mean spike trains in different classes are the same. Here we provide a toy example to illustrate the use of covariance in classification. To simplify the problem, we assume

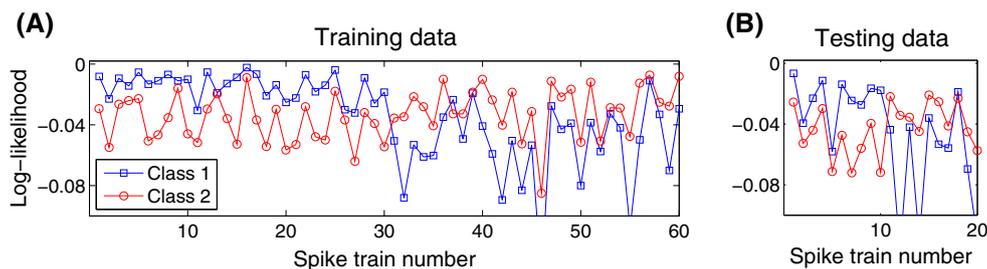


Fig. 7 (a) The log-likelihood of the 60 training spike trains in the model for Class 1 (blue line with squares) and Class 2 (red line with circles). Spike trains 1–30 are from Class 1, and spike trains

31–60 are from Class 2. (b) Same as (a) except for the 20 testing spike trains, where trains 1–10 are from Class 1, and trains 11–20 are from Class 2

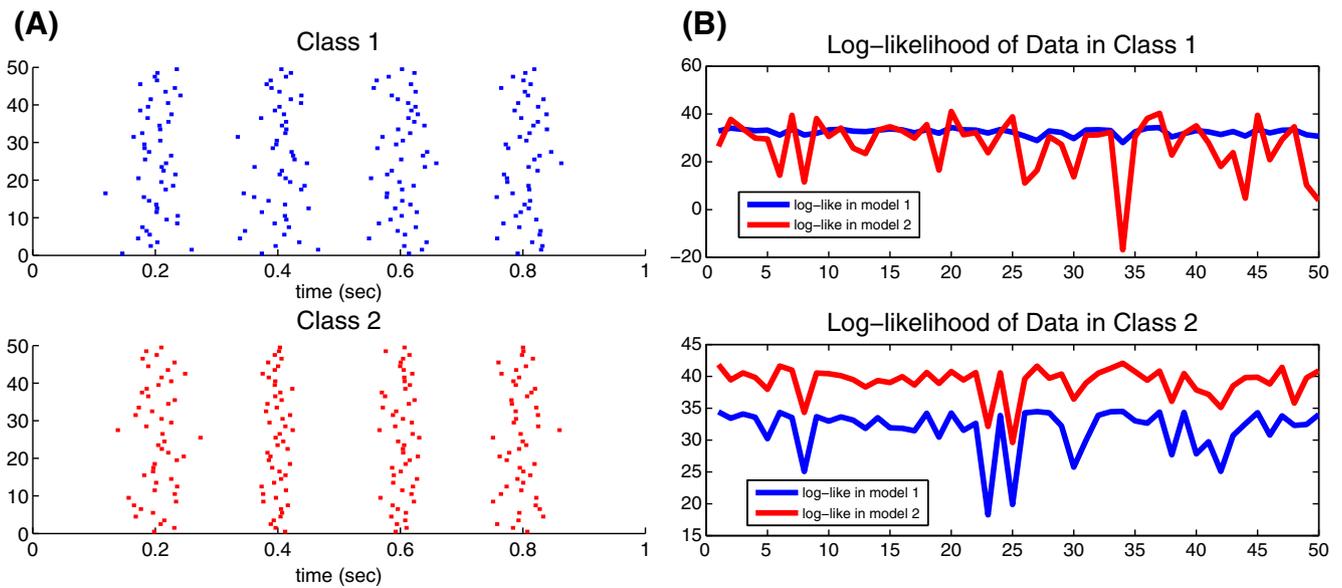


Fig. 8 (a) 50 spike trains generated from Class 1 (upper panel) and Class 2 (lower panel), respectively. Each dot indicates the time of a spike, and one row is for one trial. (b) The log-likelihoods of these 100 spike trains in the model for Class 1 (blue

line) and the model for Class 2 (red line). The upper panel shows the 50 spike trains from Class 1, and the lower panel shows the other 50 spike trains from Class 2

all trains have the same number of spikes and the distribution in the projection space follows a Gaussian model (by Eq. (12)). Specifically, let the time period be $[0, 1]$ s.

