

An information-geometric framework for statistical inferences in the neural spike train space

Wei Wu · Anuj Srivastava

Received: 25 January 2011 / Revised: 31 March 2011 / Accepted: 19 April 2011 / Published online: 17 May 2011
© Springer Science+Business Media, LLC 2011

Abstract Statistical inferences are essentially important in analyzing neural spike trains in computational neuroscience. Current approaches have followed a general inference paradigm where a parametric probability model is often used to characterize the temporal evolution of the underlying stochastic processes. To directly capture the overall variability and distribution in the space of the spike trains, we focus on a data-driven approach where statistics are defined and computed in the function space in which spike trains are viewed as individual points. To this end, we at first develop a parametrized family of metrics that takes into account different warpings in the time domain and generalizes several currently used spike train distances. These new metrics are essentially penalized L^p norms, involving appropriate functions of spike trains, with penalties associated with time-warping. The notions of means and variances of spike trains are then defined based on the new metrics when $p = 2$ (corresponding to the “Euclidean distance”). Using some restrictive conditions, we present an efficient recursive algorithm, termed Matching-Minimization algorithm, to compute the sample mean of a set of spike trains with arbitrary numbers of spikes. The proposed metrics as well as the mean spike trains are demonstrated using simulations as well as an experimental recording from the motor

cortex. It is found that all these methods achieve desirable performance and the results support the success of this novel framework.

Keywords Statistical inferences · Spike train space · Information geometry · Spike train metrics · L^p norms

1 Introduction

In the field of computational neuroscience, spike trains have been extensively examined with detailed biophysical models and artificial neural networks (Dayan and Abbott 2001). Statistical inferences play an essential role in the analysis; they provide estimates of model parameters, assess uncertainty about them, and provide a framework for prediction of future activity (Perkel et al. 1967a, b; Curran-Everett and Benos 2004; Kass et al. 2005). In this regard, the parametric probability models, such as Poisson processes or point processes in general, have been used predominantly (Tukey 1977; Box et al. 1978; Kass and Ventura 2001). Indeed, current statistical analysis on neural data can be highlighted in Fig. 1(a), where a probability model is typically a prerequisite to formal inference procedures. This analysis paradigm is now a standard for production of scientific conclusions from experimental results (Brown et al. 2002; Kass et al. 2005).

However, these analysis methods only focus on parametric representations at each given time and therefore have a limited use in evaluating data-driven statistics on spaces where full spike trains are treated as individual data points. Given a sample of spike trains, one may naturally pose questions such as “What is the mean trend of the sample?” and “What is the variability of

Action Editor: Rob Kass

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

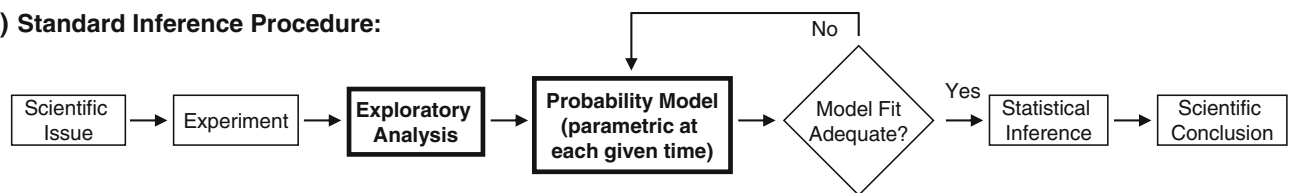
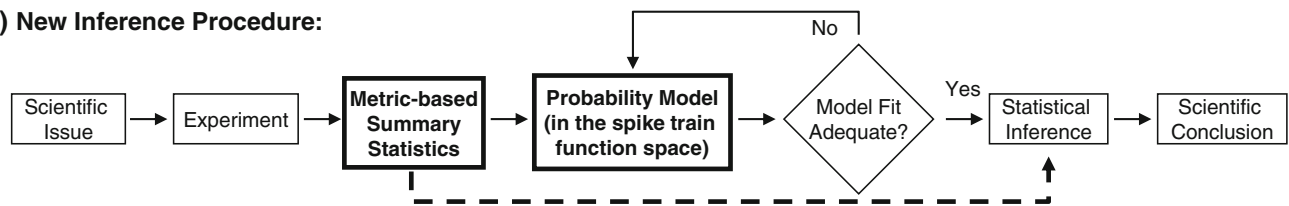
(A) Standard Inference Procedure:**(B) New Inference Procedure:**

Fig. 1 (a) A standard statistical inference procedure in Kass et al. (2005). (b) A new inference procedure where the difference is at pre-analysis and probability model (boxes with boldface

font). The new dashed-line indicates metric-based statistics could directly build statistical inferences

the sample?” Although these questions are very basic, the past models may not be able to answer them as they only characterize variability at each specific time. In this case, an overall measurement is desired for spike train variability across the entire time domain. In this article, we propose a principled framework to address this issue, where we compute metric-based summary statistics in the spike train domain, and then suggest their use in developing statistical inferences. The new procedure is illustrated in Fig. 1(b). In contrast to some informal exploratory analysis for a parametric model in the classical case, we propose to compute metric-based summary statistics in the spike train domain, and then use the results to build a (parametric or nonparametric) probability model in the function space. Moreover, inferences can also be done by directly using summary statistics (see dashed line in Fig. 1(b)). This will be shown in Section 5 of the paper.

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

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

where $d(\cdot, \cdot)$ is a distance between two spike trains. The issue is to choose a distance measure that provides intuitive and tractable solutions to the minimization problem. Not all metrics lead to interesting solutions. For example, the commonly-used Victor–Purpura metrics

(Victor and Purpura 1996, 1997) would result in infinite number of solutions to the minimization in Eq. (1). This is because these metrics are like Manhattan distances and lack the convexity needed to find the unique minimum (e.g. the Karcher mean of two points (0, 0) and (1, 1) under the Manhattan distance is any point (x, y) that satisfies $0 \leq x \leq 1, 0 \leq y \leq 1$, and $x + y = 1$, whereas the mean under the classical Euclidean distance has a unique solution at (0.5, 0.5)). This example underscores the fact that nature of metric is important in defining statistical quantities.

Over the past two decades, a number of approaches have been developed to measure distances or dissimilarities between spike trains, for example, the distances in discrete state, discrete time models (Lim and Capranica 1994; MacLeod et al. 1998; Rieke et al. 1997), those in 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. 2009), those in 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). The applications of these methods have mainly focused on the clustering or classification of spike trains with respect to different stimuli. Their relevance in deriving statistical summaries remain unclear; the statistics defined with these distances may not provide interesting intuition or tractable computations.

This implies that new distance metrics, especially the ones with Euclidean properties, are needed for computing means. In this paper we develop a comprehensive

framework involving a parametrized family of metrics that: (1) incorporate the commonly-used metrics such as the Victor-Purpura metrics and van Rossum metric (van Rossum 2001), and (2) allow us to compute intuitive solutions for means of spike trains. Depending on the values of the parameter, the new metrics can have properties as in a Manhattan distance or a Euclidean distance. We will show that the resulting mean spike train can reasonably summarize the firing pattern of a set of spike trains and can be used to provide an efficient spike train classification.

The computation of summary statistics and derivation of probability models are motivated in part by the recent progress in computing statistics on constrained function spaces, especially in shape analysis of curves, using tools from geometry (Kass and Vos 1997; Klassen et al. 2004; Michor and Mumford 2007; Younes et al. 2008). In these studies, the computation of sample means and covariances of functions are intimately based on the concept of geodesic paths under the selected metrics. In the current paper, we propose a new metric framework that is also motivated by the incorporation of time warping in information geometry and that can be used to compute statistics in the spike train space directly. Although this manuscript does not fully develop the desired statistical framework, it does provide building blocks towards that goal. In the future, these results can be further developed for various classical statistical inferences such as hypothesis tests, confidence intervals, regressions, FANOVA (functional ANOVA), FPCA (functional PCA), and regressions and bootstraps on functional data.

This study significantly extends our recent investigation (Wu and Srivastava 2011) on summary statistics in the function space of spike trains: (1) The metrics are now defined for both smoothed and unsmoothed spike trains with smoothing parameter $\sigma \in [0, \infty)$; (2) The mean statistics are demonstrated using classical homogeneous and inhomogeneous Poisson processes; (3) The new metrics and mean statistics are applied to neural decoding problems in simulated as well as real data; and (4) All mathematical details and theoretical proofs are given in Appendices A–E.

The rest of of this manuscript is organized as follows. Section 2 provides a detailed theory for the proposed metrics, including background, definition, and relationships to the previous metrics. Section 3 presents algorithms for finding optimal time-warplings, a step that is necessary for evaluating the proposed metrics. In Section 4 we introduce certain basic descriptive statistics of spike trains based on the proposed metrics.

Section 5 examines the performance of the new metrics as well as the computed mean spike trains on simulated spike trains and an real experimental recording from motor cortex. Section 6 discusses related studies and presents future work. Appendices A–E provide the mathematical details of the metric space proofs, limiting forms for distance upper bounds, theoretical facts for optimal time warpings, closed-form distance representation in some restrictive condition, and proof of the convergence of the computational algorithm for mean spike trains.

2 Theory

In this section we describe the mathematical framework for establishing a new family of metrics for measuring distance or dissimilarity between spike trains. We start with some background in information geometry and then develop the theory for deriving the proposed metrics.

2.1 Background

The proposed new distances have been motivated in part by use of Fisher–Rao (Riemannian) metric for quantifying probability densities (Srivastava et al. 2007). This metric has many fundamental advantages, including the fact that it is the only Riemannian metric that is invariant to the warping of the time axis (Čencov 1982), and has played an important role in information geometry (Kass and Vos 1997). One of the problems of interest in information geometry is to compute distances between probability densities, by viewing them as points on a certain appropriate space of densities. For instance, it can be used to quantify differences between two densities on $[0, T]$ using geodesic distances under this metric. However, this metric is somewhat complicated to deal with, since it changes from point to point, and it is not straightforward to derive geodesic paths. Luckily, a simple square-root transformation simplifies the task enormously. Bhattacharya (1943) showed that if we analyze the square-root of the densities functions, rather than the densities themselves, then the Fisher–Rao metric becomes the standard L^2 metric and the distance between such square-root densities can be computed using the L^2 norm of their distances. In defining the new metrics, we generalize this idea to include L^p norm for any $p \geq 1$. The other change that we make is that we add a penalty term to the resulting L^p distance that penalizes the

amount of warping. We will show that this generalization builds a family of metrics that not only incorporates the previous distance systems, such as the Victor–Purpura D^{interval} metric and the van Rossum metric, but also provides newer options.

2.2 Definition of the metrics

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 $\mathcal{S} = \cup_{M=0}^{\infty} \mathcal{S}_M$.

Similar to the van Rossum metric (van Rossum 2001), we smooth the spike trains to better capture the time correlation between spikes. As indicated in Schrauwen and van Campenhout (2007), various smoothing functions such as the piecewise linear kernel, the Laplacian kernel, and the Gaussian Kernel, can be used. In this paper, we smooth the spike trains using a bump kernel function

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

which is infinitely differentiable with support $[-\sigma, \sigma]$, $\sigma > 0$ and Z is a normalizing constant such that $\int_0^T K_{\sigma}(t) dt = 1$. Therefore, the smoothed spike train is

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

Note that when σ is sufficiently small, all bump functions $K_{\sigma}(t - t_i)$ are supported within $[0, T]$. Therefore, $\int_0^T f(t) dt = M$. We define the space of all smoothed spike-trains containing M spikes to be:

$$\mathcal{S}_M^{(\sigma)} = \left\{ \sum_{i=1}^M K_{\sigma}(t - t_i) \mid 0 < t_1 < \dots < t_M < T \right\},$$

and the set of all smoothed spike-trains to be $\mathcal{S}^{(\sigma)} = \cup_{M=0}^{\infty} \mathcal{S}_M^{(\sigma)}$.

Similar to the treatment of functions in information geometry, a smoothed spike train function is viewed as a “point” in an infinite-dimensional space of real-valued functions. The set of all trains, $\mathcal{S}^{(\sigma)}$, forms a “manifold” in the space whose geometry characterizes the intrinsic non-Euclidean nature of the spiking

activity (Aronov and Victor 2004). In this study, we propose to use warping in the time domain to measure the variability in the function space. Let Γ be the set of all 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:

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

It is easy to verify that Γ is a *group* with the operation being the composition of functions. Basically, this means that: (1) for any $\gamma_1, \gamma_2 \in \Gamma$, their composition $\gamma_1 \circ \gamma_2 \in \Gamma$; and (2) for any $\gamma \in \Gamma$, its inverse $\gamma^{-1} \in \Gamma$, where the identity is the self-mapping $\gamma_{\text{id}}(t) = t$.

Using the set of time-warping functions Γ , the new metrics will be defined by an optimal matching between two smoothed functions:

Definition 1 We define the metric between $f, g \in \mathcal{S}^{(\sigma)}$ in the following form:

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

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

The integration in Eq. (2) includes two terms, $|f(t)^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p}|^p$ and $|1 - \dot{\gamma}(t)^{1/p}|^p$. The first term, called *matching term*, is used to find a time warping of g , and to measure the L^p distance between f and the warped g ; it measures the *goodness-of-match* between f and g in presence of warping. Here we make use of the term $\dot{\gamma}$ for two reasons: (1) to ensure that the invariance of the integration, $\int_0^T g(\gamma(t))\dot{\gamma}(t) dt = \int_0^T g(t) dt$, i.e. the area under the function curve does not change after warping; and (2) to ensure the symmetry of D_p (see Appendix A). We clarify that we will get the same result if we warp either f , g , or both in the definition of D_p . With the infimum taken over all warpings in Γ , the matching term attempts to find a mapping that minimizes distance between f and the warped g , whereas the second term, called *penalty term*, penalizes the amount of warping. λ is the weight of this penalty term and its value is of great importance as it is used to balance the contributions of the two terms in D_p . We emphasize that D_p in Eq. (2) is a proper metric; that is, it satisfies *non-negativity*, *symmetry*, and *triangle-inequality*. Indeed, D_p shares a lot of similarities to the classical L^p norm in functional analysis. The proof is presented in Appendix A. Basically, the proof is based

on the facts that Γ is a *group* and the L^p norm follows the *triangle-inequality* when $1 \leq p < \infty$.

