Evaluating dependence in spike train metric spaces

Sohan Seth, Austin J. Brockmeier, John S. Choi, Mulugeta Semework, Joseph T. Francis, José C. Príncipe

Abstract— **Assessing dependence between two sets of spike trains or between a set of input stimuli and the corresponding generated spike trains is crucial in many neuroscientific applications, such as in analyzing functional connectivity among neural assemblies, and in neural coding. Dependence between two random variables is traditionally assessed in terms of mutual information. However, although well explored in the context of real or vector valued random variables, estimating mutual information still remains a challenging issue when the random variables exist in more exotic spaces such as the space of spike trains. In the statistical literature, on the other hand, the concept of dependence between two random variables has been presented in many other ways, e.g. using copula, or using measures of association such as Spearman's ρ, and Kendall's τ. Although these methods are usually applied on the real line, their simplicity, both in terms of understanding and estimating, make them worth investigating in the context of spike train dependence. In this paper, we generalize the concept of association to any abstract metric spaces. This new approach is an attractive alternative to mutual information, since it can be easily estimated from realizations without binning or clustering. It also provides an intuitive understanding of what dependence implies in the context of realizations. We show that this new methodology effectively captures dependence between sets of stimuli and spike trains. Moreover, the estimator has desirable small sample characteristic, and it often outperforms an existing similar metric based approach.**

I. INTRODUCTION

Many neuroscientific applications such as identifying functional connectivity among neural assemblies [1], building stimulus response models based on input-output observations [2], and neural coding [3], rely on assessing dependence between two sets of spiking activity, or a set of stimuli and the corresponding spike trains. As in communication systems and machine learning [4], neuroscientists often rely on mutual information as a measure of dependence [5], where the use of mutual information is motivated by Shannon's mathematical interpretation of information and entropy.

Formally, mutual information between a pair of random variables (X, Y) that assume values in the joint space $\mathcal{X} \times \mathcal{Y}$ with joint probability law $(X, Y) \sim \mathbb{P}_{xy}$ and marginal probability laws $X \sim \mathbb{P}_x$ and $Y \sim \mathbb{P}_y$ respectively, is defined as,

$$\mathrm{MI} = \int \log \left(\frac{\mathrm{d}\mathbb{P}_{xy}}{\mathrm{d}\mathbb{P}_{x \perp y}} \right) \mathrm{d}\mathbb{P}_{xy}$$

where $\mathbb{P}_{x \perp y} = \mathbb{P}_x \mathbb{P}_y$. Given another measure μ on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbb{P}_{xy} \ll \mathbb{P}_{x \perp y} \ll \mu$, mutual information can also be expressed as

$$\mathrm{MI} = \int \log \left(\frac{\mathrm{d}\mathbb{P}_{xy}/\mathrm{d}\mu}{\mathrm{d}\mathbb{P}_{x \perp y}/\mathrm{d}\mu} \right) \frac{\mathrm{d}\mathbb{P}_{xy}}{\mathrm{d}\mu} \mathrm{d}\mu$$

where $\ll$ implies absolute continuity.

Estimating mutual information in the Euclidean space i.e., $\mathcal{X} = \mathbb{R}^m$ and $\mathcal{Y} = \mathbb{R}^n$ is well explored, and it is comparatively simpler since in Euclidean space, μ can be chosen to be the Lebesgue measure i.e., $\mu(\mathrm{d}x, \mathrm{d}y) = \mathrm{d}x \mathrm{d}y$, which simplifies the definition of mutual information to

$$\mathrm{MI}_e = \int \log \left(\frac{f_{xy}(x, y)}{f_x(x) f_y(y)} \right) f_{xy}(x, y) \mathrm{d}x \mathrm{d}y$$

where $f_z(z) = \mathrm{d}\mathbb{P}_z/\mathrm{d}z$ defines the probability density function. Mutual information in Euclidean space is usually estimated by estimating the individual probability densities by either a Parzen type approach [6] or a nearest neighbor based approach [7]. Moreover, if the random variables are categorical, then μ can be considered to be the counting measure, and the definition further simplifies to

$$\mathrm{MI}_c = \sum_{(x,y)} \mathbb{P}_{x,y} \log \left(\frac{\mathbb{P}_{xy}}{\mathbb{P}_x \mathbb{P}_y} \right)$$

where these probabilities can be easily estimated by counting.

On the other hand, there are no well-defined estimators of mutual information in non-Euclidean spaces, such as space of spike trains, that lack an appropriate base measure μ[1]. Mutual information between two sets of spike trains are often computed by representing them either in the Euclidean or in the categorical space. The simplest approach to represent a spike train in the Euclidean space is to use binning where each bin acts as an independent axis in the Euclidean space. However, this approach has several limitations. First, binning destroys the time structure of the spike trains since it disregards the ordering of the bins. Second, the precision of spike times is controlled by the size of the bin, and third, although spike time precision can be preserved by making the binsize arbitrarily small, a smaller binsize effectively increases the dimensionality of the resulting Euclidean space that in turn makes the estimation poor due to sparsity. To alleviate these issues, Victor and Purpura proposed an alternate approach [8]. In their approach, the set of spike trains is divided into clusters using a suitable binless distance metric, and the labels of the clusters are treated as random

---

[1]Notice that mutual information is invariant to the choice of μ, however the estimators of mutual information usually rely on a certain μ to estimate the density function

variables, i.e., effectively representing a subset of spike trains as a category. Although this approach prevents binning, it has several limitations on its own. First, this method relies on the clustering method, second, it is only sensitive to the number of spike trains that fall in a cluster, rather than the relative positions of these spike trains in the cluster, and third, it requires comparatively more spike train realizations per cluster to effectively estimate the mutual information. Recently Victor has proposed yet another approach for capturing mutual information by dividing the spike trains in groups by their respective spike counts, and computing the mutual information of the individual groups by mapping them into Euclidean spaces with the dimension corresponding to the number of spikes [9]. Although technically precise, this approach suffers from poor estimation when the spike trains have a flat spike count distribution, i.e., while dividing them in groups, each group becomes sparse. This problem is usually tackled by limiting the maximum number of spikes, and projecting any spike trains with more number of spikes to a lower dimensional spike train i.e., spike train having less spikes. However, the process of projecting the spike trains is non-invertible, and hence, lossy. Therefore, the validity of the estimated mutual information remains questionable (see section II-A for details).