- In Class 1, the mean spike train is $[0.2 \ 0.4 \ 0.6 \ 0.8]$, and the covariance is $0.001 \cdot I_5$. We generate 50 spike trains with these mean and covariance. The data is shown in Fig. 8(a) (upper panel).
- In Class 2, the mean spike train is also $[0.2 \ 0.4 \ 0.6 \ 0.8]$, but the covariance is

$$0.001 \cdot \begin{bmatrix} 1 & -0.9 & 0.9 & -0.9 & 0.9 \\ -0.9 & 1 & -0.9 & 0.9 & -0.9 \\ 0.9 & -0.9 & 1 & -0.9 & 0.9 \\ -0.9 & 0.9 & -0.9 & 1 & -0.9 \\ 0.9 & -0.9 & 0.9 & -0.9 & 1 \end{bmatrix}.$$

We also generate 50 spike trains with these mean and covariance. The data is shown in Fig. 8(a) (lower panel).

Note that both classes have the same mean spike train. So if we only use mean-based criterion (Eq. (13)) for classification, the accuracy rate should be around 50 %. In Fig. 8(b), we show that the covariance-based log-likelihoods (by Eq. (12)) of these 100 trains using the models in Classes 1 and 2, respectively. For the 50 trains in Class 1, 34 of them have larger likelihood in model 1, and for the 50 trains in Class 2, all of

them have larger likelihood in model 2. Therefore, the classification accuracy is $(34 + 50)/100 = 84 \%$.

3 Experimental results

In the Theory and Methods section, 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 . Based on the optimal warping functions from each train to the mean, we then compute the covariance matrix using the tangent projection in the SRVF space. In this section, we will apply these new tools to perform some statistical analysis on a real experimental recording.

The neural data was previously used and clearly described when we first developed the metrics d_p between spike trains (Wu and Srivastava 2011). Briefly, a microelectrode array was implanted in the arm area of primary motor cortex (MI) in a juvenile male macaque monkey (*Macaca mulatta*). Signals were filtered, amplified and recorded digitally using a Cerebus acquisition system (Cyberkinetics Inc.). Single units were manually extracted using Offline Sorter (Plexon Inc.). 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. Basically, the targets in the SP task are all fixed at the four corners of a square and the movement is stereotyped. At the

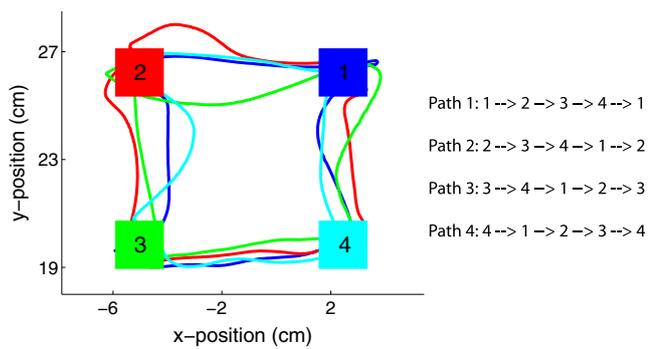


Fig. 9 Four trajectories of hand movement in the SP task. The four colors (*blue, red, green, and cyan*) indicate the trajectories started at corners 1, 2, 3, and 4, respectively, where the corners are also shown in the correspond colors. The four paths are also described with text legend on the right side of the plot

start of a trial, the target appeared at one corner and the monkey had to move the cursor to it. Once reached, the target jumped counterclockwise to the next corner and the monkey had to move the cursor again. The trial ended when the starting corner was reached again. The example movement in each path is shown in Fig. 9 (figure is extracted from Wu and Srivastava 2011). Each sequence of five 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 s. As the new metrics and statistics are defined on a fixed time interval, we normalize the kinematics and spiking

activity in each trial to 5 s. All of the surgical and behavioral procedures were approved by the University of Chicago's IACUC and conform to the principles outlined in the guide for the care and use of laboratory animals.

Here we will apply the spike train statistics and the associated probability models to perform classification on the behaviors (four different paths). We at first selectively choose spiking activity from one neuron which shows significant tuning over four different paths. For illustration, ten spike trains from each movement path as well as the reaching times at the five corners are shown in Fig. 10(a). 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 Wu and Srivastava 2011). Here we choose three different values of λ when estimating the mean spike trains in each path. These values are $(0.1, 1, 10)\lambda_0$, or $(0.65, 6.5, 65)$. 0.65 is a very very small penalty which allows most time warpings in the time domain. When the value of λ gets large (e.g. when $\lambda = 65$), only moderate time warping will be allowed for optimal matching.