The distance in Eq. (2) is defined for smoothed spike train with smoothing parameter σ . The value of σ is a choice here and its *optimal* value may depend on a specific application. In the limiting case when $\sigma \rightarrow 0$, the distance $D_p[\sigma, \lambda]$ should measure the distance in the original spike trains. Based on this idea, we directly define the distance between two (unsmoothed) spike trains as follows.

Definition 2 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 , we define the metric between f and g in the following form:

$$d_p[\lambda](f, g) = \inf_{\gamma \in \Gamma} \left(M + N - 2 \sum_{i=1}^M \sum_{j=1}^N \mathbf{1}_{\{t_i = \gamma(s_j)\}} + \lambda \int_0^T |1 - \dot{\gamma}(t)|^{1/p} dt \right)^{1/p}, \quad (3)$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function and $1 \leq p < \infty$ and $0 < \lambda < \infty$.

In Eq. (3), the penalty term is exactly the same as that in Eq. (2), whereas the matching term is already integrated out and independent of the parameter p . In Appendix B, we show that $d_p[\lambda]$ is also a proper metric for any $p \geq 1$ and $\lambda > 0$. Furthermore, we show that $d_p[\lambda]$ is indeed a limiting form of $D_p[\sigma, \lambda]$. That is, for $f, g \in \mathcal{S}$,

$$d_p[\lambda](f, g) = \lim_{\sigma \rightarrow 0} D_p[\sigma, \lambda](f * K_\sigma, g * K_\sigma).$$

To simplify the notation, we write $D_p[0, \lambda] = d_p(\lambda)$ and build a general family of spike train metrics $D_p[\sigma, \lambda]$ with smoothing parameter $\sigma \in [0, \infty)$.

Next, we consider in detail two specific cases of D_p for $p = 1$ and $p = 2$. In each case we will suggest values of λ for practical use and will relate the metric to one of the previous metrics.

2.2.1 “Euclidean distance” \mathbf{D}_2

At first we examine the metric distance when $p = 2$, where the squared distance has the form:

$$D_2^2[\sigma, \lambda](f, g) = \inf_{\gamma \in \Gamma} \left(\int_0^T [\sqrt{f(t)} - \sqrt{g(\gamma(t))\dot{\gamma}(t)}]^2 dt + \lambda \left[1 - \sqrt{\dot{\gamma}(t)} \right]^2 dt \right) \quad (4)$$

Due to its similarity to the L^2 norm, this distance can be viewed as a penalized classical “Euclidean distance” between two spike trains.

Relation to the van Rossum metric When there is no time warping (i.e. $\gamma(t) = t$), it is apparent that

$$D_2^2[\sigma, \lambda](f, g) = \int_0^T [\sqrt{f(t)} - \sqrt{g(t)}]^2 dt.$$

This is the L^2 distance between $\sqrt{f(t)}$ and $\sqrt{g(t)}$ and, therefore, D_2 can be looked as a generalization over the commonly-used van Rossum metric (van Rossum 2001). Beyond the van Rossum metric, however, D_2 also measures the separation between spikes by warping the time domain and penalizes the amount of warping. We claim that such a penalty better captures the temporal dissimilarities between spike trains. In other words, D_2 will be able to capture the temporal spacing between spikes using time warping and describe the dissimilarity as an (increasing) function with respect to such spacing.

Appropriate value of λ The weight λ plays a key role in this analysis and an “ideal” value of λ is the one that provides a good balance between the goodness-of-match and the warping penalty. To strike this balance, we examine the range of values each component in Eq. (4) takes. Assume there are M spikes in f and N spikes in g . Then

$$\begin{aligned} & \int_0^T [\sqrt{f(t)} - \sqrt{g(\gamma(t))\dot{\gamma}(t)}]^2 dt \\ & \leq \int_0^T [f(t) + g(\gamma(t))\dot{\gamma}(t)] dt = M + N. \end{aligned}$$

The upper bound $M + N$ can, in principle, be reached when all spike signals are well-separated and $\gamma(t) = t$. Moreover, we have

$$\begin{aligned} & \lambda \int_0^T [1 - \sqrt{\dot{\gamma}(t)}]^2 dt \\ & = \lambda \left(T + T - 2 \int_0^T \sqrt{\dot{\gamma}(t)} dt \right) \leq 2\lambda T. \end{aligned}$$

This upper bound is exact as it can be reached in some limiting form (see Appendix C).

If $\lambda = \lambda_0 = (M + N)/(2T)$, then $M + N = 2\lambda T$; that is, the two terms will have the same exact upper bounds and therefore contribute equally to the metric distance. However, we note that λ_0 is a very conservative weight. The optimal distance with this value of λ will tend

to match two spike trains, even if the spikes are far apart. For example, for two single-spike trains with non-overlapping spikes, if there is no warping to match, the cost is 2, whereas for any time warping with $\lambda = (1 + 1)/(2T)$, the cost is at most 2. In practice, a larger λ would be needed to constrain larger warpings. Empirically, we have found $\lambda = c\lambda_0$, $5 \leq c \leq 25$ provides a good weight to balance between the goodness-of-match and the warping cost.

Insertion and deletion of a spike Here we examine the metric distance when an elementary operation on a spike, i.e. insertion, deletion, and shift as defined in Victor and Purpura (1997), is performed. Let f be any smoothed spike train, and g be one smoothed spike train with single spike. Then $f + g$ and f can represent a spike train and the signal when one spike is inserted to or deleted from the train. Therefore, the squared distance is

$$\begin{aligned} D_2^2[\sigma, \lambda](f + g, f) &\leq \int_0^T \left[\sqrt{f(t) + g(t)} - \sqrt{f(t)} \right]^2 dt \\ &\leq \int_0^T \left[\sqrt{g(t)} \right]^2 dt = 1. \end{aligned}$$

When the smoothing parameter $\sigma = 0$, $f + g$ has one more spike than f . Any time warping cannot fully match all spikes in $f + g$, but only adds penalty in the total distance. Therefore,

$$d_2^2[\lambda](f + g, f) = 1.$$

Shifting a spike In general, shifting a spike may result in overlapping within one spike train which is difficult to analyze. So we simplify the problem by examining the distance using two single-spike trains, f and g , as is done in van Rossum (2001). Assume that the centers of the two smoothed spikes are μ_1 and μ_2 , respectively, and the kernels are fully matched (i.e. the interval $[\mu_1 - \sigma, \mu_1 + \sigma]$ in f is mapped to the interval $[\mu_2 - \sigma, \mu_2 + \sigma]$ in g , using time warping), we compute the optimal transformation between the two signals. Based on Lemma 1 (see Appendix D), with a full spike match, the squared distance is

$$\begin{aligned} D_2^2[\sigma, \lambda](f, g) &= \inf_{\gamma \in \Gamma} \left[\int_0^{\mu_1 - \sigma} \lambda \left(1 - \sqrt{\dot{\gamma}(t)} \right)^2 dt \right. \\ &\quad \left. + \int_{\mu_1 + \sigma}^T \lambda \left(1 - \sqrt{\dot{\gamma}(t)} \right)^2 dt \right] \\ &= \lambda \left(\sqrt{\mu_1 - \sigma} - \sqrt{\mu_2 - \sigma} \right)^2 \\ &\quad + \lambda \left(\sqrt{T - \mu_1 - \sigma} - \sqrt{T - \mu_2 - \sigma} \right)^2. \end{aligned}$$

In practical time warping, μ_1 and μ_2 are often close to each other. In this case, the squared distance can be approximately simplified as follows:

$$\begin{aligned} D_2^2[\sigma, \lambda](f, g) &= \lambda \left[2(T - 2\sigma) - 2 \left[\sqrt{\mu_1 - \sigma} \sqrt{\mu_2 - \sigma} \right. \right. \\ &\quad \left. \left. + \sqrt{T - \mu_1 - \sigma} \sqrt{T - \mu_2 - \sigma} \right] \right] \\ &\approx \lambda \left[2(T - 2\sigma) \right. \\ &\quad \left. - 2\sqrt{(T - 2\sigma + \mu_2 - \mu_1)(T - 2\sigma - \mu_2 + \mu_1)} \right] \\ &\approx \lambda(\mu_2 - \mu_1)^2 / (T - 2\sigma). \end{aligned} \quad (5)$$

Equation (5) indicates that the metric distance between two single-spike signals is approximately linear with respect to their spike time distance, $|\mu_1 - \mu_2|$. Moreover, we see that with a full spike match, the squared metric resulting from the warping penalty is approximately $\lambda(\mu_2 - \mu_1)^2 / (T - 2\sigma)$. If there is no warping, the squared metric between the two non-overlapped spikes will be 2. Hence, the two spikes can be matched with a time warping if and only if $\lambda(\mu_2 - \mu_1)^2 / (T - 2\sigma) \leq 2$, or equivalently $\lambda \leq 2(T - 2\sigma) / (\mu_2 - \mu_1)^2$.

2.2.2 “Manhattan distance” D_1

We obtain another useful metric when we set $p = 1$:

$$\begin{aligned} D_1[\sigma, \lambda](f, g) &= \inf_{\gamma \in \Gamma} \left(\int_0^T |f(t) - g(\gamma(t))\dot{\gamma}(t)| \right. \\ &\quad \left. + \lambda |1 - \dot{\gamma}(t)| dt \right). \end{aligned} \quad (6)$$

Being a first-order distance like an L^1 norm, the D_1 distance corresponds to a penalized version of the classical “Manhattan distance”.

Appropriate value of λ Similar to metric distance D_2 , we examine the upper bound of each component in Eq. (6) to find an appropriate value for λ . Assume there are M spikes in f and N spikes in g . Then

$$\begin{aligned} \int_0^T |f(t) - g(\gamma(t))\dot{\gamma}(t)| dt &\leq \int_0^T [f(t) + g(\gamma(t))\dot{\gamma}(t)] dt = M + N. \end{aligned}$$

The upper bound $M + N$ can also, in principle, be reached when all spike signals are well-separated and $\gamma(t) = t$. Moreover, we have

$$\lambda \int_0^T |1 - \dot{\gamma}(t)| dt \leq \lambda \int_0^T [1 + \dot{\gamma}(t)] dt = 2\lambda T.$$

This upper bound is exact as it can be reached in some limiting form (see Appendix C). These two upper bounds are exactly the same as those for the D_2 metric. Therefore, we can use the same strategy to choose appropriate value of λ . Empirically, we have found $\lambda = c(M + N)/(2T)$, $2 \leq c \leq 5$ provides a good weight to balance between the goodness-of-matching and warping cost.

Insertion and deletion of a spike What is the D_1 distance when we insert or delete a spike? As in D_2 , let f be any smoothed spike train, and g be one smoothed spike train with single spike. Then $f + g$ and f can represent a spike train and the signal when one spike is inserted to or deleted from the train. The D_1 distance between them is

$$D_1[\sigma, \lambda](f + g, f) \leq \int_0^T |f(t) + g(t) - f(t)| dt = \int_0^T |g(t)| dt = 1.$$

When the smoothing parameter $\sigma = 0$, $f + g$ has one more spike than f . Any time warping cannot fully match all spikes in $f + g$, but only adds penalty in the total distance. Therefore,

$$d_1[\lambda](f + g, f) = 1.$$

Shifting a spike As in D_2 , we simplify the problem by studying the distance between two single-spike signals, f and g . Assuming that the two spikes are fully matched (i.e. the interval $[\mu_1 - \sigma, \mu_1 + \sigma]$ in f is mapped to the interval $[\mu_2 - \sigma, \mu_2 + \sigma]$ in g), we compute the optimal transformation between the two signals. Based

on Lemma 1 (see Appendix D), with a full spike match, the metric distance is

$$\begin{aligned} D_1[\sigma, \lambda](f, g) &= \inf_{\gamma \in \Gamma} \left[\int_0^{\mu_1 - \sigma} \lambda |1 - \dot{\gamma}(t)| dt + \int_{\mu_1 + \sigma}^T \lambda |1 - \dot{\gamma}(t)| dt \right] \\ &= \lambda |(\mu_1 - \sigma) - (\mu_2 - \sigma)| \\ &\quad + \lambda |(T - \mu_1 - \sigma) - (T - \mu_2 - \sigma)| \\ &= 2\lambda |\mu_1 - \mu_2| \end{aligned} \quad (7)$$

Equation (7) indicates that the metric distance D_1 between two single-spike signals is linear with respect to their spike time distance $|\mu_1 - \mu_2|$ (no approximation in contrast to the case for D_2). Moreover, the two spikes can be fully matched with a time warping if and only if $2\lambda |\mu_2 - \mu_1| \leq 2$, or equivalently $\lambda \leq 1/|\mu_2 - \mu_1|$.