Due to the inherent difficulty in estimating mutual information, it is natural to ask whether the dependence between sets of spike trains can be assessed in other ways. In the statistical literature the concept of dependence has been explored beyond the concept of mutual information, especially in the context of real valued random variables, e.g. copulas or measures of association such as Spearman's ρ and Kendall's τ, have been used to understand and quantify dependence [10]. Although, mostly defined on the real line, the simplicity of these approaches, both in terms of understanding and computation, make them worth exploring in the context of capturing dependence between two sets of spike trains. In this paper, we propose a generalization of the concept of association to arbitrary metric spaces that is suitable for capturing dependence between sets of spike trains, since the space of spike train can be treated as a metric space [11], [8], [12]. Moreover, the proposed approach satisfies many desired criteria of a measure of dependence such as (1) it is parameter free, (2) it is bounded between 0.5 and 1, where 0.5 is reached upon independence and 1 is reached upon strict dependence, and (3) it conveys a clear understanding of dependence in the context of the realizations. This is advantageous in contrast to the concept of mutual information which, although well understood in the context of the probability laws, is often obscure after being estimated from the realizations, especially in the case of continuous random variables.

A measure of dependence can be applied in many different areas of neuroscientific exploration such as functional connectivity analysis and neural coding. In this paper, we explore the concept of dependence for spiking activity due to stimuli or due to other spiking activity. In particular,

we use a simulated data where the stimuli controls the phase of periodic spiking activity, and an experimental data where the stimuli is electrical microstimulation with different amplitudes and durations. In the experimental data, we apply the proposed method to assess the ranges of parameters that increase the dependence between the stimulation and the resulting spiking activity. In addition, we also explore a synthetic example of dependence between two sets of spiking activity. Our simulations reveal that the proposed approach is a suitable tool for neuroscientific exploration, and is potentially useful for further refining the experimental design in this context.

The rest of paper is organized as follows. In section II we start by briefly describing the concept of association for real valued random variables, and then, generalize this idea to introduce a generalized measure of association. In section III, we apply the proposed approach on synthetic data, and compare its performance against an existing clustering-based method for estimating mutual information (or transmitted information [11]), and in section IV we analyze its utility on the experimental data. In section V, we conclude the paper with a brief summary of the proposed approach, and an overview of the future work.

II. METHOD

We start this section by briefly discussing the existing literature on measures of association, and then, describe a generalization of this approach to capture dependence in the context of spike trains.

A. Background

The simplest measure of association is the correlation as introduced by Pearson [10]. Given realizations $\{(x_i, y_i)\}_{i=1}^n$ of a pair of random variables, correlation is defined as $\sum_i x_i y_i$. The intuitive purpose of defining correlation in such a way is to capture whether large values of random variable X are associated with large values of random variable Y. This approach, however, only captures linear dependence between two random variables i.e., it is maximum when the random variables are linear functions of each other. Spearman generalized this idea and considered the correlation between the ranks of $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ as a measure of association [13]. Unlike Pearson's correlation, Spearman's approach, also known as Spearman's ρ, captures monotone dependence i.e., the dependence is maximum when the random variables are monotonic functions of each other and not just linear. Kendall also discussed a similar idea but from a different perspective, and quantified dependence by the difference between the number of concordant and discordant realization pairs, where two pairs (x_i, y_i) and (x_j, y_j) are called concordant if $(x_i - x_j)$ and $(y_i - y_j)$ have the same signature or they are called discordant [14]. Both Spearman's and Kendall's approach, also known as Kendall's τ, capture the same notion of dependence as correlation, i.e., whether large values of X are associated with large values of Y, but unlike Pearson's approach, these approaches judge a value being large only in the context of the rest of the realizations,

i.e., by considering its rank rather than its absolute value. An interesting attribute of measures of association is that they provide a clear understanding of what dependence implies in the context of realizations. This attribute is often missing in the context of other measures of dependence such as mutual information, where, although the statistical meaning, i.e., the function of the random variable is well understood, the corresponding estimator is not. For example, mutual information as a statistic is invariant to one-to-one transformation of the random variables, and it reaches its maximum when two random variables are functionally related, but these attributes can not be satisfied by estimators of mutual information since the estimators must be regularized. Therefore, in the context of the realizations, the exact meaning of mutual information still remains to be explored. The concept of association, on the other hand, is well understood in the context of the realizations, and therefore, is an attractive alternative to the mutual information.

B. Generalized association

We generalize the idea of association as follows: given a set of realizations $\{(x_i, y_i)\}_{i=1}^n$, we consider it to be dependent *if pairs of close realizations of one random variable, say X, are associated with pairs of close realization of the other random variable, Y*. This approach is intuitive in the sense that if two random variables are related by a continuous mapping then we expect this statement to be true. However, defining dependence in this way does not imply an appropriate approach to quantify it. In fact, this statement can be materialized in several intuitive ways, e.g. using graph-theoretic approaches as explored in [15]. We discuss a conceptually different approach for quantifying dependence, such that it satisfies some intuitive properties that a measure of dependence should possess. For example, the dependence value should be bounded, and the two extremes should belong to independence and strict dependence. Moreover, the notion of strict dependence should be well understood, and finally, since we aim at finding dependence between two sets of spike trains, the proposed measure should be well defined in the context of two metric spaces, and not just the Euclidean space.