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 in each path, we apply the MCP-algorithm to compute their mean for $\lambda = 0.65, 6.5$, and 65. The result is shown in Fig. 10(b). We see that the number of spikes in the mean is a decreasing function with respect to the

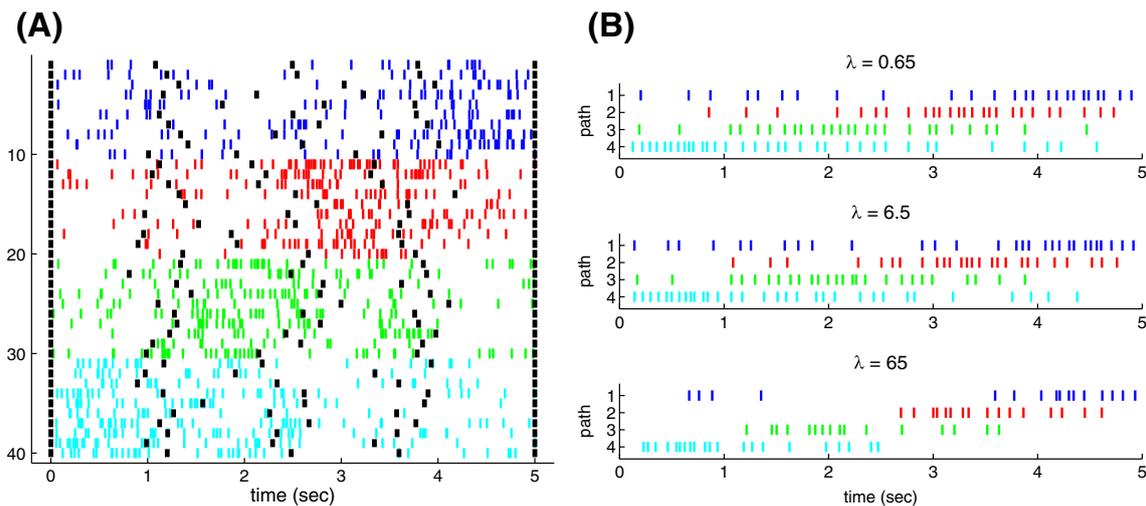


Fig. 10 (a) Ten example spike trains generated from each path. The four colors (*blue, red, green, and cyan*) indicate spiking activity during hand movement in paths 1, 2, 3, and 4, respectively. Each *thin vertical line* indicates the time of a spike, and one row

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)

value of λ . Nevertheless, all these means appropriately represent the firing patterns in the corresponding path. 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 (Victor and Purpura 1996, 1997) has been done with the d_2 metric where we choose $\lambda = 10(E + E)/(2T) = 65$ (see Wu and Srivastava 2011). 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 compute N^2 pairs of distances (between training and testing trains). To simplify the computation, we propose a mean-based method 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, the computation in classification will only have the order $O(N)$. This efficient computational strategy has been used in (Wu and Srivastava 2011). However, in that case, the mean was estimated by the MM-algorithm where we need to assume λ is sufficiently small. In the classification, we have to reassign a large value for λ (e.g. $\lambda = 65$ in the SP data). This large value is needed here because a very small λ would lose the contribution in the warping term which will lead to poor classification.

To overcome this inconsistency, we, in this paper, have proposed the general MCP-algorithm to estimate the mean spike train of a set of spike trains for any value of $\lambda (> 0)$. The estimation result from training data is shown in Fig. 10(b). For the 120 test trains (30 trains from each path), we label each train by the shortest distance over the four distances to the four

Table 1 Classification performance by various methods

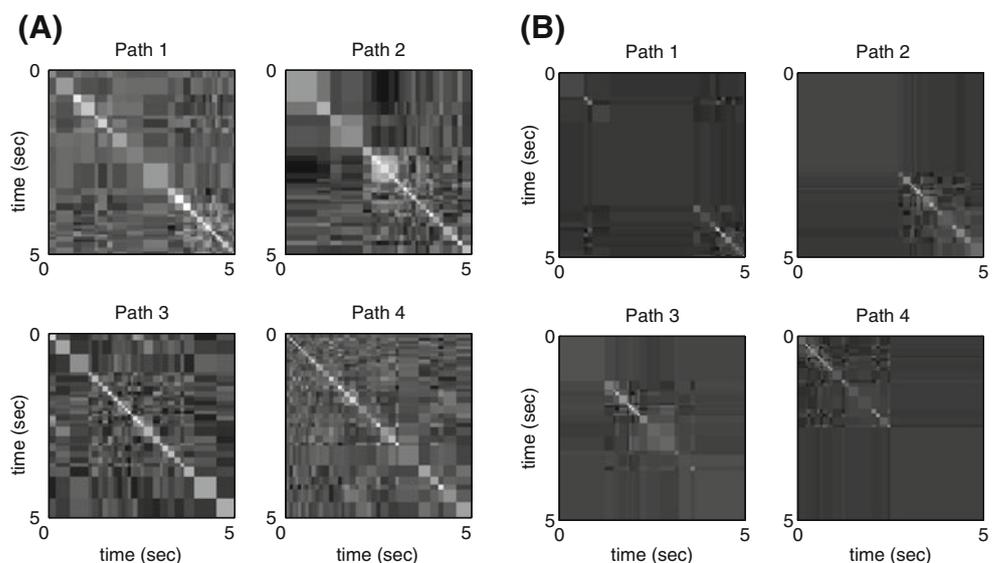
Method	Comp cost	$\lambda = 0.65$ (%)	$\lambda = 6.5$ (%)	$\lambda = 65$ (%)
Pairwise distance	$O(N^2)$	30.0	70.8	90.0
Mean-distance	$O(N)$	28.3	68.3	87.5
Warping-distance	$O(N)$	65.0	74.2	20.0
Warping-based likelihood	$O(N)$	53.3	51.7	13.3