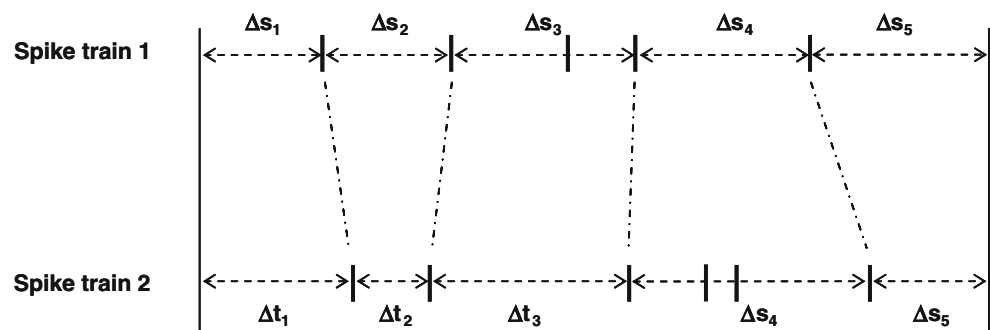
Relation to D^{interval} The Victor–Purpura metric D^{interval} focuses on the duration variation in the inter-spike intervals, or ISIs (in particular, we examine $D^{\text{interval:fix}}$ where an auxiliary spike is placed at time 0 and time T (Victor and Purpura 1997)). A change in an ISI from Δ_1 to Δ_2 is nothing but a time warping between the corresponding spikes. Here we show that when σ is sufficiently small, the proposed D_1 metric provides results equivalent to D^{interval} in nature. In the extreme case when $\sigma = 0$, we compare the metrics for the three elementary operations using $d_1[\lambda]$.

1. Inserting or deleting one spike $\delta(t - t_i)$: There is no time warping in this operation. Then the cost is

$$\int_0^T |\delta(t - t_i)| dt = 1,$$

which is the same as the cost in D^{interval} .

Fig. 2 An illustration of the relationship between ISI variation and time warping



2. Shifting an ISI from Δ_1 to Δ_2 : This is a time warping, and the cost is

$$\inf_{\gamma \in \Gamma} \int_{\Delta_1} \lambda |1 - \dot{\gamma}(t)| dt = \lambda |\Delta_2 - \Delta_1|,$$

which is the same as the cost in $D^{\text{interval}}[q]$ with $q = \lambda$.

Indeed, an optimal mapping between two spike trains can be looked as a path from one to the other via elementary steps, and vice versa, as was suggested by Victor and Purpura (1996). See Fig. 2 for an illustration. The ISI change between the spike trains can be looked as a time warping between the ISIs (Δs_i vs. Δt_i , $i = 1, \dots, 5$). The spikes that are not used in the mapping (the 3rd one in train 1, and the 4th and 5th ones in train 2) can be looked as the ones inserted in or deleted from the spike trains, respectively.

Therefore, $d_1[\lambda] = D_1[0, \lambda] = D^{\text{interval}}[\lambda]$. Here $\sigma = 0$ basically requires that two spikes either fully match or do not match at all. This condition will also be satisfied if σ is sufficiently small such that the spike times can be equivalently represented using ISIs. In real neural data, two consecutive spikes should be at least 1 ms apart. Therefore, for $\sigma < 0.5$ ms, we also have $D_1[\sigma, \lambda] = D^{\text{interval}}[\lambda]$.

3 Matching algorithms

3.1 Dynamic programming for $D_p[\sigma, \lambda]$

In order to evaluate D_p , we need to find an optimal diffeomorphism (time warping) that minimizes the penalized warping distance between two spike train signals. Note that the space of warpings, Γ , is an infinite-dimensional space, so it would be impractical to search for the optimum in the entire space. Luckily, there exists an efficient algorithm, namely the dynamic programming, to solve for the optimal γ , albeit on a finite grid in the interval $[0, T]$. We briefly describe the approach below and refer the reader to Bertsekas (1995) for more detailed description.

At first, we will describe a general procedure for any but fixed parameters $\sigma > 0$, $p \geq 1$, and $\lambda > 0$. The time domain $[0, T]$ is discretized into N equally-sized bins. Then, the diffeomorphism γ is approximated by a piecewise linear homomorphism with path from $(0, 0)$ to (T, T) ($(1, 1)$ to (N, N) in the discretized space), where all segments in the path pass through the grid and have positive slopes (this is to assure the mapping is invertible). An example path is shown in Fig. 3. Now

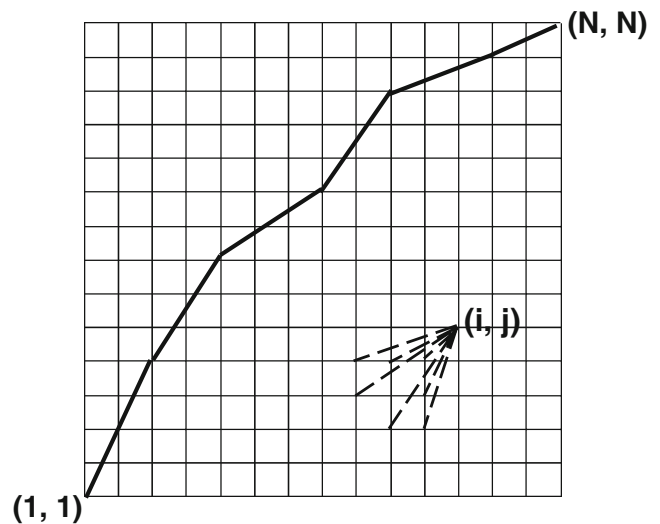


Fig. 3 An illustration of a path from $(1, 1)$ to (N, N) . The plot also shows the seven neighbors for each node (i, j)

the problem is reduced to finding the allowable path that minimizes the energy or cost function,

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

Note that the total energy associated with a γ is an accumulation of energies associated with piecewise linear segments contained in γ . For the nodes (k, l) and (i, j) where $k < i$ and $l < j$, the energy of the linear segment connecting them is

$$E(k, l; i, j) = \int_{[k, i]} \left(\left| f(t)^{1/p} - \left[g(\gamma(t)) \frac{j-l}{i-k} \right]^{1/p} \right|^p + \lambda \left| 1 - \left(\frac{j-l}{i-k} \right)^{1/p} \right|^p \right) dt.$$

To decrease computational cost, one often restricts (k, l) to a small subset N_{ij} that lies below and to the left of (i, j) . A possible choice of seven nodes is shown in Fig. 3. Using this small neighborhood, the optimal γ can be found using standard dynamic programming that has the computational cost of $O(N^2)$.

We remark that although this small neighbor provides convenience and efficiency in the optimal path finding, it restricts the path to a limited number of warpings in each segment. Here the warping velocity, $\dot{\gamma}$, only takes the following seven discrete values: $1/3$, $1/2$, $2/3$, 1 , $3/2$, 2 , and 3 . This limitation may introduce certain numerical errors in practical computations of metric distances between spike trains.

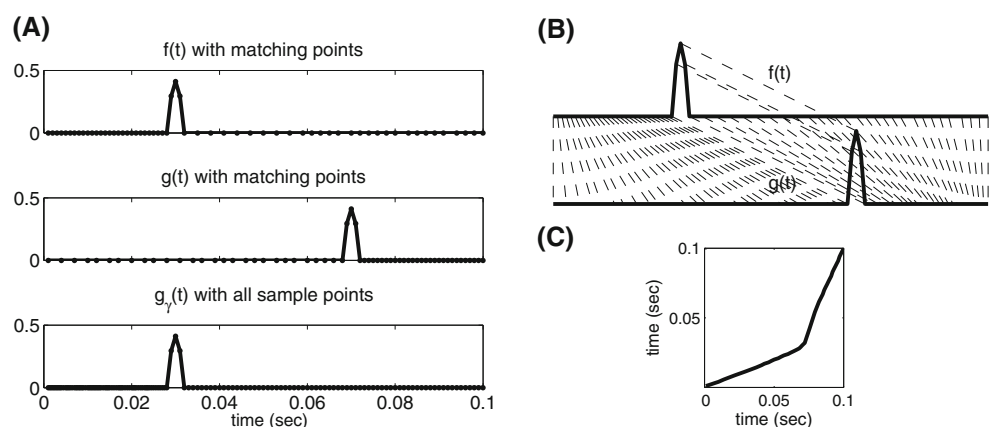
3.2 Dynamic programming for $d_p[\lambda]$

The proposed dynamic programming provides an effective procedure for computing distances between two smoothed spike train functions. Though the method can be used for any value of $\sigma (> 0)$, the time domain has to be discretized which may result in inefficient computation (particularly when N is large). When $\sigma = 0$, the spike times can be equivalently represented by the corresponding ISIs. To compute $d_p[\lambda]$, we can let the grid be the same as the true spike times of the two spike trains (including times 0 and T). As indicated in Lemma 1 in Appendix D, a linear mapping between each pair of ISIs will indeed be an optimal warping. With this new grid, for two spike trains f and g that have m and n spikes respectively, the computational cost will be $O(mn)$. This dynamic programming has the same computational efficiency as that for the Victor–Purpura metric D^{interval} (Victor and Purpura 1996), whereas it can be conducted for any value of parameter $p \geq 1$ ($p = 1$ in D^{interval}).

In particular, when λ is sufficiently small, there is only little penalty for the time warping and the distance is dominated by the matching term. Without loss of generality, we let $m \geq n$. one can show that when $\lambda < 1/(2^{p-1}T)$, an optimal time warping must have the n (of m) spikes in f match with the n spikes in g (this is the most efficient way to lower the matching cost). In this case, we have $n+1$ pairs of ISIs associated the n spikes in f and n spikes in g . Denoting these $n+1$ pairs of ISIs are $\{\Delta s_k\}_{k=1}^{n+1}$ and $\{\Delta t_k\}_{k=1}^{n+1}$, the distance between f and g can be further represented in the following form:

$$d_p[\lambda](f, g) = \left((m - n) + \lambda \sum_{k=1}^{n+1} |\Delta s_k^{1/p} - \Delta t_k^{1/p}|^p \right)^{1/p}. \quad (8)$$

Fig. 4 Matching between two single-spike trains based on D_2 . **(a)** The upper and middle panels show the smoothed spike trains $f(t)$ and $g(t)$ with all matching points in time warping, respectively. The lower panel shows $g(t)$ with all sample points after time warping; that is, $g_\gamma(t) = g(\gamma(t))\dot{\gamma}(t)$. **(b)** The point-to-point time warping between $f(t)$ and $g(t)$. **(c)** The warping path from $[0, 0]$ to $[T, T]$



The detailed demonstration is described in Appendix D. Note that here in addition to computing the distance between f and g , we identify n spikes in f that match the n spikes in g . This identification will be crucially important in computing the summary statistics in Section 4.

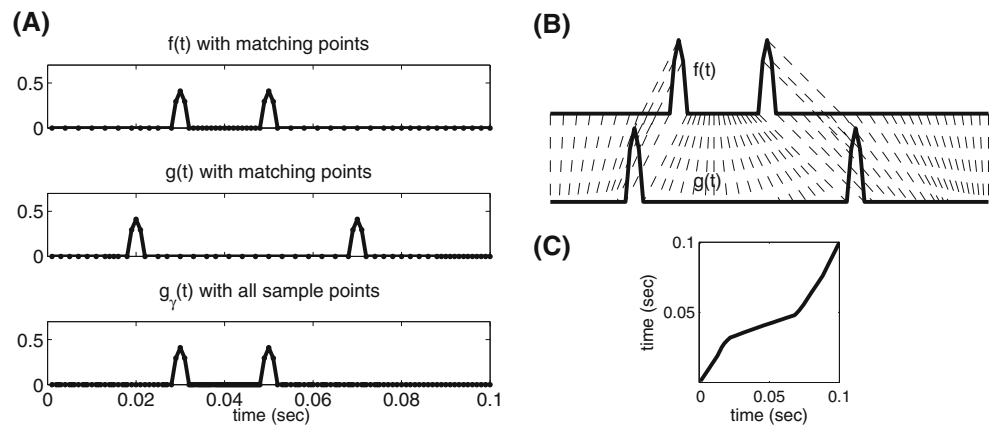
3.3 Examples of matching spike trains

Our goal in this study is to provide a systematical framework to measure the distances between spike trains. We have provided theoretical formulations as well as practical algorithms for the new metrics. Here we will illustrate them using a few simulated examples.

Metric distance between two single-spike trains We start with the computation of the D_2 metric between two single-spike trains $f(t)$ and $g(t)$, shown in Fig. 4. The total time is $T = 0.1$ s with bin size = 1 ms. The spike in $f(t)$ is at time 0.03 s and the spike in $g(t)$ is at time 0.07 s. The smoothing parameter $\sigma = 2$ ms, and each bump kernel is represented by five discrete points, $[-\sigma, -\sigma/2, 0, \sigma/2, \sigma]$. For balancing the cost in time warping and mismatching, we choose $\lambda = (1 + 1)/(2T) = 10$ in this example. Using the dynamic programming algorithm, the squared distance $D_2^2(f, g)$ is found to be 0.185. Figure 4(b) shows the optimal matching between the two spike trains and C shows the optimal γ^* . Using the same parameters, we found the metric $D_1(f, g)$ equals 0.8 and the resulting optimal warping (omitted) is nearly identical to the one shown in Fig. 4.

Let $\sigma = 0$, we can also compute the d_p distance between the original spike trains f and g . It is found that $d_2^2(f, g) = 0.167$. This is slightly different from the D_2 distance (0.185), which is due to the discretized computation in D_2 . We also find that $d_1(f, g) = 0.8$, which equals the D_1 distance.