Following these requirements, we define the *closeness* of two realizations in terms of the respective distance metrics of the spaces where the random variables assume values in. Notice that these two spaces $\mathcal{X}$ and $\mathcal{Y}$ can be very different with different distance metrics. For example, $\mathcal{X}$ could be the real line for stimulus parameters, whereas $\mathcal{Y}$ could be the space of spike trains. Then given realizations $\{(x_i, y_i)\}_{i=1}^n$, we capture dependence by finding the nearest neighbor, say x_j, of a realization, say x_i, in one of the spaces $\mathcal{X}$, and observing how far the corresponding realizations y_i and y_j are in the other space $\mathcal{Y}$. Say these realizations are rank r_i apart (in the context of the other realizations), then we repeat this process for all samples x_i's, and build a measure of dependence by quantifying an appropriate attribute of these ranks. We simplify our approach in the following algorithm,

1) For all $i \in \{1, \ldots, n\}$, repeat the following;

2) Find x_{j^*} closest to x_i in terms of $\mathfrak{d}_{\mathcal{X}}$ i.e.,

$$j^* = \arg\min_{j \neq i} \mathfrak{d}_{\mathcal{X}}(x_i, x_j),$$

3) Find rank r_i of y_{j^*} in terms of $\mathfrak{d}_{\mathcal{Y}}$ i.e.,

$$r_i = \#\{j : j \neq i, \mathfrak{d}_{\mathcal{Y}}(y_j, y_i) \leq \mathfrak{d}_{\mathcal{Y}}(y_{j^*}, y_i)\}.$$

where $\mathfrak{d}_{\mathcal{Z}}$ denotes the associated metric of the space $\mathcal{Z}$. Notice that if two random variables are independent, then any rank r_i should take any value in the set $\{1, \ldots, n-1\}$ with equal probability, i.e., the ranks should be uniformly distributed. On the other hand, we say that the two variables are strictly dependent if these ranks are always 1. This happens when the nearest neighbors of any realizations of X are also the nearest neighbor of the corresponding realizations of Y. For example, a trivial case is $X = Y$. Finally, if the two random variables are dependent, then these ranks would tend to be close to 1, i.e., the more dependent two random variables are, the more skewed the distribution of ranks would be.

Therefore, following the requirements mentioned above, we quantify dependence as the area under the empirical cumulative distribution function of the realizations r_i's. We call this estimate the generalized measure of association or GMA. Notice that the minimum value of this estimator is, thus, 0.5 when r_i's are uniformly distributed, whereas the maximum value of the estimator is 1 when the random variables are strictly dependent. However, notice that we do not claim the minimum dependence value to be a sufficient condition for independence. This attribute is very similar to the measures of association where zero association does not guarantee independence. Moreover, since we are working with empirical estimates, the minimum value is not strictly 0.5 but it can be less than 0.5, and approaches 0.5 when n increases. This nature is very similar to any consistent estimator of independence. Finally, although we have not found a formal proof, we conjecture that the proposed measure does not assume value between 0 and 0.5 (except for the inaccuracy's due to finite samples), and therefore, we always use a one-sided test to judge the significance of an acquired dependence value.

C. Adjustment for ties

In the derivation of generalized association, we have implicitly assumed that the estimated ranks do not overlap, i.e., the distances of the rest of the realizations from a certain realization are distinct. Although this often happens when the random variables are continuous, this condition can be easily violated. For example, consider the realizations $\{(1, 2), (2, 1), (3, 2)\}$. Then if we take $(x_i, y_i) = (2, 1)$, then both $(1, 2)$ and $(3, 2)$ share the same distance from $x_i = 1$ in terms of $\mathfrak{d}_{\mathcal{X}}$. Therefore, we have to decide which is the nearest neighbor i.e., $x_j = 1$ or $x_j = 3$. On the other hand, even if we randomly decide on either of them then for any of the two cases, $y_j = 2$, and therefore, we need to decide whether $r_i = 1$ or $r_i = 2$. Although this example is trivial, similar situation can be easily encountered in practice.

For example, if one uses a spike metric with no temporal precision, then two spike trains with same number of spikes are treated the same [8]. To tackle this situation, therefore, it is important to design an estimator that effectively captures the dependence when several realizations overlap or share the same distance from a certain realization. To capture this, we consider a probabilistic approach of ranking. In essence, instead of providing a certain realization (x_i, y_i) a rank r_i with probability 1, we spread the probability evenly to several ranks, say $r_i^{(1)}, \ldots, r_i^{(k)}$. To simplify our explanation, we start by considering a discrete random variable R for the ranks, and estimate its probability mass function over the iterations as follows,

1) Assign $P(R = r) = 0 \; \forall r \in \{1, \ldots, (n-1)\}$
2) For all $i \in \{1, \ldots, n\}$, repeat the following;
3) Find the set of points $\{x_{j^*} : j^* \in \mathcal{J}\}$ closest to x_i in terms of $\mathfrak{d}_\mathcal{X}$ i.e.,

$$\mathcal{J} = \{j^* : j^* = \arg\min_{j \neq i} \mathfrak{d}_\mathcal{X}(x_i, x_j)\}.$$

4) For all $j^* \in \mathcal{J}$, find the spread of ranks i.e., $r_{i,\max}$ and $r_{i,\min}$ of y_{j^*} in terms of $\mathfrak{d}_\mathcal{Y}$ such that

$$r_{i,\max} = \#\{j : j \neq i, \mathfrak{d}_\mathcal{Y}(y_j, y_i) \leq \mathfrak{d}_\mathcal{Y}(y_{j^*}, y_i)\} \text{ and}$$
$$r_{i,\min} = \#\{j : j \neq i, \mathfrak{d}_\mathcal{Y}(y_j, y_i) < \mathfrak{d}_\mathcal{Y}(y_{j^*}, y_i)\}.$$