means in the training set. When $\lambda = 65$, the accuracy of this mean-distance classification is 87.5 % (105/120), about the same as that based on pairwise-distances. The result with other two values of λ is also shown in Table 1. In contrast to the pairwise classification, the mean-distance method has comparable accuracies (28.3 %, 68.3 %, 87.5 % vs. 30.0 %, 70.8 %, 90.0 % for $\lambda = 0.65, 6.5, 65$, respectively), but significantly outperforms in terms of efficiency (linear order vs. quadratic order).

Based on the mean spike train for each value of λ , we compute the covariance matrices using Eq. (10). The images of the covariance for $\lambda = 0.65$ and $\lambda = 65$ are shown in Fig. 11. We see that when $\lambda = 0.65$, the covariances has more detailed structure. This is because the means in this case have more spikes (due to little warping penalty). Nevertheless, the general patterns in covariance are the same for different values of penalty. For example, more variability for both λ values when the hand moves upward.

The classification on the test trains can also be done with the warping distance and the associated covariance. We here use a Gaussian likelihood method on

Fig. 11 (a) The covariance images in the four paths when $\lambda = 0.65$, where *light colors* in the images indicate higher variability whereas *dark colors* indicate lower variability. (b) Same as (a) except for $\lambda = 65$. It is apparent that the covariance has more detailed structure when $\lambda = 0.65$



the warping distance and covariance. If we ignore the covariance (i.e. let the covariance be identity matrix), then the model will be isometric and the likelihood will only depend on the warping distance from the mean. The classification result is also shown in Table 1. As the model only depends on the warping term (not about the cost on matching), small values of λ (e.g. 0.65, 6.5) result in better classification performance than large ones (e.g. 65). We also note that incorporating covariance in this case actually makes the performance worse. We believe this is due to insufficient training data for the covariance estimation. The numbers of spikes in the means are around 20 to 30, and the covariance matrices will be in their quadratic orders. However, we only have 30 (or up to 60) spike trains in the training data for each path. We plan to address this issue by making some constraint on the covariance (e.g. assuming sparsity or other local structures). This is further discussed in the Discussion Section.

We here also present a preliminary comparison with the model-based methods which are widely used for modeling and analysis of spike train data. In this case, we assume the spike train data in each movement path follows an inhomogeneous Poisson process. We use the same 30 training trials in each path for estimating the Poisson rate function $\rho_d(t)$, $t \in [0, T]$, $d = 1, 2, 3, 4$ (with bin size 100 ms). For each spike train in the test data, we perform the standard likelihood-based classification based on the estimated $\rho(t)$. It turns out that the accuracy rate over 120 test trials is 86.7 %, which is comparable to the Mean-Distance classification (87.5 %) when $\lambda = 65$.

4 Discussion

Based on the novel framework introduced in Wu and Srivastava (2011), 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 proof 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 where λ is assumed sufficiently small. (3) By adopting the standard theory in Riemannian manifold, we propose a framework to estimate covariance of the spike trains using optimal warping functions from each train in the set to the mean spike train. (4) We further propose a simple Gaussian

model on the estimated mean and covariance, and perform likelihood-based classification on simulations as well as real experimental data.

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 statistically 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 which are often motivated by real neurophysiological mechanisms (Abbott and Sejnowski 1999; Dayan and Abbott 2001). The new set of tools in this project will provide an alternative approach for spike train analysis to the classical methods such as firing rate models and temporal models (Rieke et al. 1997; Brown et al. 2002; Kass et al. 2005). The new framework is expected to be beneficial to the spike train analysis and neural coding community. One key benefit in the new method is that it provides a principled measure for the variability in the entire time domain. However, we emphasize that either of the two frameworks has its advantages. For example, the statistical summaries (e.g. mean and covariance) in the new framework are model free. This, in practice, would be an advantage when the parametric models (e.g. Poisson processes or other point processes) can not fit the data well. In contrast, when the model-based methods do appropriately fit the data, they can make more detailed inferences across the time domain (e.g. infer the dynamic process at any time instant).