Fig. 5 Matching between two double-spike trains when $\lambda = 100$. (a)–(c) Follow the same description as that in Fig. 4. Note that both spikes are matched between two trains



Metric distance between two double-spike trains To illustrate the mapping in multi-spike cases, we examine the D_2 metric distance between two double-spike trains $f(t)$ and $g(t)$ shown in Fig. 5. The total time is $T = 0.1$ s with bin size = 1 ms. The two spikes in $f(t)$ are at 0.03 and 0.05 s, respectively and the two spikes in $g(t)$ are at 0.02 and 0.07 s, respectively. The smoothing parameter $\sigma = 2$ ms and $\lambda = 5(2 + 2)/(2T) = 100$ was used in this example. Using the dynamic programming algorithm, the squared distance $D_2^2(f, g)$ is found to be 1.19. We see that the ISIs can be warped such that both spikes in $f(t)$ can match the spikes in $g(t)$. As there are total of 4 spikes in the two trains, the squared distance between them is at most 4. If we increase the weight λ to 400, we will not see matching for both spikes. This is because $D_2^2(f, g)$ in the full match will be $1.19 \times 4 > 4$. Indeed, Fig. 6 shows the matching when $\lambda = 400$. We see that only the first spike is matched, and the squared distance $D_2^2(f, g) = 2.74 < 4$.

The distance between the spike-trains can also be measured in metric D_1 . We let all parameter values be the same as that in Fig. 5 except that $\lambda = (2 + 2)/(2T) = 20$. We found the consistent result that both spikes are matched and the distance $D_1(f, g) = 1.2$.

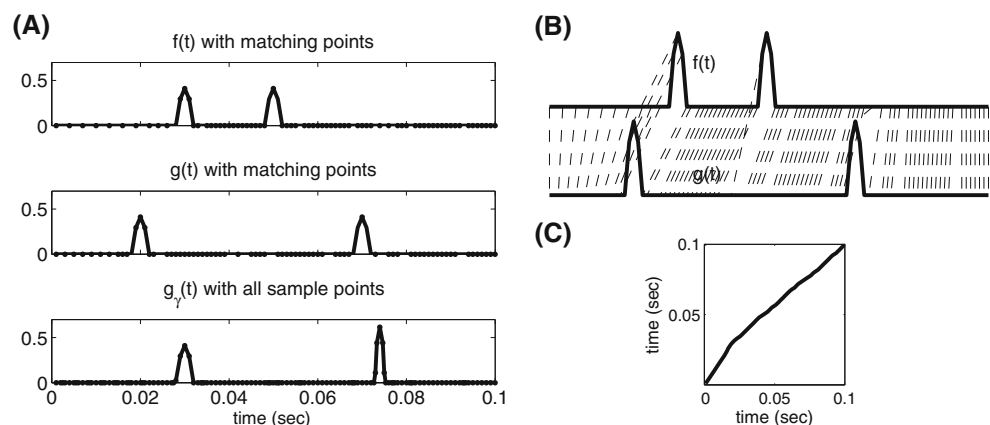
As there are totally four spikes in both trains, the distance $D_1(f, g) \leq 4$. If we increase the weight λ to 80, we will not see matching for both spikes. This is because $D_1(f, g)$ in the full match will be $1.2 \times 4 > 4$. Indeed, we find that when $\lambda = 80$, only the first spike is matched, and the distance $D_1(f, g) = 3.6 < 4$. The matchings in both cases are very similar to that in Figs. 5 and 6 and are omitted here.

We also compute the d_p distance for $\sigma = 0$. For $p = 2$, it is found that $d_2^2[100](f, g) = 1.03$ and $d_2^2[400](f, g) = 2.53$. These distances are also slightly different from the estimated D_2 distances. For $p = 1$, we have $d_1[20](f, g) = 1.2$ and $d_1[80](f, g) = 3.6$, which are exactly the same as those D_1 distances.

4 Summary statistics based on metric distances

So far we have introduced a family of metrics for quantifying differences between spike trains. Based on these metrics, we will explore the notion of descriptive statistics for observed spike trains. For example, we can ask questions like “What is the mean trend of a set of spike trains?” and “What is the dispersion from the

Fig. 6 Matching between two double-spike trains when $\lambda = 400$. (a)–(c) Follow the same description as that in Fig. 4. Note that only the first spike is matched between two trains



sample mean?” The answers to these questions are of fundamental importance in understanding the structure and statistical variability of spike trains. Note that if spike trains are smoothed, one can borrow tools from function spaces to address these problems. Applying tools from information geometry on the function space of smoothed spike train, we can define items such as mean, covariance, and principal component analysis. However, the estimated statistics, in general, may be smoothed, bounded functions and lose the special properties of the spike trains (i.e. a sequence of kernel functions).

Since our interest is in statistical summaries on the spike train space, we focus our attention on the more difficult case of computing statistics when $\sigma = 0$. In particular, using the *Euclidean distance* $d_2[\lambda]$, we will establish the notions of mean and variance, respectively, of spike trains and will demonstrate them using examples. We will at first illustrate the mean of a set of spike trains in \mathcal{S}_M to help provide some intuition. Then we will have some in-depth discussion on the mean spike trains in the general domain of \mathcal{S} .

We note that the proposed metrics depend on the penalty parameter λ . If the value of λ is large, any time warping on ISIs would be discouraged. Such constraint may make the statistics difficult to estimate. If we let λ be sufficiently small, then the framework becomes fully elastic and spikes can move relatively freely in the time domain and this leads to a more intuitive distance and efficient estimations. Here we assume that $\lambda < 1/(2NT)$ for all the following statistics computation, where N is the number of spike trains in a given set. The solution we provide below is under this assumption.

4.1 Mean spike train in \mathcal{S}_M under \mathbf{d}_2

In standard descriptive statistics, the mean of a set of sample points in a Euclidean space is defined as the point whose elements are averages of the corresponding elements in the given set. It is also the point that has the minimum sum of squared Euclidean distances to all points in the set. For spike trains $S_1, \dots, S_N \in \mathcal{S}_M$ (all trains with M spikes), their sample *mean* can be analogously defined using the classical Karcher mean (Karcher 1977) as follows:

$$C^* = \arg \min_{C \in \mathcal{S}_M} \sum_{k=1}^N d_2(S_k, C)^2 \quad (9)$$

To explore an analytical solution to this minimization problem, each train S_k can be equivalently represented by an ISI vector $(s_{k,1}, \dots, s_{k,M+1})$ with $\sum_{i=1}^{M+1} s_{k,i} = T$. The mean spike train C is also as-

sumed to have M spikes (this assumption will be verified in Section 4.2) and its ISI vector is denoted as (c_1, \dots, c_{M+1}) . If $\lambda < 1/(2NT) \leq 1/(2T)$, then the condition in Theorem 1 in Appendix D is satisfied for $p = 2$. Therefore, the mean spike train in Eq. (9) can be computed as follows:

$$C^* = \arg \min_{c_1 + \dots + c_{M+1} = T} \sum_{k=1}^N \sum_{j=1}^{M+1} (\sqrt{s_{k,j}} - \sqrt{c_k})^2. \quad (10)$$

Equation (10) is a constrained optimization problem with a unique closed-form solution. Using the Lagrange multiplier technique, we obtain the solution:

$$c_j^* = \frac{\left(\sum_{k=1}^N \sqrt{s_{k,j}}\right)^2}{\sum_{j=1}^{M+1} \left(\sum_{k=1}^N \sqrt{s_{k,j}}\right)^2} T, \quad j = 1, \dots, M+1. \quad (11)$$

With the estimated sample mean C^* , the squared sum $V^* = \frac{1}{N} \sum_{k=1}^N d_2(S_k, C^*)^2$ is often defined as the *variance* for the spike trains S_1, \dots, S_N . Basically, it describes the dispersion from the sample mean.

As a simple example, assume that $T = 1$ s, $N = 2$, and we want to find the mean of two ISI vectors (0.14, 0.52, 0.34) s and (0.42, 0.36, 0.22) s, respectively. Using Eq. (11), the ISI vector of the mean spike train is (0.268, 0.448, 0.284) s. The *variance* with this mean is 2.58×10^{-2} s². This *variance* is minimal in the sense that it is less than or equal to any squared sum computed using any 3-tuple ISI vector. For example, if we simply take the element-wise average of the two ISI vectors, then their mean will be (0.28, 0.44, 0.28) s and the squared sum for this vector will be 2.60×10^{-2} s².

4.2 Mean spike train in \mathcal{S} under \mathbf{d}_2

Now we consider the mean computation for a set of spike trains $S_1, S_2, \dots, S_N \in \mathcal{S}$ where the number of spikes in each train is an arbitrary non-negative integer. The definition of their mean spike train extends that in Eq. (9):

$$S = \arg \min_{C \in \mathcal{S}} \sum_{k=1}^N d_2(S_k, C)^2 \quad (12)$$

Our algorithm for finding the mean is based on the previous solution (Eq. (11)) and the assumption that $\lambda < 1/(2NT)$. Note that to standardize the end points, here we have inserted spikes at both time ends, i.e. at times 0 and T . Therefore, if the number of spikes is n_k in the k th train, then the number of ISIs in this train is $n_k + 1, k = 1, \dots, N$.

For the mean spike train S , there are two unknowns: the number of spikes n and the placements of these

spikes in $[0, T]$. We start with the first unknown. According to the definition of the mean, we need to minimize the cost function $\sum_{k=1}^N d_2(S_k, S)^2$. When $\lambda < 1/(2NT) \leq 1/(2T)$, using Theorem 2, this cost can be computed in the following form:

$$\sum_{k=1}^N [|n_k - n| + \lambda \cdot (\text{optimal warping cost in the associated ISIs})]. \quad (13)$$

As the warping cost in each match is up to $2T$ (see Appendix C), the total penalty in the cost function is upper bounded by $2\lambda NT$. Therefore, when $2\lambda NT < 1$, or $\lambda < 1/(2NT)$, the cost function is dominated by the sum on $|n_k - n|$. The problem then reduces to solving for the minimum of:

$$\sum_{k=1}^N |n_k - n|.$$

Therefore, the optimal number of spikes for S is the median of (n_1, n_2, \dots, n_N) . In case the median is not unique (this could happen when N is even and the two middle values are not the same), we can use any integer between the two middle values, and then break the tie by incorporating the time warping cost in each case.

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

1. Set initial times for the n spikes in $[0, T]$ to form an initial S .
2. **Matching Step:** Using the dynamic programming for $d_p[\lambda]$, find the optimal time matching between S and S_k , $k = 1, \dots, N$. There are two cases:
 - Case 1: $n_k \geq n$,
 - (a) Identify n spikes in S_k that are optimally matched to the n spikes in S .
 - (b) Compute the $n + 1$ ISIs in S_k using these n spikes.
 - Case 2: $n_k < n$,
 - (a) Identify the n_k spikes in S that are optimally matched to the n_k spikes in S_k .
 - (b) Based on the n_k matched pairs of spikes (between S_k and S), compute the spike time in S_k as a function of spike time in S via linear interpolation. Insert $n - n_k$ (imaginary) spikes in S_k using the interpolated values computed on the other $n - n_k$ spike times in S .

- (c) Compute the $n + 1$ ISIs in S_k using the n spikes (n_k real and $n - n_k$ imaginary).

3. **Minimization Step:** Compute the optimal $(n + 1)$ ISIs in S based on Eq. (11). These ISIs minimize the sum of squared distance for the given matching. Then update the mean spike train with these ISIs.
4. If the sum of squared distance stabilizes over steps 2 and 3, then the mean spike train is the current estimate and stop the process. Otherwise, go back to step 2.

Denote the estimated mean in the i th iteration as $S^{(i)}$. We can show that the sum of squared distance $\sum_{k=1}^N d_2^2(S_k, S^{(i)})$ decreases iteratively. The detail is described in Appendix E. As 0 is a natural lower bound, $\sum_{k=1}^N d_2^2(S_k, S^{(i)})$ will always converge when i goes to infinity. In practical applications, we find the process only takes a few iterations to reach a reasonable convergence of the sum of squared distance and the mean spike train. However, note that there may be multiple local minima in Eq. (12), and more work will be needed to investigate that issue in the future.

5 Experimental results

In previous sections, we have provided theoretical formulations as well as practical algorithms on the new metrics and the mean statistics. Here we will provide more detailed analysis when these methods are applied to simulated and real experimental data. Note that the estimation of mean spike trains is entirely data-driven, which does not depend on any specific model on spike train processes. For comparison with the underlying density functions, we will investigate the mean statistics in classical Poisson processes. For practical use, we will apply both the metrics and statistics in classification problems in one simulated dataset and one real experimental recording from primate motor cortex. Focusing on applications on original spike trains, we let the smoothing parameter $\sigma = 0$ in all examples. All computations are done using Matlab (Mathworks Inc.) on a laptop computer with a 2.00GHz Intel(R) CPU.

5.1 Mean statistics in Poisson processes

Homogeneous Poisson process At first, we test the MM-algorithm on spike trains generated from a homogeneous Poisson process. When λ is sufficiently small ($\leq 1/(2NT)$ in Eq. (13)), the estimation process is independent of the specific value of λ . Therefore, we here only assume λ is small, but do not give any specific value. Let total time $T = 1$ (s), Poisson rate

$\rho = 10$ (spikes/s) and randomly generate 20 spike trains from this process. The individual spike trains are shown in Fig. 7(a) (upper panel). The numbers of spikes in these trains vary from 5 to 15, with the median value of 10. Therefore, n , the number of spikes in the mean, is set to 10. To initialize the algorithm, we randomly generate ten spikes in $[0, T]$ and used the iterative process to update the estimate. The initial and the convergent train are also shown in Fig. 7(a) (middle panel and lower panel). We can see 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. Interestingly, such evenness is hardly visible in any of the individual samples.

The estimation of the mean usually takes approximately 5–10 iterations. The evolution of the time warping cost versus the iteration index is shown in Fig. 7(b). We see that the cost function decreases rapidly and converges to a minimum. We also show the mean spike train during the estimation process in Fig. 7(c). The first row (0th) is the random initial, and the last row (9th) is the final estimate. Significant changes are observed

in the first few (3–5) iterations, and then the process stabilizes.