5) For all rank values $r_{i,\min} < r \leq r_{i,\max}$, assign

$$P(R = r) \leftarrow P(R = r) + 1/n/|\mathcal{J}|/(r_{i,\max} - r_{i,\min}).$$

It can be easily verified that $\sum P(R = r) = 1$. Then, we compute GMA as the area under the CDF of the rank variable R. The adjusted measure is bounded between 0.5 and 1 where the lower and the upper limit is achieved under independence and strict dependence. Moreover, when the samples and distance values are distinct, then the former approach is a special case of the latter approach.

D. Examples

To further illustrate the idea of generalized association, we provide two simple examples in Euclidean space, and show how the proposed method captures dependence.

1) Clayton copula: Consider the bivariate Clayton copula i.e., $C(u, v) = (u^{-\rho} + v^{-\rho} - 1)^{-1/\rho}$ with zero mean unit variance Gaussian marginals. Here the coefficient ρ controls the dependence, and a higher value of ρ implies higher dependence between X and Y. Figure 1 shows the distribution of the rank variable, and the corresponding dependence captured by GMA for different values of ρ. We observe that the distribution of the rank variable is flat when ρ is small, and it gets more skewed when ρ increases, whereas the estimated GMA values show a monotonic increment.

2) Multivariate Gaussian: Consider two 3-dimensional zero mean identity covariance Gaussian vectors, where the first element of the first vector is correlated with the second element of the second vector. Figure 2 shows the distribution of the rank variable, and the estimated dependence value by the proposed framework for different values of the correlation coefficients. As expected, we observe a similar effect as
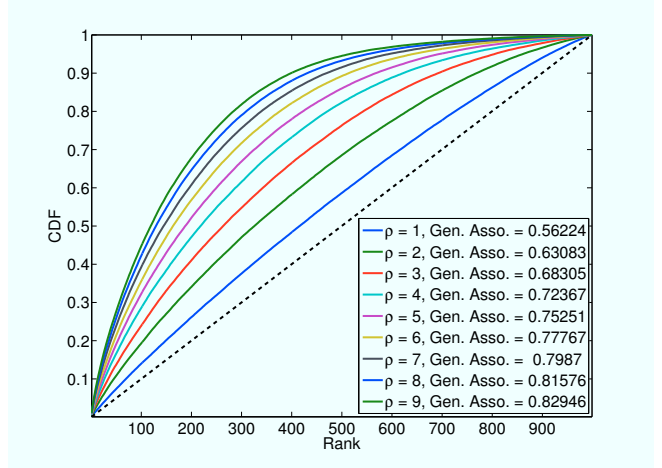

Fig. 1. Illustration of rank distribution and estimated dependence values. For this example, the data has been generated from a bivariate Clayton copula.
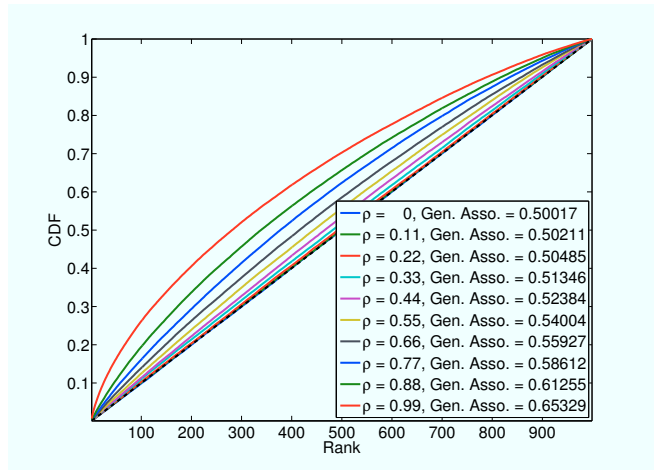

Fig. 2. Illustration of rank distribution and estimated dependence values. For this example, the data has been generated from a multivariate Gaussian distribution.

in the previous example. However, we notice that in this particular example, the dependence value is not close to 1 even though ρ is close to 1. This happens since when $\rho = 1$, only two elements of the random vectors become identical, whereas the rest of the elements still remain independent, and therefore, the two random vectors are only weakly dependent.

III. SIMULATION

In this section, we apply the proposed approach on simulated data, and compare its performance against the method proposed by [8]. Before proceeding, we briefly discuss the notion of distance on the space of spike trains. The distance between two spike trains can be defined in several ways, for example, see [8] and [16]. However, we choose to explore the measure suggested by [8], and leave a detailed study on the effect of other distance metrics as future work.
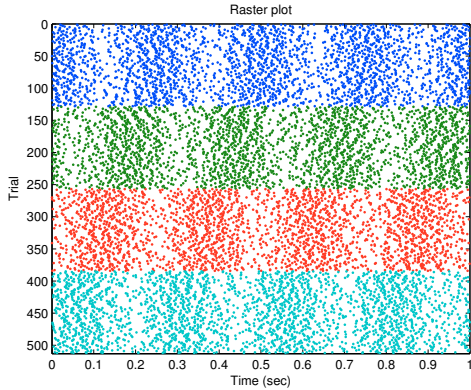
Fig. 3. Simulation II-D.2: Rasters for $\phi \in \{0, \pi/2, \pi, 3\pi/2\}$ and $m = 0.5$. Inhomogeneous Gamma renewal process with average spike rate of 20 Hz.