The penalty coefficient λ in $d_p[\lambda]$ balances the goodness-of-matching and the degree of time warping in measuring distance between two spike trains. An appropriate value of λ is essentially important. Based on the exact bound in the matching term and penalty term (Wu and Srivastava 2011), we have provided an empirical value range for λ when $p = 1$ and $p = 2$. The MCP-algorithm in this paper estimates mean spike trains for any $\lambda > 0$. However, we note that when λ is significantly large, only moderate warpings are allowed, which will result in very few number of spike trains in the mean. In this case, the distance will be mainly based on the numbers of spikes in the two trains, and the resulting mean cannot reasonably characterize the firing pattern in the data (e.g. as shown in Fig. 3(d)). In contrast, if λ is too small, any time warping would be allowed in the distance computation. The distance between two spike trains will mainly depend on the matching term and the information from the warping term can be largely ignored. This will often result in poor classification performance. For example, in the

spike train data from motor cortex, we have show that the classification performance is much better when $\lambda = 65$ (87.5 %) than when $\lambda = 0.65$ (28.3 %).

Based on the optimal time warping functions to the mean, we further compute the covariance matrix by the classical projection in the tangent space in a Riemmanian manifold. We show that a Gaussian model can be applied on the estimated mean and covariance, and it provides reasonable performance on the classification in simulation as well as real data. However, due to great number of parameters of the covariance matrix and limited sample size, the estimation of covariance often has the non-robustness issue: With number of ISIs in mean being M , the number of parameters in the covariance matrix is $M(M + 1)/2$. As the sample size N is often small (e.g. $N = 30$ in the motor cortical training data), the estimation of all entries in the covariance may be greatly overfit. In Section 3, we observe that the classification performance by the likelihood (when mean and covariance are used) is even worse than that by warping distance (when only mean is used). To address this issue, we will investigate some constraint on the covariance. Some local structures (e.g. banded, tapered, sparse, etc) will be examined for a robust and accurate estimation of the covariance (and its inverse) (Bickel and Levina 2008; Levina et al. 2008).

Note that so far the covariance is only based on the warping functions from the mean, with the matching term ignored. To maximize the information characterized by warping, it is necessary to have a small value of λ (e.g. the covariances when $\lambda = 0.65$ provides more details than those when $\lambda = 0.65$ in the motor cortical data). In the future, we will investigate a more general covariance that includes information for both warping and matching terms. That is, we not only describe that warping variability from any one train to the sample mean, but also describe the unmatched spikes between them. If such a covariance can be well formulated, we should be able to build a generative model on the spike trains. In that case, the spike train can be generated from a random sample (by the covariance) starting from the center (or mean spike train). This generative model could have important application in cases when real data is too complicated to be well characterized by conventional time-based models (Rieke et al. 1997; Brown et al. 2002; Kass et al. 2005).

Acknowledgements 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 at the University of Chicago for providing experimental data.

Appendix A: Existence of the optimal warping function

We at first state one theoretical fact on the optimal time warping on the metric d_p , which is actually the same as Lemma 1 in Wu and Srivastava (2011). We restate it here to emphasize its importance.

Lemma 1 For $p \geq 1$ and any positive diffeomorphism γ from $[a, c]$ to $[b, d]$ (i.e. $\gamma(a) = b, \gamma(c) = d, 0 < \dot{\gamma}(t) < \infty$), the time warping has the following optimal cost:

$$\inf_{\gamma \in \Gamma} \int_a^c |1 - \dot{\gamma}(t)^{1/p}|^p dt = |(c - a)^{1/p} - (d - b)^{1/p}|^p.$$

The infimum is reached if the warping is linear from $[a, c]$ to $[b, d]$ (i.e. $\dot{\gamma}(t) = \frac{d-b}{c-a}$). This condition is also necessary for any $p > 1$.

Based on this result, we have the following theorem on the existence of the optimal warping function between any two spike trains.

Theorem 1 Assume $f(t) = \sum_{i=1}^M \delta(t - t_i) \in \mathcal{S}_M$ and $g(t) = \sum_{j=1}^N \delta(t - s_j) \in \mathcal{S}_N$ are two spike trains in $[0, T]$. Then there exists a warping function $\gamma^* \in \Gamma_{f,g}$, such that

$$d_p[\lambda](f, g) = \left(X([f], [g \circ \gamma^*]) + \lambda \int_0^T |1 - \dot{\gamma}^*(t)^{1/p}|^p dt \right)^{1/p}, \quad (14)$$

where $1 \leq p < \infty$ and $0 < \lambda < \infty$.