To further demonstrate the mean estimation, we randomly generate another 200 spike trains with the same process. The mean of these trains are shown in Fig. 7(d). We see that with a larger sample size, the variability gets lower and all spikes in the mean are more evenly distributed.

Inhomogeneous Poisson process Next we test our method on spike trains generated from an inhomogeneous Poisson process. We still have total time $T = 1$ (s), but the Poisson rate (in the units of spikes/s) varies over time with the function

$$\rho(t) = 4 \sin(2\pi t) + 20, \quad 0 \leq t \leq T$$

This function is shown in Fig. 8(a) (first panel). We randomly generate 15 spike trains from this process. The data are shown in Fig. 8(a) (second panel). The numbers of spikes in these trains vary from 16 to 29, with the median value of 20 implying that n is 20. We randomly generate 20 spikes in $[0, T]$ to initialize the algorithm and implement the iterations. The initial and

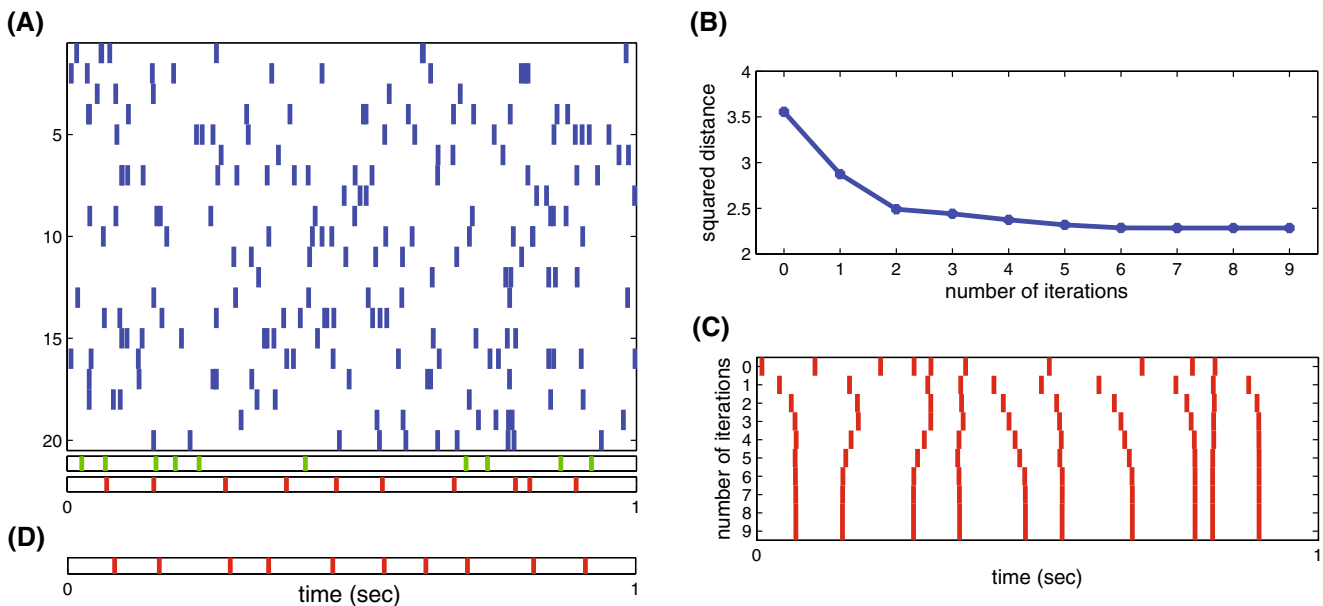


Fig. 7 (a) *Upper panel*: 20 spike trains generated from a homogeneous Poisson process. Each *blue vertical line* indicates a spike. *Middle panel*: A (uniformly) random initial with ten spikes. Each *green vertical line* indicates a spike in the initial. *Lower panel*: The mean spike train estimated using the proposed algorithm. Each *red vertical line* indicates a spike in the estimated mean. We can see that the mean is approximately evenly distributed which reasonably captures the homogeneous nature of the process. (b)

The sum of squared warping distance over all iterations. (c) 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 sum of squared distance decreases over the iterations and the mean spike trains converges very fast. (d) The mean of another randomly generated 200 spike trains

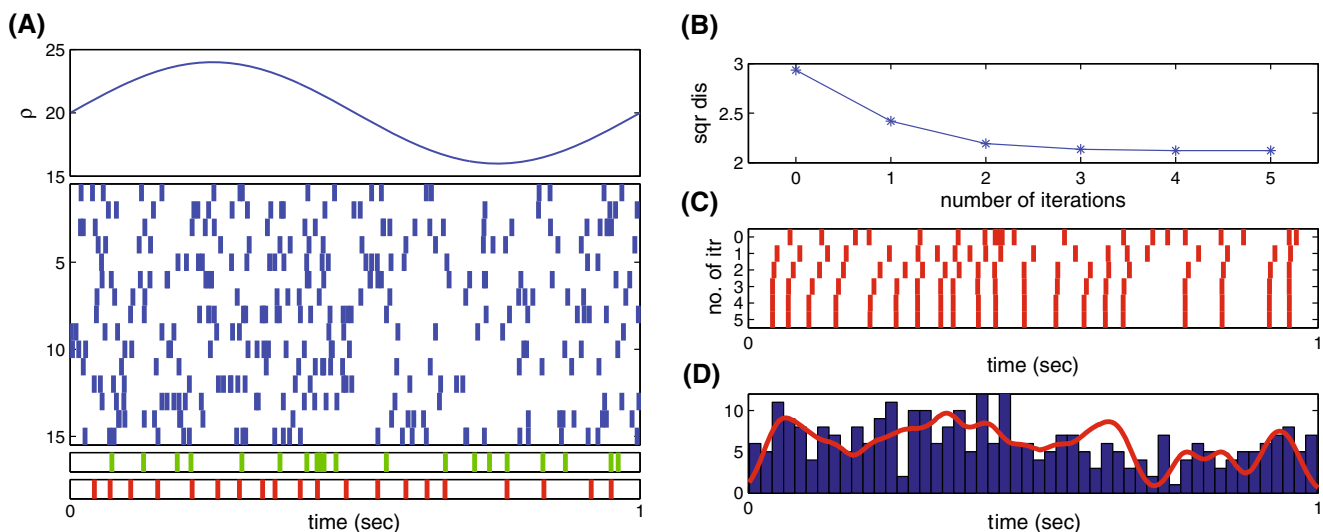


Fig. 8 (a) First panel: Poisson rate $\rho(t)$ varies on $[0, T]$. Second panel: 15 spike trains generated from the Poisson process with rate $\rho(t)$. Each blue vertical line indicates a spike. Third panel: A random initial of mean with 20 spikes. Each green vertical line indicates a spike. Fourth panel: The mean spike train estimated using the proposed algorithm. Each red vertical line indicates a spike. We can see the mean is distributed approximately fol-

lowing the Poisson rate (more spikes when $\rho(t)$ is large, and fewer spikes when $\rho(t)$ is small). (b) The sum of squared warping distance over all iterations. (c) The estimated mean spike train over all iterations. (d) PSTH of the 15 spike trains with binsize 20 ms and a smoothed version (using a Gaussian kernel) of the estimated mean spike train

the final convergent trains are also shown in Fig. 8(a) (third and fourth panels). We can see the mean approximately follows the Poisson rate $\rho(t)$ (dense when the rate is large, and sparse when the rate is small). This distribution well captures the inhomogeneous nature of the process.

The estimation of the mean in this case takes only five iterations and the evolution of the cost function is shown in Fig. 8(b). Similar to the homogeneous case, the cost decreases rapidly and converges to a constant after a few iterations. We also show the estimated mean spike train in Fig. 8(c). Most of the changes are observed in the a few (around 3) iterations, and then the process stabilizes.

The mean spike train can also be related to the PSTH (peri-stimulus time histogram) of the observed spike trains. We compare them here and the result is shown in Fig. 8(d). The bars are the PSTH of the 15 spike trains with binsize 20 ms, and the red line is a smoothed function by the estimated mean spike train and a Gaussian kernel function. We see that the smoothed spike train approximates the histogram. Therefore, we can look the mean as a representation of the density in the spike train space. One advantage of the mean spike train over the standard PSTH method is that the mean spike train is still in the continuous time domain, whereas the PSTH is a histogram which needs to be represented in a discretized time domain for some appropriate bin size. Note that PSTH can be converted into continuous

time process using some kernel functions, but one still needs to address the tricky choice of the bandwidth parameter.

5.2 Classification in simulated spike trains

We have illustrated the proposed metrics using single- and double-spike trains. Here we extend this investigation to a more general case where spiking activity is randomly generated from stimulus-dependent point processes. In particular, we simulate spike trains of neurons in the arm area of primary motor cortex when the hand moves from point *S* to point *E* with four equal-distance paths on the horizontal plane with a constant speed (see Fig. 9(a)). Assume the movement time $T = 2$ s, the x- and y-position in each path follows the equations:

$$\text{Path 1: } x(t) = -\cos(0.5\pi t), y(t) = \sin(0.5\pi t)$$

$$\text{Path 2: } x(t) = -\cos(0.5\pi t), y(t) = -\sin(0.5\pi t)$$

$$\text{Path 3: } x(t) = 0.5(\cos(\pi t) + 1)(1_{t \geq 1} - 1_{t < 1}), \\ y(t) = 0.5 \sin(\pi t)$$

$$\text{Path 4: } x(t) = 0.5(\cos(\pi t) + 1)(1_{t \geq 1} - 1_{t < 1}), \\ y(t) = -0.5 \sin(\pi t), 0 \leq t \leq T$$

Spiking activity is simulated using a generalized linear model (GLM) with Poisson density function where the

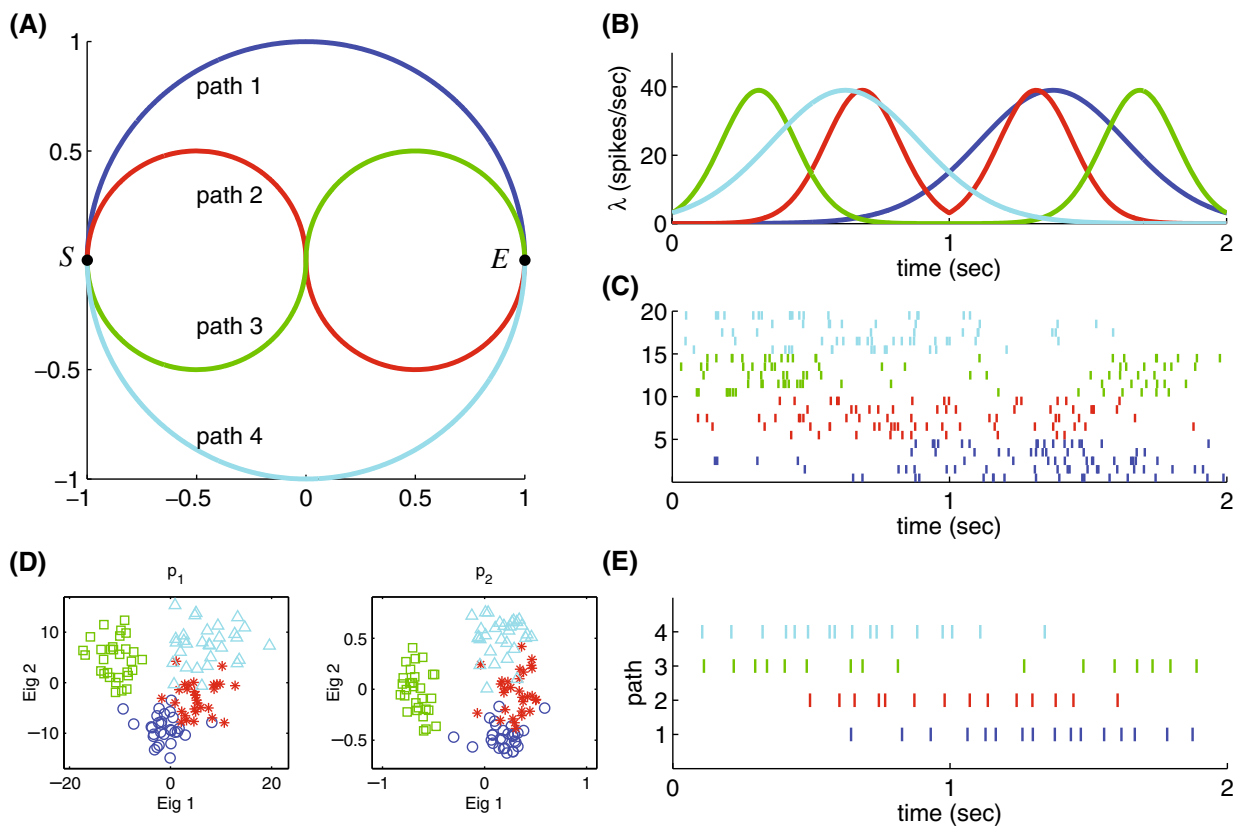


Fig. 9 (a) The four movement paths from S to E , colored by blue, red, green, and cyan, respectively. (b) Poisson density functions of these paths with same colors in (a). (c) Five spike trains generated from each of the density functions. Each thin vertical line indicates the time of a spike, and one row is for one trial. The colors are the same as that in (a). (d) The 2-dimensional points associated with the two largest eigenvalues based on metrics d_1

(left panel) and d_2 (right panel) in the MDS analysis. Blue circles indicate spike trains for path 1, red stars are for path 2, green squares are for path 3, and cyan triangles are for path 4. (e) The mean spike trains of the four paths in the training data. 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)

rate is given in the following form (Brockwell et al. 2004; Truccolo et al. 2005):

$$\rho(t) = e^{a_0 + a_1 x(t) + a_2 y(t) + a_3 \dot{x}(t) + a_4 \dot{y}(t)}.$$

For illustration, we take an example case where the density is a function of velocity with coefficients $a_0 = a_1 = a_2 = 0$, $a_3 = 1.5$, $a_4 = 1$ (note that the proposed method does not depend on the model form and any parameter values can be substituted here). The density function of each path is shown in Fig. 9(b). We see that the four density functions have different forms; two are unimodal, the other two are bimodal, and all peaks are at different times. 50 spike trains are randomly generated with each density function. For illustration, five spike trains from each movement path are shown in Fig. 9(c).