A. Victor-Purpura spike train metric

The spike-distance metric explored by Victor and Purpura [8] is a cost-based edit distance between sequences of time events. The distance is defined as the minimum cost of transforming one spike train to the other, by three simple operations i.e., shifting, adding, and removing, where each operation has an associated cost. Given two spike trains $x = \{x_i\}_{i=1}^m$ and $y = \{y_i\}_{i=1}^n$, each spike, say x_i, is either matched to a single spike from the other spike train, say y_j, to form a matching (x_i, y_j) with a cost of $q|x_i - y_j|$, or the spike is deleted from x_i at cost 1. Any unmatched spikes in y is assessed a cost of 1 each. Therefore, the sole parameter q controls temporal precision, i.e., the cost of moving a spike in time versus simply removing it and perhaps adding it somewhere else. Let $M = \{(x_i, y_j), \ldots\}$ be a set of matchings where a particular spike can appear in at most 1 element. Then the distance is defined as [17]

$$\mathfrak{d}^{(q)}(x,y) = \min_M \left[\sum_{(x_i,y_j) \in M} q|x_i - y_j| + (m + n - 2|M|) \right].$$

For the case $q = 0$, if $m = n$ then $\mathfrak{d}^{(0)}(x,y) = 0$, and if $m \neq n$, then $\mathfrak{d}^{(0)}(x,y) = |m - n|$. For $q = \infty$, unless spikes exactly align in time, they must be removed and re-added to align since no shifting is allowed, thus the minimum cost is $\mathfrak{d}^{(\infty)}(x,y) = m + n$. For $0 < q < \infty$, this metric takes into account the temporal positions of the spikes, and q controls the temporal resolution i.e., how much separated the spikes are in time. It has been recently shown that this metric is essentially a L_1 metric and that it can be extended to L_p spaces [17]. In addition, the metric has also been extended to multiple spike trains [18], but here we only concentrate on single spike trains.

1) Implementation: The framework for the simulation analysis was conducted using the Spike Train Analysis Toolkit, a neuroinformatics resource funded by the NIH's Human Brain Project. The toolbox has efficient C/C++ implementations of many information theoretic quantities [19]; of interest to us is its implementation of Victor Purpura spike train metric and the clustering-based mutual information estimator that we use for a baseline comparison in a simulated experiment. The default parameters of the clustering-based mutual information algorithm were used. For the actual metric we used Victor Purpura spike metric across a range of q, which controls the temporal precision. The generalized association code was implemented in MATLAB and the simulations were run concurrently using the same distance evaluations provided by the Spike Train Analysis Toolkit.

B. Dependence between stimulus and spike trains

We explore the characteristics of the generalized association to capture the dependence between a variable that controls the characteristics of simulated spike trains. We expand on an experiment on temporal phase discrimination used in [8]. The goal is to identify the dependence the phase imposes on cyclic spiking activity. The underlying spiking is an inhomogeneous Gamma renewal process, with underlying marginal intensity function

$$\lambda(t) = R_0[1 + m\cos(2\pi t/T + \phi)].$$

To form the Gamma process with shape k, where k is a positive integer value, a Poisson process with rate $k\lambda(t)$ is formed and every kth spike is kept. The resulting spike train is much more regular than an inhomogeneous Poisson process, which lends itself for more precise temporal comparisons. Spikes were generated with different phase ϕ chosen from a discrete set Φ of uniformly spaced phases in the interval $(0, 2\pi)$. Like the original experiment the base firing rate is $R_0 = 20$ Hz, the period was $T = 0.25$ ms, and length of spike trains generated is 1s. The tuning depth m was varied to alter the difficulty of the dependence estimation. The set of spike trains used for $m = 0.5$ and $|\Phi| = 4$ is shown in Figure 3.

Each value of ϕ serves as a category for estimating mutual information, but for generalized association these values are treated as continuous with distance metric, $\mathfrak{d}(\phi_1, \phi_2) = \cos^{-1}(\cos(\phi_1 - \phi_2))$. Thus, the measure of association can use the ordering to better quantify the dependence in this case. We investigate how well the methods capture dependence between ϕ and the resulting spike trains as the tuning depth m and the number of discrete phases $|\Phi|$ varies. Having only discrete phases Φ actually *under utilizes* GMA, as GMA can be applied to continuously distributed data, but using discrete values facilitates a direct comparison with the existing clustering-based method.

Each dependence measurement was calculated on 100 bootstrap trials. On each run the dependence was also calculated on shuffled categories to generate an independent surrogate distribution. We use a hypothesis test for independence with size 0.05 to test the statistical power of the methods. For tuning depth $m = 0$ the spike trains are independent of ϕ.

The simulation results for a tuning depth of $m = 0.5$ (Figure 4) show the dependence value across temporal precisions q. It peaks at a temporal precision of $q = 16\text{s}^{-1}$, this is the quarter-wave of the underlying period of the spike generation. Similar results for mutual information are
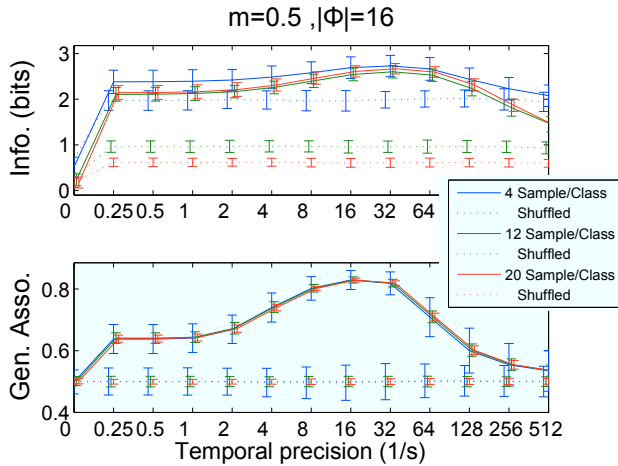
5