Proof For any warping function $\gamma \in \Gamma$, we have two sets of spike times, $[f]$ and $[g \circ \gamma]$. Let $\{a_l\}_{l=1}^L$ be the intersection of $[f]$ and $[g \circ \gamma]$, where L is the cardinality of $\{a_l\}$. This implies $X([f], [g \circ \gamma]) = M + N - 2L$. Denoting $a_0 = 0$ and $a_{L+1} = T$, we define a warping function $\tilde{\gamma} \in \Gamma_{f,g}$ as follows:

$$\tilde{\gamma}(t) = \frac{\gamma(a_l) - \gamma(a_{l-1})}{a_l - a_{l-1}}(t - a_{l-1}) + \gamma(a_{l-1}), \text{ if } a_{l-1} \leq t \leq a_l, l = 1, \dots, L + 1.$$

For illustration, we show the warping functions γ and $\tilde{\gamma}$ in Fig. 12, where $\tilde{\gamma}$ is basically a piecewise linear version of γ by linearly connecting all grid points $(a_l, \gamma(a_l)), l = 0, \dots, L + 1$. We note that $\tilde{\gamma}$ remains the time matches between $\{a_l\}$ and $\{\gamma(a_l)\}$. This implies that

$$X([f], [g \circ \tilde{\gamma}]) \leq M + N - 2L = X([f], [g \circ \gamma]).$$

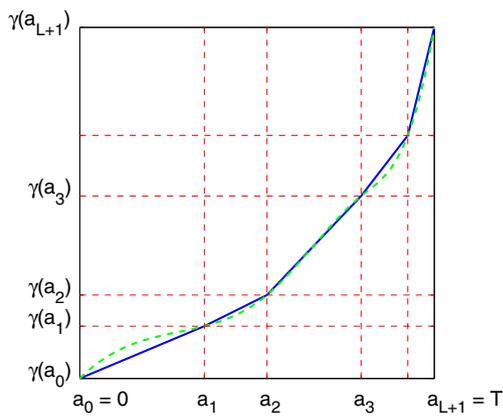


Fig. 12 Any warping function $\gamma \in \Gamma$ (green dashed line) and a function $\tilde{\gamma} \in \Gamma_{f,g}$ (blue solid line) obtained by linearly interpolating γ at the intersection of $[f]$ and $[g \circ \gamma]$

As $\tilde{\gamma}$ is linear from $[a_{l-1}, a_l]$ to $[\gamma(a_{l-1}), \gamma(a_l)]$, using Lemma 1, we have

$$\int_{a_{l-1}}^{a_l} |1 - \dot{\tilde{\gamma}}(t)|^{1/p} dt \leq \int_{a_{l-1}}^{a_l} |1 - \dot{\gamma}(t)|^{1/p} dt, \quad l = 1, \dots, L + 1.$$

Therefore,

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

This indicates $\tilde{\gamma} \in \Gamma_{f,g}$ provides a better time warping to measure the distance between f and g than γ . Note that the set $\Gamma_{f,g}$ is finite. This implies that,

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

Therefore, there exists $\gamma^* \in \Gamma_{f,g} \subset \Gamma$, such that

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

Appendix B: Convergence of the MCP-algorithm

Theorem 2 Denote the estimated mean in the j th iteration of the MCP-algorithm as $S^{(j)}$. Then the sum of squared distances $\sum_{i=1}^N d_2(S_i, S^{(j)})^2$ decreases iteratively. That is,

$$\sum_{i=1}^N d_2(S_i, S^{(j+1)})^2 \leq \sum_{i=1}^N d_2(S_i, S^{(j)})^2.$$

Proof In the j th iteration, we at first perform the *Matching Step* and found the optimal warping γ_i from S_i to $S^{(j)}$, $i = 1, \dots, N$. That is,

$$d_2(S_i, S^{(j)})^2 = X([S_i \circ \gamma_i], [S^{(j)}]) + \lambda \left\| 1 - \sqrt{\dot{\gamma}_i} \right\|^2.$$

Then, we perform the *Centering Step* to compute the extrinsic mean of warping functions $\{\gamma_i\}_{i=1}^N$. As shown in Section 2.2, the mean warping function, $\bar{\gamma}$, of $\{\gamma_i\}_{i=1}^N$ is also piecewise linear and its SRVF has the closed form given by Eq. (6). By the definition of the extrinsic mean, we have

$$\begin{aligned} \sum_{i=1}^N \left\| 1 - \sqrt{\dot{\gamma}_i} \right\|^2 &\geq \sum_{i=1}^N \left\| \sqrt{\dot{\bar{\gamma}}} - \sqrt{\dot{\gamma}_i} \right\|^2 \\ &= \sum_{i=1}^N \left\| 1 - \sqrt{\dot{\gamma}_i \circ \bar{\gamma}^{-1}} \right\|^2 \end{aligned}$$