We apply both d_1 and d_2 metrics to measure the distance between each pair of spike trains. Let E denote the average number of spikes in all four paths; In our experiment, we find $E = 15.98$. Then we let $\lambda =$

$3(E + E)/(2T) = 23.97$ in metric d_1 and $\lambda = 10(E + E)/(2T) = 79.90$ in metric d_2 . Within the 50 trials for each path, we choose 30 as training data, and the rest 20 as testing data.

With 120 training spike trains over four paths, there are $120 \times 119/2 = 7,140$ distances. As a summary, we found that the average distances within the same paths are $18.0(\pm 3.0)$ and $1.90(\pm 0.10)$ for d_1 and d_2 , respectively, and between different paths are $23.7(\pm 4.8)$ and $2.07(\pm 0.13)$, respectively. This indicates that the metrics appropriately address the dissimilarity between spike trains. To visually look at the clustering based on the pairwise distances, we use a multidimensional scaling (MDS) technique (Seber 2004) to show the 2-dimensional points associated with the two largest eigenvalues (see Fig. 9(d)). The plots indicate that the spike trains can be approximately partitioned into four clusters via either the d_1 or d_2 metric, though a small number of spike trains are around the borderlines among clusters.

The 80 testing spike trains are used to evaluate the classification performance. For each test train, we compute its average distance to each of the training path; that is, compute the distance to the 30 training trains, and then take the average. The classification is based on the shortest averaged distance to the four path categories. It is found that when the distance is d_1 , the accuracy is 75/80, and when d_2 is used, the accuracy is around the same at 74/80. This result indicates that both d_1 and d_2 metrics appropriately quantify distance between spike trains and can be used in classification of spiking activity with respect to different stimuli/behaviors.

The classification above needs to compute the pairwise distance between 80 testing trains and 120 training trains ($80 \times 120 = 9,600$ distances). Let M and N denote the number of trains in the training and testing set, respectively. Then the computational cost is $O(MN)$. This cost may be expensive when M and N are large. Here we will lower this cost by using mean spike trains. Based on the proposed MM-algorithm, we compute the mean spike train for each path in the training data (Fig. 9(e)). 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 use these mean spike trains for classification where each test train is classified by the shortest distance over the four distances to the four means in the training set. In this case, we need to compute $80 \times 4 = 320$ distances (in the order of $O(N)$ only). It is found that the accuracy of this mean-distance classification is 73/80, about the same as that based on pairwise-distances. As a summary, we show the classification accuracy and efficiency of each method in Table 1. We see all methods have nearly the same accuracy, whereas the mean-distance method takes only about 1/30 of time of the other two methods.

5.3 Experimental recording from motor cortex

In addition to the simulation examples shown above, we here demonstrate the use of the metrics and mean spike trains in a real experimental recording. The motor

behavior is similar to an open squared-path movement that we investigated recently (Chi et al. 2007; Wu and Srivastava 2011). Briefly, silicon microelectrode arrays containing 100 platinized-tip electrodes were 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., Foxboro, MA). Single units were manually extracted using Offline Sorter (Plexon Inc., Dallas, TX). The bin size is chosen as 10 ms for the firing rates. Approximately 99.93% of bins have at most one spike, and therefore this rate sequence can be approximately looked as a binary spike train sequence. The monkey 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 (Fig. 10(a)). At the 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. That is, in each trial the subject reached a sequence of five targets which were the four corners of the square with the first and last ones overlapped. 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.

To illustrate the new metrics in measuring distances between spike trains, we selectively choose spiking activity from one neuron which show significant tuning property over four different paths. 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 seconds. For illustration, six spike trains from each movement path as well as the reaching times at the five corners are shown in Fig. 10(b).

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. We at first perform the classical pairwise-distance classification with the d_1 and d_2 metrics. Let E denote the average number of spikes in all four paths, and we find $E = 32.6$. Then we let $\lambda = 3(E + E)/(2T) = 19.6$ in metric d_1 and $\lambda = 10(E + E)/(2T) = 65.2$ in metric d_2 . For each test trial, we compute its distance to all $30 \times 4 = 120$ training trials. As a summary, we found that the average distances within the same paths are $40.2(\pm 6.1)$ and $5.53(\pm 0.49)$ for d_1 and d_2 , respectively,

Table 1 Classification accuracy and efficiency (total computational time) for three methods: pairwise-distance using d_1 , pairwise-distance using d_2 , mean-distance based on d_2

Method	Accuracy (%)	Efficiency (s)
Pairwise, $d_1 (= D^{\text{interval}})$	93.75 (75/80)	747
Pairwise, d_2	92.50 (74/80)	763
Mean, d_2	91.25 (73/80)	25

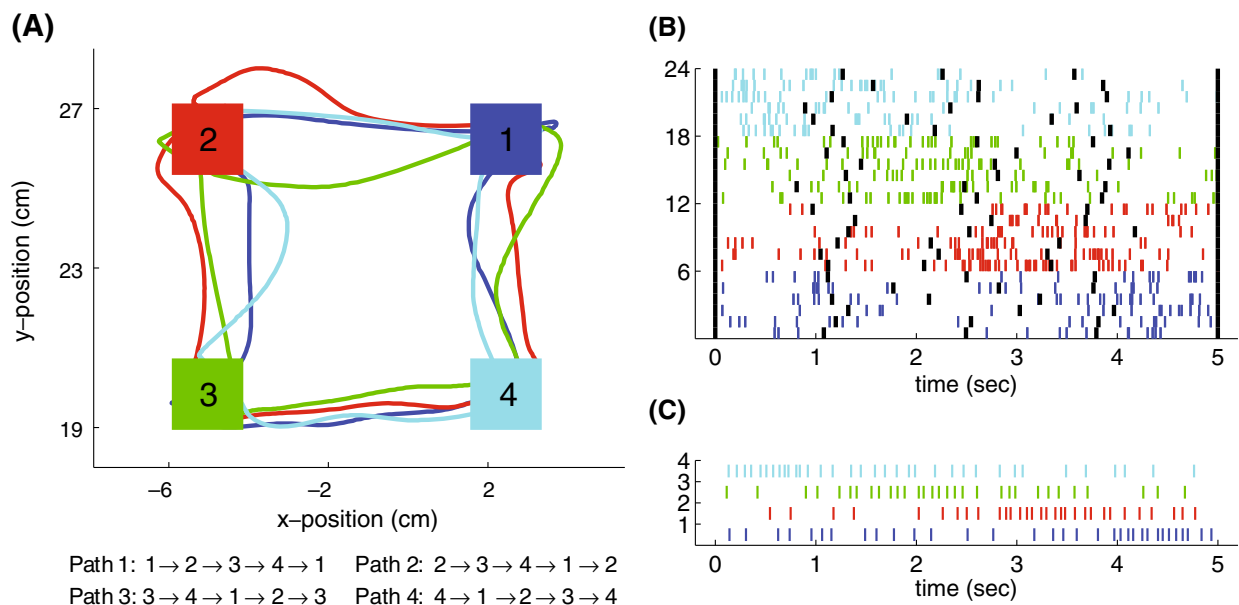


Fig. 10 (a) Four trajectories of hand movement in the SP task. The four colors (blue, red, green, and cyan) indicate the trajectories started at corners 1, 2, 3, and 4, respectively, where the corners are also shown in the correspond colors. The four paths are also described with text legend under the plot. (b) Six spike trains randomly chosen from each path. Each thin vertical line

indicates the time of a spike, where the colors are the same as that in (a). Each black, thick vertical line is the time of reaching a corner. One row is for one trial. (c) The mean spike trains of the four paths in the training data. 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)

and between different paths are $47.0(\pm 7.2)$ and $6.14(\pm 0.56)$, respectively. This indicates that the metrics appropriately address the dissimilarity between spike trains. Same as in the simulated data, the classification is conducted based on the shortest averaged distance to the four path categories. It is found that when the distance is d_1 , the accuracy is 89.2% (107/120), and when d_2 is used, the accuracy is 90.0% (108/120). This result further indicates that both d_1 and d_2 metrics appropriately quantify distances between spike trains and can be used for classification of spike trains.

As shown in the simulated data, the computation can be greatly reduced when a mean-distance is used. Here we also compute the mean spike train in each path using the proposed MM-algorithm. The result is shown Fig. 10(c). Comparing with the original spike trains, the mean spike trains appropriately represent the firing patterns in the corresponding movement. For example, the spiking frequency is higher when the hand moves upward, which is apparent in all four means. 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 means in the training set. This is significantly more efficient than the 120 distances in the pairwise case. It is found that the accuracy of this mean-distance classification is 87.5% (105/120), about the same as that based on pairwise-distances.

Finally, we note that the correlation coefficient between all d_2 and d_1 distances is 0.94. This suggests both metrics have somewhat equivalent performance when measuring the spike distances. Intuitively, one may look this equivalence as that between an L^2 norm and an L^1 norm in the functional space (i.e. “Euclidean distance” vs. “Manhattan distance”).

6 Discussion and future work

Motivated by the flexibility of Fisher–Rao metric in information geometry in handling time warping, we have developed a parametrized family of metrics, D_p , $p \geq 1$, that takes into account different time warping of spike trains. When the smoothing parameter σ equals 0, these metrics are also represented as d_p , $p \geq 1$. Using the “Euclidean distance” d_2 , we define the mean (and variance) of a set of spike trains. In particular, we propose an iterative numerical procedure, the MM-algorithm, for the mean computation for spike trains with arbitrary numbers of spikes. There are many uses of this statistic, two of which are demonstrated here: (1) the computation of a template or a representative spike train for a neuron under a fixed stimulus/behavior, and (2) to classify spike trains efficiently by comparing them to the mean templates for different behaviors. In our

future work, these building blocks will be used to derive various classical statistical analyses such as hypothesis tests, confidence intervals, regressions, FANOVA (functional ANOVA), FPCA (functional PCA), and regressions and bootstraps on functional data. For example, we can decide whether a new spike train belongs to a set of spike trains in a statistically principled fashion by using a confidence interval or a hypothesis test (with estimated mean and variance).

We point out that the mean spike train can be defined for any of the previous metrics, in principle. Computing the mean, however, may be more complicated in some situations than others. Since our metrics are simply penalized L^p metrics, the average value corresponds to more conventional statistics, such as the Euclidean mean and variance. Based on the theoretical closed-form result in optimal time warping (Theorems 1 and 2 in Appendix D), we are able to derive an efficient, and theoretically-proven convergent, MM-algorithm for computing mean of a set of spike trains with arbitrary numbers of spikes. Note that the mean spike train estimation is under the assumption that λ is sufficiently small and $\sigma = 0$. If we allow the full space of bounded, non-negative continuous functions, the constraints on σ and λ can be relaxed. In that case, the statistics and paths can be estimated via state-of-the-art numerical methods in information geometry.

We would like to emphasize that the basic idea in this paper is to build a data-driven statistical inference system. It is based on treating each spike train as a single point in the spike train space, and we propose this as an alternative to (point process) model-based inferences. Although we have not provided a comprehensive comparison with the model-based methods, some quick comparisons are as follows: (1) The mean computation in the new framework is model free. In practice, this could be an advantage when the commonly-used models (e.g. Poisson processes or other point processes) can not fit the data well or they involve complicated estimations; (2) When the firing rate is high, the smoothed mean spike train may approximate the PSTH or the underlying mean firing rate function; (3) The mean firing rate function in the model-based methods is often based on Poisson assumption, but this assumption will not be satisfied in history-dependent spiking activity (commonly observed in real data); (4) The standard mean firing rate function (under a Poisson model) is also metric-based where the metric is implicitly Euclidean in a vector space; and (5) This new framework incorporates a measure of elasticity between spikes, whereas the firing rate methods do not.

The new metrics quantify distances for both unsmoothed (when $\sigma = 0$) and smoothed (when $\sigma > 0$) versions of the spike trains. The metrics on unsmoothed spike trains can be more efficiently computed, and have direct use in defining summary statistics such as means and variances. In contrast, spike trains are often modeled as point processes based on continuous density functions. A smoothing procedure may better represent the underlying functions. In practical use, a choice of “optimal” σ may depend on specific needs and/or research goals.

Based on the proposed d_2 and d_1 metrics, we have introduced the notions of mean and variance for a set of spike trains. To the best of our knowledge these ideas have not been discussed in the literature before. These quantities provide more detailed information about the statistical distribution of the underlying spike train space than merely using pairwise distances. Note that the variance defined in the paper is a scalar value, which represents the total variation and can be looked as the trace of the underlying covariance matrix. The definition of the full covariance matrix has been left out for future.

The proposed metric, D_p ($p \geq 1$), is similar to a penalized L^p norm in functional analysis and it provides a generalization of some previous metrics. For example, it generalizes the Victor-Purpura D^{interval} metric when $p = 1$ and the van Rossum metric when $p = 2$. Another form of L^p norm was recently developed in a point process framework (Dubbs et al. 2009) where the method extends the l^1 norm in Victor-Purpura metric to an l^p norm and minimizes the cost based on bipartite graph matching theory. This particular metric does not have any direct relation to the van Rossum metric even when p equals 2.