Fig. 4. Simulation II-D.2: Dependence measures across values of q when the number of classes is 16.
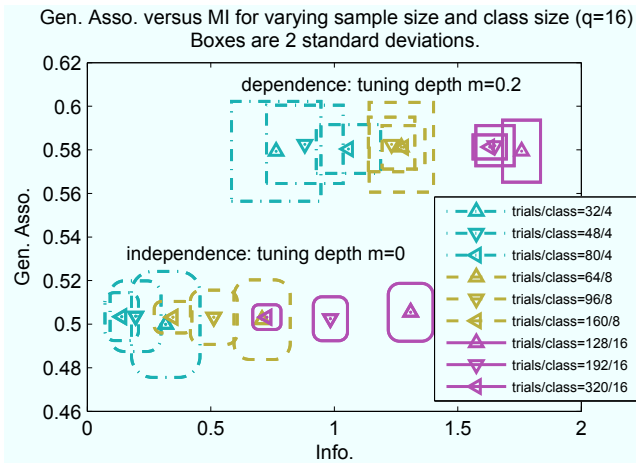

Fig. 5. Simulation II-D.2: Dependence measures for (in)dependence for different number of classes and samples when $q = 16\text{s}^{-1}$.
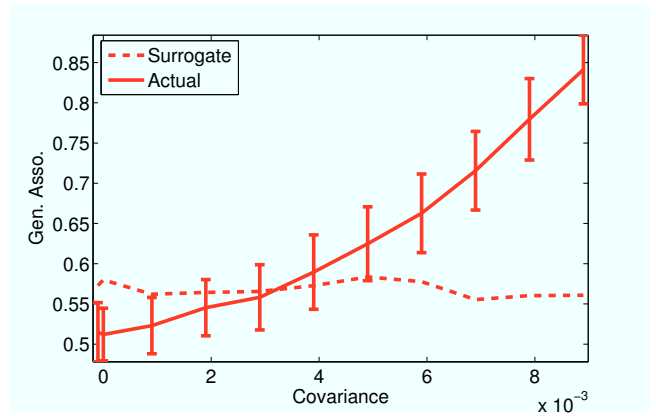

Fig. 6. Simulation III-C: Dependence between two sets of spike trains. Dotted line is the 0.95 quantile of the surrogate data, whereas solid line is the mean $\pm$ standard deviation of the actual GMA values. Both surrogate and actual values have been computed 128 times.

shown in [8]. GMA shows a similar trend across q, but it has a sharper peak surrounding the peak temporal precision, whereas, the mutual information is peaked for larger range of values; thus, GMA better highlights the optimal temporal precision. Across different sample sizes the variance of both methods decrease. Note that the small sample size bias developed in the shuffled surrogate is not removed from the mutual information calculation to illustrate its behavior across parameters; whereas, GMA has no change in bias for different sample sizes. The means of estimated GMA stay the same over different sample sizes whereas the means of the estimated mutual information decrease when sample size is increased. It should be noted that for different numbers of categories, $|\Phi|$, the theoretical value of mutual information should be $\log_2 |\Phi|$ bits.

We further examine how the methods are able to assess (in)dependence at tuning depth $m = 0.2(m = 0)$ while the temporal precision for the metric is held at $q = 16$ (Figure 5). It is evident that both methods are able to detect the dependence. In addition, both methods show a decrease in

variance as the number of samples per category increases; however, at small number of samples per category GMA has much lower variance; an important characteristic for neuroscience applications where the number of trials per stimulus or exactly repeated stimuli are limited. In addition, there is no change in the means of the estimated GMA values for different sample sizes. Overall, the results lend confidence that GMA is a stable measure that is able to capture association across changes in number of categories and samples, and is not plagued by need for bias correction that is needed with mutual information estimators.

C. Dependence between sets of spike trains

In this paper, we mostly focus on assessing dependence between input stimulus and the resulting spike trains. However, in this section, we provide a simple example to show that the proposed approach can also be applied to capture dependence between two sets of spike trains i.e., when both the spaces $\mathcal{X}$ and $\mathcal{Y}$ are spaces of spike trains. In order to do that, we follow [20] to generate two sets of spike trains with known covariance structure. Notice that [20] only proposes an approach to generate binned spike trains. Therefore, to collect the time instances of the events, we generate binary spike trains i.e., spike trains with at most one spike per bin, and consider the beginning of the bin to be the spike timing. For each $i = 1, \ldots, n$, we set the mean spiking probability for both spike trains to 0.05 and generate 100 bins each. This is equivalent to generating two 1s long spike trains with mean firing rate 5Hz each. We vary the covariance of the binary spike trains in the allowed range, from -10^{-4} to 9.9×10^{-3}, and present the estimated dependence values in Figure 6 using Victor Purpura metric with $q = 20$ and $n = 40$ realizations. It can easily observed that GMA can successfully detect the dependence among the spike train observations. Notice that, in a practical situation the two sets of spike trains might require two different metrics to capture their respective statistics. However, here we have used the same q for both since they have the same marginal statistics.
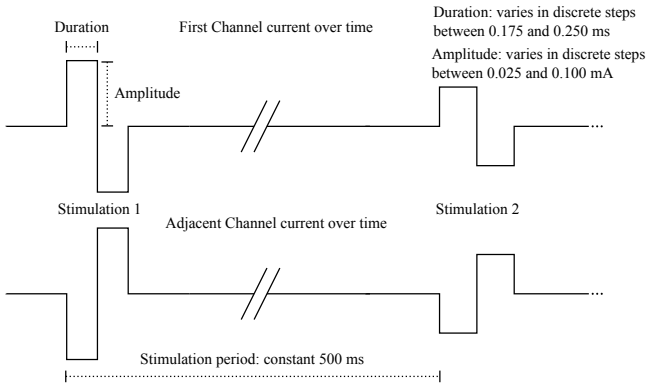
6