We then apply $\bar{\gamma}^{-1}$ to both $S_i \circ \gamma_i$ and $S^{(j)}$. In this operation, the spikes in both trains are moved with the same warping function. Using the definition of the d_2 metric, we have

$$\begin{aligned} \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 &\leq \sum_{i=1}^N X([S_i \circ (\gamma_i \circ \bar{\gamma}^{-1})], [S^{(j)} \circ \bar{\gamma}^{-1}]) \\ &\quad + \lambda \left\| 1 - \sqrt{\dot{\gamma}_i \circ \bar{\gamma}^{-1}} \right\|^2. \end{aligned}$$

One can easily verify that

$$\square \quad X([S_i \circ \gamma_i], [S^{(j)}]) = X([S_i \circ \gamma_i \circ \bar{\gamma}^{-1}], [S^{(j)} \circ \bar{\gamma}^{-1}]).$$

Therefore,

$$\begin{aligned} \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 &\leq \sum_{i=1}^N X([S_i \circ \gamma_i], [S^{(j)}]) \\ &\quad + \lambda \left\| 1 - \sqrt{\gamma_i} \right\|^2 \\ &= \sum_{i=1}^N d_2(S_i, S^{(j)})^2. \end{aligned}$$

Next, we perform the *Pruning Step* and update the number of spikes in the mean. $[S^{(j)} \circ \bar{\gamma}^{-1}]$ is the set of spikes in $[S^{(j)} \circ \bar{\gamma}^{-1}]$ that appear more than $N/2$ times in $\{[S_i \circ \gamma_i \circ \bar{\gamma}^{-1}]\}$. Base on the definition of the d_2 metric, we have

$$\begin{aligned} \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 \\ \leq \sum_{i=1}^N X([S_i \circ (\gamma_i \circ \bar{\gamma}^{-1})], [S^{(j)} \circ \bar{\gamma}^{-1}]) \\ + \lambda \left\| 1 - \sqrt{\gamma_i \circ \bar{\gamma}^{-1}} \right\|^2. \end{aligned}$$

Using the basic rule of the Exclusive OR, it is easy to find that

$$\begin{aligned} \sum_{i=1}^N X([S_i \circ (\gamma_i \circ \bar{\gamma}^{-1})], [S^{(j)} \circ \bar{\gamma}^{-1}]) \\ \leq \sum_{i=1}^N X([S_i \circ (\gamma_i \circ \bar{\gamma}^{-1})], [S^{(j)} \circ \bar{\gamma}^{-1}]). \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 \\ \leq \sum_{i=1}^N X([S_i \circ (\gamma_i \circ \bar{\gamma}^{-1})], [S^{(j)} \circ \bar{\gamma}^{-1}]) \\ + \lambda \left\| 1 - \sqrt{\gamma_i \circ \bar{\gamma}^{-1}} \right\|^2 \leq \sum_{i=1}^N d_2(S_i, S^{(j)})^2 \end{aligned}$$

Finally, we test whether one spike in $[S^{(j)} \circ \bar{\gamma}^{-1}]$ that appears the minimal number of times in $\{[S_i \circ \gamma_i \circ \bar{\gamma}^{-1}]\}$ can be removed.

- (i) If the spike can be removed, the remaining set of spikes is denoted as $[S^{(j)} \circ \bar{\gamma}^{-1}]$. This is un-

der the condition that $\sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 \leq \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2$. In this case, we update the mean spike train $S^{(j+1)} = S^{(j)} \circ \bar{\gamma}^{-1}$, and the number of spikes in the mean $n = |S^{(j+1)}|$. In this case, we have

$$\begin{aligned} \sum_{i=1}^N d_2(S_i, S^{(j+1)})^2 &= \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 \\ &\leq \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 \\ &\leq \sum_{i=1}^N d_2(S_i, S^{(j)})^2 \end{aligned}$$

- (ii) If the spike cannot be removed, we update the mean spike train $S^{(j+1)} = S^{(j)} \circ \bar{\gamma}^{-1}$, and the number of spikes in the mean $n = |S^{(j+1)}|$. In this case, we also have

$$\begin{aligned} \sum_{i=1}^N d_2(S_i, S^{(j+1)})^2 &= \sum_{i=1}^N d_2(S_i, S^{(j)} \circ \bar{\gamma}^{-1})^2 \\ &\leq \sum_{i=1}^N d_2(S_i, S^{(j)})^2. \end{aligned}$$

In summary, we have shown that the sum of squared distances $\sum_{i=1}^N d_2(S_i, S^{(j)})^2$ decreases iteratively. \square

References

- Abbott, L.F., & Sejnowski, T.J. (1999). *Neural codes and distributed representations: foundations of neural computation*. The MIT Press.
- Aronov, D. (2003). Fast algorithm for the metric-space analysis of simultaneous responses of multiple single neurons. *Journal of Neuroscience Methods*, *124*, 175–179.
- Aronov, D., & Victor, J. (2004). Non-Euclidean properties of spike train metric spaces. *Physical Review E*, *69*, 061905.
- Aronov, D., Reich, D.S., Mechler, F., Victor, J. (2003). Neural coding of spatial phase in v1 of the macaque monkey. *Journal of Neurophysiology*, *89*, 3304–3327.
- Bickel, P.J., & Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, *36*, 199–227.
- Bilodeau, M., & Brenner, D. (1999). *Theory of multivariate statistics*. Springer.
- Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York: Wiley.
- Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E., Frank, L.M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, *14*, 325–346.
- Dayan, P., & Abbott, L.F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. The MIT Press.