We wish to point out that the proposed framework for analyzing spike trains is not a Riemannian analysis on the space of spike trains. Our framework is purely distance based in the sense that the basic element here is a distance between any two spike trains. We have not defined a Riemannian metric for spike trains although it would be natural to use the Riemannian framework to reach some of the same goals (sample statistics, optimal paths, etc). The difficulty lies in that fact that of all L^p spaces, only L^2 is a Hilbert space, i.e. has an inner-product, while the others are only Banach spaces. We conjecture that for $p = 2$ one can particularize our framework into a Riemannian framework with full inner products, geodesics, and geodesic-based statistics, it does not seem possible for other values of p . Once we have a Hilbert manifold, other extensions, including the

use of reproducing Kernel Hilbert spaces (Paiva et al. 2009b), can be applied to spike train data.

This study has focused on spike train statistics for a given sample. It would be equally important to investigate the statistics for the underlying population. Assuming that all spike trains in \mathcal{S} follow a probability distribution with density function f , the average of squared distance from each spike train S_k in the sample to any spike train C will have the following limiting form (when $N \rightarrow \infty$):

$$\frac{1}{N} \sum_{k=1}^N d_2(S_k, C)^2 \rightarrow \int d_2(S, C)^2 f(S) dS.$$

This is because $d_2(S_k, C)^2$ is a real-valued function, and the sample mean on the left converges to its mean value using the standard law of large numbers (as long as the variance is finite). Therefore, the Karcher sample mean will have the following limit,

$$\arg \min_{C \in \mathcal{S}} \sum_{k=1}^N d_2(S_k, C)^2 \rightarrow \arg \min_{C \in \mathcal{S}} \int d_2(S, C)^2 f(S) dS,$$

where the RHS can be defined as the population mean for the distribution $f(S)$. In our future work, we will investigate a formal definition as well as theoretical properties of this population mean.

Though the new metrics are defined for any $p \geq 1$, the study has been focused on when p equals 1 and 2. This is essentially in that these two metrics (1) have intuitive correspondences (“Manhattan distance” and “Euclidean distance”); (2) generalize previous metrics; and (3) are natural choices to build summary statistics (such as mean and median) for spike trains. Our future work will explore the properties for distances with other p 's. In particular, the current focus has been on a well-defined metric and we have limited $p \geq 1$ to ensure the triangle-inequality. The case when $p < 1$ would have certain interesting properties and may be of great importance to pursue. We will also systematically compare the performance of these metrics with established distances (Paiva et al. 2009a).

Finally, so far the proposed metrics are only for distance between two unlabeled spike trains (single-unit case). We will investigate its natural extensions to the multineuronal case (e.g. using a Hausdorff-type distance involving sets of spike trains) which would be more desirable in neural decoding as well as other practical problems (Aronov et al. 2003; Victor et al. 2007; Houghton and Sen 2008).

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 for providing experimental data. We also thank Prof. Eric Klassen for helpful discussions.

Appendix A: Metric $D_p[\sigma, \lambda]$

Here we need to show the distance $D_p[\sigma, \lambda]$ satisfies the following three conditions in the definition of a metric space, where $p \geq 1$, $\sigma > 0$, and $\lambda > 0$.

- (a) Non-negativity: $D(f, g) \geq 0$ for any $f, g \in \mathcal{S}^{(\sigma)}$. The equality holds if and only if $f = g$.

Proof The non-negativity is apparent. \square

- (b) Symmetry: $D(f, g) = D(g, f)$ for any $f, g \in \mathcal{S}^{(\sigma)}$.

Proof For any $\gamma(t) \in \Gamma$, let $s = \gamma(t)$. Then $t = \gamma^{-1}(s)$, and $s = \gamma(\gamma^{-1}(s))$, $1 = \dot{\gamma}(\gamma^{-1}(s))\dot{\gamma}^{-1}(s)$. Therefore, $\int |f(t)^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p}|^p + \lambda |1 - \dot{\gamma}(t)^{1/p}|^p dt = \int |[f(\gamma^{-1}(s))\dot{\gamma}^{-1}(s)]^{1/p} - g(s)^{1/p}|^p + \lambda |\dot{\gamma}^{-1}(s)^{1/p} - 1|^p ds$. Taking infimum over $\gamma(t)$ and $\gamma^{-1}(s)$ on both sides, the symmetry holds. \square

- (c) Triangle-inequality: $D(f, g) \leq D(f, h) + D(h, g)$ for any $f, g, h \in \mathcal{S}^{(\sigma)}$.

Proof The triangle-inequality will hold if for any $\gamma(t), \gamma_1(t) \in \Gamma$ we have

$$\begin{aligned} & \left[\int |f(t)^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p}|^p + \lambda |1 - \dot{\gamma}(t)^{1/p}|^p dt \right]^{1/p} \\ & \leq \left[\int |f(t)^{1/p} - [h(\gamma_1(t))\dot{\gamma}_1(t)]^{1/p}|^p \right. \\ & \quad \left. + \lambda |1 - \dot{\gamma}_1(t)^{1/p}|^p dt \right]^{1/p} \\ & \quad + \left[\int |h(s)^{1/p} - [g(\gamma \circ \gamma_1^{-1}(s))\dot{\gamma} \circ \dot{\gamma}_1^{-1}(s)]^{1/p}|^p \right. \\ & \quad \left. + \lambda |1 - \dot{\gamma} \circ \dot{\gamma}_1^{-1}(s)^{1/p}|^p ds \right]^{1/p} \end{aligned} \quad (14)$$

where \circ denotes function composition in Γ (and therefore $\gamma \circ \gamma_1^{-1}$ is still a warping in the group Γ). Let

$s = \gamma_1(t)$, the second term on the right hand side of Eq. (14) is equal to

$$\begin{aligned} & \left[\int \left| [h(\gamma_1(t))\dot{\gamma}_1(t)]^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p} \right|^p \right. \\ & \quad \left. + \lambda |\dot{\gamma}_1(t)^{1/p} - \dot{\gamma}(t)^{1/p}|^p dt \right]^{1/p} \\ & \text{Therefore, Eq. (14) will hold if we can show that for any } a, b, c, d, e, \text{ and } f \text{ in the } L^p \text{ space} \\ & \left[\int |a - b|^p + |d - e|^p dt \right]^{1/p} \\ & \leq \left[\int |a - c|^p + |d - f|^p dt \right]^{1/p} \\ & \quad + \left[\int |c - b|^p + |f - e|^p dt \right]^{1/p} \end{aligned} \quad (15)$$

Indeed, Eq. (15) is a direct result from the triangle inequalities in the L^p space and the L^p norm in the Euclidean \mathfrak{R}^2 vector space. \square

Appendix B: Metric $d_p[\lambda]$

(B.1) At first, we show the distance $d_p[\lambda]$ also satisfies the following three conditions in the definition of a metric space, where $p \geq 1$ and $\lambda > 0$.

- (a) Non-negativity: $d(f, g) \geq 0$ for any $f, g \in \mathcal{S}$. The equality holds if and only if $f = g$.

Proof The non-negativity is apparent. \square

- (b) Symmetry: $d(f, g) = d(g, f)$ for any $f, g \in \mathcal{S}$.

Proof Assume $f \in \mathcal{S}_M$ and $g \in \mathcal{S}_N$. For any $\gamma(t) \in \Gamma$, let $s = \gamma(t)$. Then $t = \gamma^{-1}(s)$, and $s = \gamma(\gamma^{-1}(s))$, $1 = \dot{\gamma}(\gamma^{-1}(s))\dot{\gamma}^{-1}(s)$. Therefore, $M + N - 2 \sum_{i=1}^M \sum_{j=1}^N \mathbf{1}_{t_i=\gamma(s_j)} + \lambda \int |1 - \dot{\gamma}(t)^{1/p}|^p dt = N + M - 2 \sum_{j=1}^N \sum_{i=1}^M \mathbf{1}_{s_j=\gamma^{-1}(t_i)} + \lambda \int |\dot{\gamma}^{-1}(s)^{1/p} - 1|^p ds$. Taking infimum over $\gamma(t)$ and $\gamma^{-1}(s)$ on both sides, the symmetry holds. \square

- (c) Triangle-inequality: $d(f, g) \leq d(f, h) + d(h, g)$ for any $f, g, h \in \mathcal{S}$.

Proof Assume $f = \sum_{i=1}^L \delta(t - r_i)$, $g = \sum_{j=1}^M \delta(t - s_j)$, and $h = \sum_{k=1}^N \delta(t - t_k)$. The triangle-

inequality will hold if for any $\gamma(t), \gamma_1(t) \in \Gamma$ we have

$$\begin{aligned} & \left[L + M - 2 \sum_{i=1}^L \sum_{j=1}^M \mathbf{1}_{r_i=\gamma(s_j)} \right. \\ & \quad \left. + \lambda \int |1 - \dot{\gamma}(t)^{1/p}|^p dt \right]^{1/p} \\ & \leq \left[L + N - 2 \sum_{i=1}^L \sum_{k=1}^N \mathbf{1}_{r_i=\gamma_1(t_k)} \right. \\ & \quad \left. + \lambda \int |1 - \dot{\gamma}_1(t)^{1/p}|^p dt \right]^{1/p} \\ & \quad + \left[N + M - 2 \sum_{k=1}^N \sum_{j=1}^M \mathbf{1}_{t_k=\gamma_1^{-1} \circ \gamma(s_j)} \right. \\ & \quad \left. + \lambda \int |1 - \dot{\gamma}_1^{-1} \circ \dot{\gamma}(t)^{1/p}|^p dt \right]^{1/p} \end{aligned} \quad (16)$$

where \circ denotes function composition in Γ . Let $s = \gamma(t)$ in the first and third integrations and let $s = \gamma_1(t)$ in the second integration, Eq. (16) can be written as

$$\begin{aligned} & \left[L + M - 2 \sum_{i=1}^L \sum_{j=1}^M \mathbf{1}_{r_i=\gamma(s_j)} \right. \\ & \quad \left. + \lambda \int |1 - \dot{\gamma}^{-1}(s)^{1/p}|^p ds \right]^{1/p} \\ & \leq \left[L + N - 2 \sum_{i=1}^L \sum_{k=1}^N \mathbf{1}_{r_i=\gamma_1(t_k)} \right. \\ & \quad \left. + \lambda \int |1 - \dot{\gamma}_1^{-1}(s)^{1/p}|^p ds \right]^{1/p} \\ & \quad + \left[N + M - 2 \sum_{k=1}^N \sum_{j=1}^M \mathbf{1}_{\gamma_1(t_k)=\gamma(s_j)} \right. \\ & \quad \left. + \lambda \int |\dot{\gamma}^{-1}(s)^{1/p} - \dot{\gamma}_1^{-1}(s)^{1/p}|^p ds \right]^{1/p} \end{aligned}$$

It is straightforward to verify that

$$\begin{aligned} N + \sum_{i=1}^L \sum_{j=1}^M \mathbf{1}_{r_i=\gamma(s_j)} & \geq \sum_{i=1}^L \sum_{k=1}^N \mathbf{1}_{r_i=\gamma_1(t_k)} \\ & \quad + \sum_{k=1}^N \sum_{j=1}^M \mathbf{1}_{\gamma_1(t_k)=\gamma(s_j)}. \end{aligned}$$

This implies that

$$\begin{aligned} L + M - 2 \sum_{i=1}^L \sum_{j=1}^M \mathbf{1}_{r_i=\gamma(s_j)} \\ \leq L + N - 2 \sum_{i=1}^L \sum_{k=1}^N \mathbf{1}_{r_i=\gamma_1(t_k)} \\ + N + M - 2 \sum_{k=1}^N \sum_{j=1}^M \mathbf{1}_{\gamma_1(t_k)=\gamma(s_j)}. \end{aligned}$$

Therefore, Eq. (16) will hold if we can show that for any non-negative values x, y, z with $x \leq y + z$ and a, b, c in the L^p space

$$\begin{aligned} \left[x + \int |a - b|^p dt \right]^{1/p} \leq \left[y + \int |a - c|^p dt \right]^{1/p} \\ + \left[z + \int |c - b|^p dt \right]^{1/p} \end{aligned} \quad (17)$$

Indeed, Eq. (17) is also a direct result from the triangle inequalities in the L^p space and the L^p norm in the Euclidean \mathbb{R}^2 vector space. \square

(B.2) Then, we show that when $\sigma \rightarrow 0$, the limiting form of $D_p[\sigma, \lambda]$ is $d_p[\lambda]$.

Proof Assume $f(t) = \sum_{i=1}^M K_\sigma(t - t_i) \in \mathcal{S}_M^{(\sigma)}$ and $g(t) = \sum_{j=1}^N K_\sigma(t - s_j) \in \mathcal{S}_N^{(\sigma)}$ are two smoothed spike trains in $[0, T]$. When σ is sufficiently small, spikes in f and g will either fully match (two spike kernels are entirely overlapped) or do not match at all (two spike kernels are well separated). Moreover, when two spike kernels $K_\sigma(t - t_i)$ in f and $K_\sigma(t - s_j)$ in g are fully matched, there must be no time warping in this match. That is, $\gamma(s_j) = t_i$ and $\dot{\gamma}(t) = 1$ when $s_j - \sigma < t < s_j + \sigma$.