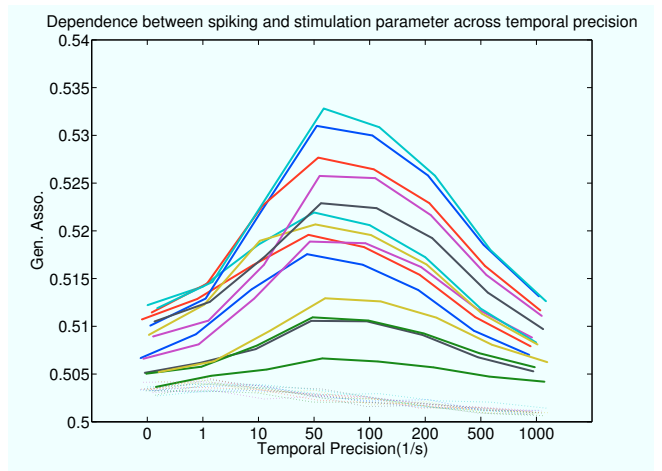Fig. 7. Experiment IV: Description of the microstimulation waveforms.

Fig. 8. Experiment IV: Generalized association across temporal precision for all 14 cortical channels. Dotted lines indicate the 95th percentile of the surrogate.

IV. EXPERIMENTAL DATA

We apply the proposed method to quantify the dependence between microstimulation parameters and the neural responses they elicit, in particular, the dependence of the neural response (spike trains) recorded in the somatosensory cortex (S1) on the electrical microstimulation administered in the thalamic somatosensory region. The neural spiking used in this analysis is from a rat with two chronically implanted 16-channel (2×8) tungsten micro-wire arrays (Tucker Davis). Neuronal activity was recorded from one array in the cortex (S1) using the Plexon multichannel acquisition processor. Action potentials were detected using a constant threshold and were sorted by a semi-automated clustering procedure (SortClient) using the first 3 principal components of the detected waveforms. The second array was positioned in the ventral posterior lateral (VPL) region of the thalamus.

Prior to the recording session, anesthesia was induced by isofluorane followed by a Nembutal injection and maintained with isofluorane. A pair of thalamic channels with response to cutaneous touch of a forepaw digit were selected as the channels for microstimulation. Bipolar microstimulation (AM Systems Model 2200 Isolator) was applied to two adjacent electrodes in the thalamic array. Each stimulation consisted of a biphasic square pulse (Figure 7). The pulse duration and current amplitude were varied, but the stimulations were always at 500 ms apart. During the session, 19 distinct pairs of pulse duration and current amplitude were applied, with 140 responses from each pair randomly permuted throughout the recording. We analyze 480 ms of spiking data after stimulus onset on 14 cortical channels after each stimulus onset for analysis. This time window is much larger than the majority of the response from microstimulation that occurs within 100 ms of stimulation.

The goal of this analysis is to assess the dependence of the neural response on the stimulation parameters. However, first, we investigate the significance of the estimated dependence values using bootstrap sampling. There are 100 bootstrap trials with 40 trials, out of 140, for each stimulation setting. We use the proposed estimator with the Victor Purpura spike distance as metric across a range of temporal precision q values. Euclidean distance is used in the two dimensional

stimulation space of amplitude and duration. Both generalized association and mutual information are calculated for each sample and for a shuffled version that destroys the dependence between stimulation and spiking. We show the mean values of GMA and the 0.95 quantiles of the corresponding shuffled data in Figure 8. It is evident that these values are relatively low, but still significant on all 14 channels over all temporal precisions. Also, we observe a clear peak around 50-100 Hz, which implies that the precision relevant to the stimulation is within 10 to 20 ms. We set the value of q to 50.

Since the stimulation is discrete valued, a conditioned dependence analysis can be performed to investigate the dependence between spiking activity and a particular parameter while fixing the other parameter to a certain value. We consider this approach to investigate the effect of effect of amplitude and duration on an individual basis. We show the conditional dependence (i.e., the dependence between spiking activity and amplitude keeping the duration fixed, and the dependence between spiking activity and duration keeping the amplitude fixed), the marginal dependence (i.e., the dependence between spiking activity and amplitude, and the dependence between spiking activity and duration), and the joint dependence (i.e., the dependence between spiking activity and amplitude–duration together) for all the 14 channels in Figure 9. As expected, we observe that the joint dependence is always statistically significant on all channels. However, the marginal and conditional dependences, on the other hand, reveal more interesting structures.

We observe that the duration parameter alone (i.e., from the marginal dependence perspective) does not have any effect on the spiking activity, whereas the amplitude parameter alone imposes a strong dependence on the spiking activity. However, it should be noted that our observation is only limited to the particular values of duration parameters used in the experiment i.e., in the range 125-250μs. Moreover, the conditional dependence between the spiking activity and the
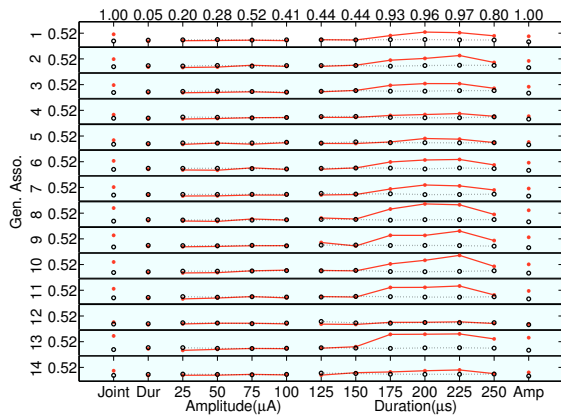
7