- Dryden, I.L., Koloydenko, A., Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, 3, 1102–1123.
- Dubbs, A.J., Seiler, B.A., Magnasco, M.O. (2010). A fast L^p spike alignment metric. *Neural Computation*, 22, 2785–2808.
- Houghton, C. (2009). Studying spike trains using a van Rossum metric with a synapse-like filter. *Journal of Computational Neuroscience*, 26, 149–155.
- Houghton, C., & Sen, K. (2008). A new multineuron spike train metric. *Neural Computation*, 20, 1495–1511.
- Hunter, J.D., & Milton, J.G. (2003). Amplitude and frequency dependence of spike timing: implications for dynamic regulation. *Journal of Neurophysiology*, 90, 387–394.
- Karcher, H. (1977). Riemann center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30, 509–541.
- Kass, R.E., & Ventura, V. (2001). A spike-train probability model. *Neural Computation*, 13, 1713–1720.
- Kass, R.E., & Vos, P.W. (1997). *Geometric foundations of asymptotic inference*. Wiley.
- Kass, R.E., Ventura, V., Brown, E.N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94, 8–25.
- Klassen, E., Srivastava, A., Mio, W., Joshi, S.H. (2004). Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 372–383.
- Kreuz, T., Haas, J.S., Morelli, A., Abarbanel, H., Politi, A. (2007). Measuring spike train synchrony. *Journal of Neuroscience Methods*, 165, 151–161.
- Kurtek, S., Srivastava, A., Wu, W. (2011). Signal estimation under random time-warpings and nonlinear signal alignment. In *Neural Information Processing Systems (NIPS)*.
- Levina, E., Rothman, A., Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested lasso penalty. *The Annals of Applied Statistics*, 2, 245–263.
- Lim, D., & Capranica, R.R. (1994). Measurement of temporal regularity of spike train responses in auditory nerve fibers of the green treefrog. *Journal of Neuroscience Methods*, 52, 203–213.
- MacLeod, K., Backer, A., Laurent, G. (1998). Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature*, 395, 693–698.
- Michor, P.W., & Mumford, D. (2007). An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23, 74–113.
- Paiva, A.R.C., Park, I., Principe, J.C. (2009a). A comparison of binless spike train measures. *Neural Computing and Applications*. doi:10.1007/s00521-009-0307-6.
- Paiva, A.R.C., Park, I., Principe, J.C. (2009b). A reproducing kernel Hilbert space framework for spike train signal processing. *Neural Computation*, 21, 424–449.
- Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25, 127–154.
- Quiroga, R.Q., Kreuz, T., Grassberger, P. (2002). Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical Review E*, 66, 041904.
- Ramsay, J.O., & Silverman, B.W. (2005). *Functional data analysis* (2nd ed.). Springer Series in Statistics.
- Rencher, A.C. (2002). *Methods of multivariate analysis*. Wiley.
- Rieke, F., Warland, D., Ruyter van Steveninck, R.R., Bialek, W. (1997). *Spikes: Exploring the neural code*. MIT Press.
- Schreiber, S., Fellous, J., Whitmer, D., Tiesinga, P., Sejnowski, T. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52–54, 925–931.
- Srivastava, A., & Jermyn, I.H. (2009). Looking for shapes in two-dimensional, cluttered point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9), 1616–1629.
- Tukey, J.W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.
- Valderrama, M.J. (2007). An overview to modelling functional data. *Computational Statistics*, 22, 331–334.
- van Rossum, M.C.W. (2001). A novel spike distance. *Neural Computation*, 13, 751–763.
- Victor, J.D., & Purpura, K.P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of Neurophysiology*, 76, 1310–1326.
- Victor, J.D., & Purpura, K.P. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Network*, 8, 127–164.
- Victor, J.D., Goldberg, D.H., Gardner, D. (2007). Dynamic programming algorithms for comparing multineuronal spike trains via cost-based metrics and alignments. *Journal of Neuroscience Methods*, 161, 351–360.
- Wu, W., & Srivastava, A. (2011). An information-geometric framework for statistical inferences in the neural spike train space. *Journal of Computational Neuroscience*, 31, 725–748.
- Younes, L., Michor, P.W., Shah, J., Mumford, D. (2008). A metric on shape space with explicit geodesics. *Rendiconti Lincei Matematica e Applicazioni*, 9, 25–57.