This indicates that

$$\begin{aligned} \left| f(t)^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p} \right|^p \\ = \begin{cases} 0 & \text{if fully matched} \\ f(t) + g(\gamma(t))\dot{\gamma}(t) & \text{if not matched} \end{cases} \end{aligned} \quad (18)$$

Therefore, when σ is sufficiently small, $\int_0^T |f(t)^{1/p} - [g(\gamma(t))\dot{\gamma}(t)]^{1/p}|^p dt = M + N - 2 \sum_{i=1}^M \sum_{j=1}^N \mathbf{1}_{t_i=\gamma(s_j)}$. This is actually the total number of spike kernels that are not overlapped after warping. As the penalty term in $d_p[\lambda]$ is exactly

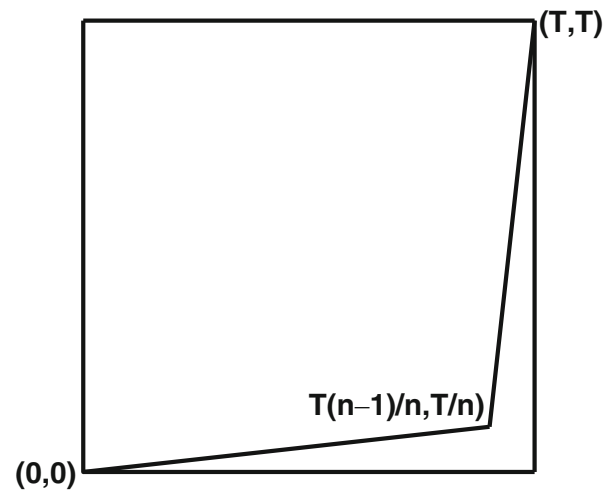


Fig. 11 A piecewise linear warping from $[0, T]$ to $[0, T]$

the same as that in $D_p[\sigma, \lambda]$, $d_p[\lambda]$ is the limiting form of $D_p[\sigma, \lambda]$ when $\sigma \rightarrow 0$. \square

Appendix C: Limiting form for upper bounds

Here we show that the upper bound is exact in some limiting form for both D_2 and D_1 distances. Our proof is based on an example plot in Fig. 11, which shows an infinite sequence of piecewise linear warpings from $[0, T]$ to $[0, T]$. Computing the integral of the piecewise linear warping, we have

$$\begin{aligned} \int_0^T \sqrt{\dot{\gamma}(t)} dt &= \int_0^{\frac{n-1}{n}T} \sqrt{\frac{1}{n-1}} dt + \int_{\frac{n-1}{n}T}^T \sqrt{n-1} dt \\ &= \frac{2T}{n} \sqrt{n-1} \rightarrow 0 (n \rightarrow \infty) \end{aligned}$$

and

$$\begin{aligned} \int_0^T |1 - \dot{\gamma}(t)| dt &= \int_0^{\frac{n-1}{n}T} \left(1 - \frac{1}{n-1}\right) dt + \int_{\frac{n-1}{n}T}^T (n-1-1) dt \\ &= \frac{2(n-2)}{n} T \rightarrow 2T (n \rightarrow \infty). \end{aligned}$$

This proves that the upper bound, $2\lambda T$, is exact for both D_2 and D_1 .

Appendix D: Optimal time warping

Here we state some theoretical facts on the optimal time warping.

Lemma 1 For any positive diffeomorphism γ from $[a, c]$ to $[b, d]$ (i.e. $\gamma(a) = b, \gamma(c) = d, 0 < \dot{\gamma}(t) < \infty$)

and $1 \leq p < \infty$, the time warping has the following optimal cost:

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

When $p > 1$, the equality holds if and only if the warping is linear from $[a, c]$ to $[b, d]$. When $p = 1$, such linear property is sufficient.

Proof In domain $[a, c]$, for any function $f(t) \in L^p$, we denote its L^p norm as $\|f(t)\|_p = (\int_a^c |f(t)|^p dt)^{1/p}$. Then,

$$\begin{aligned} \int_a^c |1 - \dot{\gamma}(t)|^{1/p} dt &= \|1 - \dot{\gamma}(t)\|_p^p \\ &\geq \left| \|1\|_p - \|\dot{\gamma}(t)\|_p \right|^p \\ &= |(c-a)^{1/p} - (d-b)^{1/p}|^p. \end{aligned}$$

Using theory in the L^p spaces, when $p > 1$, the equality holds if and only if $\dot{\gamma}(t)$ is a positive constant, i.e. the warping is linear from $[a, c]$ to $[b, d]$. When $p = 1$, the equality holds if the sign of $1 - \dot{\gamma}(t)$ does not change in $[a, c]$ (always nonnegative or always nonpositive), for example, if the warping is linear from $[a, c]$ to $[b, d]$. \square

Theorem 1 Let spike trains $f, g \in S_M$ in $[0, T]$ and the corresponding ISIs are $\Delta s_1, \dots, \Delta s_{M+1}$ and $\Delta t_1, \dots, \Delta t_{M+1}$, respectively. If $\lambda < 1/(2^{p-1}T)$, then

$$d_p[\lambda](f, g) = \left(\lambda \sum_{k=1}^{M+1} |(\Delta s_k)^{1/p} - (\Delta t_k)^{1/p}|^p \right)^{1/p} \quad (19)$$

Proof In domain $[0, T]$, for any function $h(t) \in L^p$, we denote its L^p norm as $\|h(t)\|_p = (\int_0^T |h(t)|^p dt)^{1/p}$. Then,

$$\begin{aligned} \int_0^T |1 - \dot{\gamma}(t)|^{1/p} dt &= \|1 - \dot{\gamma}(t)\|_p^p \\ &\leq \left| \|1\|_p + \|\dot{\gamma}(t)\|_p \right|^p = 2^p T. \end{aligned}$$

This indicates that the total penalty is bounded by $\lambda 2^p T$. If one spike in one train does not match with any spike in the other train, the cost in the matching term should be at least 2. Therefore, when $\lambda 2^p T < 2$, or $\lambda < 1/(2^{p-1}T)$, all M spikes in f and all M spikes in g must be one-to-one matched for the optimal distance.

When all spikes are matched, the distance between two spike trains only depends on the optimal warping cost for each pair of ISIs. The distance in Eq. (19) is a direct result from Lemma 1. \square

Theorem 2 Let spike trains $f \in S_M$ and $g \in S_N$ in $[0, T]$. If $\lambda < 1/(2^{p-1}T)$, then

$$\begin{aligned} d_p[\lambda](f, g) &= (|M - N| + \lambda \cdot (\text{optimal warping cost in the} \\ &\quad \text{associated ISIs}))^{1/p}. \end{aligned} \quad (20)$$

Proof Without loss of generality, we assume $M \geq N$. The matching cost is minimal with value $M - N$ when all N spikes in g are matched with N (of M) spikes in f . In this case, we have $N + 1$ pairs of ISIs associated the N spikes in f and N spikes in g . As shown in Theorem 1, the total penalty is always bounded by $\lambda 2^p T$. Therefore, when $\lambda < 1/(2^{p-1}T)$, the N spikes in g must match N (of M) spikes in f and the optimal total cost which results in the following distance:

$$\begin{aligned} d_p[\lambda](f, g) &= (M - N + \lambda \cdot (\text{optimal warping cost in the } N + 1 \\ &\quad \text{pairs of ISIs}))^{1/p}. \end{aligned}$$

Moreover, using Theorem 1, if the $N + 1$ pairs of ISIs are $\{\Delta s_k\}_{k=1}^{N+1}$ and $\{\Delta t_k\}_{k=1}^{N+1}$, then the optimal time warping cost is: $\sum_{k=1}^{N+1} |\Delta s_k^{1/p} - \Delta t_k^{1/p}|^p$. Therefore, the distance between f and g can be further represented in the following form:

$$d_p[\lambda](f, g) = \left((M - N) + \lambda \sum_{k=1}^{J+1} |\Delta s_k^{1/p} - \Delta t_k^{1/p}|^p \right)^{1/p}.$$

\square

Appendix E: Convergence of the MM-algorithm

Denote the estimated mean in the i th iteration of the MM-algorithm as $S^{(i)}$. Now we show that the sum of squared distance $\sum_{k=1}^K d_2^2(S_k, S^{(i)})$ decreases iteratively. That is,

$$\sum_{k=1}^K d_2^2(S_k, S^{(i)}) \leq \sum_{k=1}^K d_2^2(S_k, S^{(i-1)}).$$

As 0 is a natural lower bound, the iteration will always converge.

Proof In the i th iteration, we have the initial $S^{(i-1)}$ and let $d_{k,i-1}$ denote the distance between S_k and $S^{(i-1)}$ based on current spike train matching subset. The number of spikes in the mean, n , is known beforehand. Hence the squared distance on the matching part is always $|n_k - n|$, which is invariant with respect to any subset of spikes being used in the matching process.

Therefore, we only need to focus on how the optimal matching cost changes from $S^{(i-1)}$ to $S^{(i)}$.

In case 1, we compute the distance, r_k , between S_k and $S^{(i-1)}$ and find $n + 1$ ISIs in S_k that have the optimal time warping with the $n + 1$ ISIs in $S^{(i-1)}$. As the ISIs are optimally selected, $r_k \leq d_{k,i-1}$. In case 2, The decrease in distance also holds. However, as $n_k < n$, the number of ISIs is only $n_k + 1$, which could not be used to compute the mean using the closed-form formula in Eq. (11). In the MM-algorithm, we propose to linearly interpolate $n - n_k$ imaginary spikes in S_k based on the n_k pairs of spikes between S_k and $S^{(i-1)}$. Here we show this interpolation does not change the warping distance between S_k and $S^{(i-1)}$.

Indeed, as the warping distance is represented in the sum of squared form (see the proof in Theorem 2), we only need to show the invariance in each squared term. Let Δa be one ISI in S_k and Δb be the associated ISI in $S^{(i-1)}$. Their squared distance is

$$(\sqrt{\Delta a} - \sqrt{\Delta b})^2.$$

Assume there are h spikes in the ISI in $S^{(i-1)}$, which partition Δb into $h + 1$ ISIs, denoted by (u_1, \dots, u_{h+1}) with $\sum_{j=1}^{h+1} u_j = \Delta b$. Our algorithm inserts h spikes in the ISI in S_k using linear interpolation. Therefore the $(h + 1)$ ISIs in S_k have lengths

$$\frac{\Delta a}{\Delta b}(u_1, \dots, u_{h+1}).$$

Then the new warping distance based on these ISIs are

$$\begin{aligned} \sum_{j=1}^{h+1} \left(\sqrt{u_j} - \sqrt{\frac{\Delta a}{\Delta b} u_j} \right)^2 &= \sum_{j=1}^{h+1} u_j \left(1 - \sqrt{\frac{\Delta a}{\Delta b}} \right)^2 \\ &= \Delta b \left(1 - \sqrt{\frac{\Delta a}{\Delta b}} \right)^2 \\ &= (\sqrt{\Delta a} - \sqrt{\Delta b})^2 \end{aligned}$$

In summary, over step 2, the distance between S_k and $S^{(i-1)}$ has decreased and we have found a set of $n + 1$ ISIs in S_k that correspond to the $n + 1$ ISIs in $S^{(i-1)}$. In step 3, we update the ISIs in $S^{(i-1)}$ to $S^{(i)}$. This update is based on the closed-form formula in Eq. (11) which minimizes the sum of squared distance. Therefore,

$$\sum r_k^2 \geq \sum d_{k,i}^2.$$

As $r_k \leq d_{k,i-1}$, $k = 1, \dots, K$, we have

$$\sum d_{k,i-1}^2 \geq \sum d_{k,i}^2.$$

□

References

- 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., 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.
- Aronov, D., & Victor, J. (2004). Non-Euclidean properties of spike train metric spaces. *Physical Review E*, 69, 061905.
- Bertsekas, D. P. (1995). *Dynamic programming and optimal control*. Athena Scientific.
- Bhattacharya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.
- 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.
- Brockwell, A. E., Rojas, A. L., & Kass, R. E. (2004). Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology*, 91, 1899–1907.
- 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.
- Čencov, N. N. (1982). *Statistical decision rules and optimal inferences, translations of mathematical monographs* (Vol. 53). Providence: AMS.
- Chi, Z., Wu, W., Haga, Z., Hatsopoulos, N., & Margoliash, D. (2007). Template-based spike pattern identification with linear convolution and dynamic time warping. *Journal of Neurophysiology*, 97, 1221–1235.
- Curran-Everett, D., & Benos, D. J. (2004). Guidelines for reporting statistics in journals published by the american physiological society. *Journal of Applied Physiology*, 97, 457–459.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge: MIT.
- Dubbs, A. J., Seiler, B. A., & Magnasc, M. O. (2009). A fast Lp 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., Ventura, V., & Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94, 8–25.
- Kass, R. E., & Vos, P. W. (1997). *Geometric foundations of asymptotic inference*. New York: Wiley.
- 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.

- 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.
- Perkel, D. H., Gerstein, G. L., & Mooren, G. P. (1967a). Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophysics Journal*, 7, 391–418.
- Perkel, D. H., Gerstein, G. L., & Mooren, G. P. (1967b). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophysics Journal*, 8, 419–440.
- 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.
- Rieke, F., Warland, D., Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge: MIT.
- Schrauwen, B., & van Campenhout, J. (2007). Linking non-binned spike train kernels to several existing spike train metrics. *Neurocomputing*, 70, 1247–1253.
- Schreiber, S., Fellousb, J., Whitmerc, D., Tiesingaa, P., & Sejnowskib, T. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52–54, 925–931.
- Seber, G. A. F. (2004). *Multivariate observations*. New York: Wiley.
- Srivastava, A., Jermyn, I. H., & Joshi, S. H. (2007). Riemannian analysis of probability density functions with applications in vision. In *IEEE conference on computer vision and pattern recognition (CVPR)*
- Truccolo, W., Eden, U., Fellows, M., Donoghue, J., & Brown, E. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *Journal of Neurophysiology*, 93, 1074–1089.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.
- van Rossum, M. C. W. (2001). A novel spike distance. *Neural Computation*, 13, 751–763.
- 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.
- 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.
- Wu, W., & Srivastava, A. (2011). Towards statistical summaries of spike train data. *Journal of Neuroscience Methods*, 195, 107–110.
- 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.