Fig. 9.   Experiment IV: Conditional, joint, and marginal generalized association for each of the 14 channels. The black/dotted-line/unfilled is the 0.95 quantile of the surrogate values, whereas the red/solid-line/filled is the 0.5 quantile of the actual values. The values at the top are the average power over the 14 channels, the null hypothesis being independence. It is easy to observe that the dependence between spiking activity and joint of amplitude–duration or marginal of amplitude is strong (i.e., power 1) whereas dependence between spiking activity and marginal of duration is insignificant (i.e., power 0.05). Also, conditioned on the amplitude values, the dependence between spiking activity and duration is not very significant (i.e., highest power 0.52), whereas conditioned on certain duration values (i.e., from 175 to 225 μs) the dependence between spiking activity and amplitude is very strong (i.e., power more than 0.9).

duration given the amplitude is not significant (or relatively low) i.e., given the range of duration under investigation it is rather immaterial for how long a certain amplitude has been applied. A similar situation is also observed when the dependence between spiking activity and amplitude is assessed given particular values of durations, but only upto an extent. That is, we observe that given a small duration spiking activity does not depend on the amplitude too much. However, for a longer duration spiking activity becomes very sensitive to the amplitude applied. These observations provide more insight in the effect of these parameters on the spiking activity. Thus, they can be used to find the range of durations where there is significant dependence between the stimulation amplitude and the spiking response, and that in turn, can be exploited for further refining the experimental design in this study.

V. Summary and future work

In this paper, we have explored a novel understanding of dependence in the context of two random variables that assume values in arbitrary metric spaces. We have derived a simple estimator, and have applied this approach in the context of assessing dependence between sets of stimuli and spike trains in both simulated and experimental data. The simulations reveal promising result, and show that the proposed methodology can indeed be used for neuroscientific exploration.

However, certain areas of the proposed approach still need some attention. For example, in this paper, we have only explored the Victor Purpura metric as a distance metric between two spike trains. However, there exists a number

of different spike train metrics in the literature that can be potentially used in this framework. Next, the proposed method only addresses a measure of association and not a measure of dependence in strict sense since it does not guarantee that two random variables are independent when it achieves minimum value. The exact implication of this feature still remains unexplored. Finally, we have applied our method for exploratory purposes. The true potential of this approach will be investigated in the context of further applications.

References

[1] A. Engel, P. Fries, and W. Singer, "Dynamic predictions: oscillations and synchrony in top-down processing," *Nature Reviews Neuroscience*, vol. 2(10), pp. 704–16, 2001.

[2] B. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[3] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: exploring the neural code*. Cambridge, MA, USA: MIT Press, 1999.

[4] C. Shannon and W. Weaver, *The mathematical theory of communication*. Urbana, Illinois: University of Illinois Press, 1949.

[5] J. Shlens, M. B. Kennel, H. D. I. Abarbanel, and E. J. Chichilnisky, "Estimating information rates with confidence intervals in neural spike trains," *Neural Computation*, vol. 19, no. 7, pp. 1683–1719, 2007. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/neco.2007.19.7.1683

[6] C. Granger and J. L. Lin, "Using the mutual information coefficient to identify lags in nonlinear models," *Journal of Time Series Analysis*, vol. 15-4, pp. 371 – 384, 1991.

[7] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 69, no. 6 Pt 2, June 2004.

[8] J. D. Victor and K. P. Purpura, "Metric-space analysis of spike trains: theory, algorithms and application," *Network: Computation in Neural Systems*, vol. 8, no. 2, pp. 127–164, 1997.

[9] J. D. Victor, "Binless strategies for estimation of information from neural data," *Phys. Rev. E*, vol. 66, no. 5, p. 051903, Nov 2002.

[10] D. D. Mari and S. Kotz, *Correlation and Depedence*. London: Imperieal College Press, 2001.

[11] J. D. Victor and K. P. Purpura, "Nature and precision of temporal coding in visual cortex: a metric-space analysis," *Journal of Neurophysiology*, vol. 76, no. 2, pp. 1310–1326, 1996. [Online]. Available: http://jn.physiology.org/content/76/2/1310.short

[12] J. D. Victor, "Spike train metrics," *Current Opinion in Neurobiology*, vol. 15, pp. 585–592, 2005.

[13] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15(1), pp. 72–101, 1904.

[14] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30(1), pp. 81–93, 1938.

[15] J. Friedman and L. Rafsky, "Graph-theoretic measures of multivariate association and prediction," *The Annals of Statistics*, vol. 11, no. 2, pp. 377–391, 1983.

[16] M. C. W. van Rossum, "A novel spike distance," *Neural Computation*, vol. 13, pp. 751–763, 2001.

[17] A. J. Dubbs, B. A. Seiler, and M. O. Magnasco, "A fast $\mathcal{L}_p$ spike alignment metric," *Neural Computation*, vol. 22, no. 11, pp. 2785–2808, 2010.

[18] D. Aronov, "Fast algorithm for the metric-space analysis of simultaneous responses of multiple single neurons," *Journal of Neuroscience Methods*, vol. 124, no. 2, pp. 175 – 179, 2003.

[19] D. Goldberg, J. Victor, E. Gardner, and D. Gardner, "Spike train analysis toolkit: Enabling wider application of information-theoretic techniques to neurophysiology," *Neuroinformatics*, vol. 7, pp. 165–178, 2009, 10.1007/s12021-009-9049-y. [Online]. Available: http://dx.doi.org/10.1007/s12021-009-9049-y

[20] J. H. Macke, P. Berens, A. S. Ecker, A. S. Tolias, and M. Bethge, "Generating spike trains with specified correlation coefficients," *Neural Comput.*, vol. 21, pp. 397–423.