A Comparison of Euclidean metrics in spike train space

Sergiusz Wesolowski[1], Wei Wu[2]

[1] *Dept. of Mathematics, Florida State University, Tallahassee FL,USA*

[2] *Dept. of Statistics, Florida State University, Tallahassee FL, USA,*

Introduction

Spike trains are observables when investigating neural activity - represent the response of a neuron to stimuli and are often modeled as realizations of stochastic point processes. The spike train space is non-euclidean, recently, however, two L_2 - like distances were introduced on that space:
the **Elastic distance** and **Generalized Victor-Purpura (GVP) distance**.

On this poster we briefly review these two distances and run several comparisons, including construction of the summary statistics, corresponding in ideas to mean and variance as well as classification capabilities. To allow comparisons between metrics we propose an efficient algorithm for GVP summary statistics.

Properties

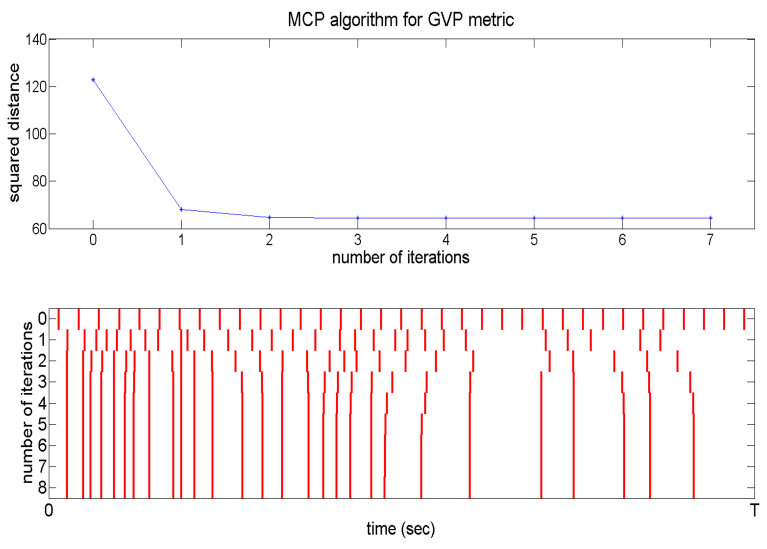
Similarities: Both metrics:

- define a proper Euclidean distance,
- account for temporal structure of spike trains by incorporating term corresponding to differences in spike positions,
- have same computational cost $O(nm)$.

Differences:

- Elastic metric can be viewed as generalization of van Rossum metric, whereas GVP is the generalization of VP metric,
- The temporal structure differences are incorporated with ISIs in Elastic metric and pure spike positions in GVP,

Picture to the right: Shows the convergence rate of the MCP algorithm (Top, blue curve) and temporal mean adjustments during algorithm iterations (Bottom, red stripes correspond to spikes).



MCP algorithm for GVP mean

1. Initialize mean S with n_S uniformly distributed spikes, ($n_S = \max\{n_1, \dots, n_N\}$)
2. **Matching Step**: Use the dynamic programming to find warping γ^k to S. $\gamma^k =$

$$argmin_{\gamma}[m + n - 2\sum_{j=1}^{n}\sum_{i=1}^{m}\mathbb{I}_{x_i=\gamma(y_j)} + \lambda^2 \sum_{x_i=\gamma(y_j)} (x_i - y_j)^2]$$

3. **Centering Step**:
   (a) Find the optimal position of spikes in current mean S and encode it in a warping function $\bar{\gamma}^{-1}$. New spike positions are averaged over all spikes matched with the spike investigated in S. Then the new spike position $\bar{s}_l$ is the average of positions of matched spikes: $\bar{s}_l = \frac{1}{|M_l|}\sum_{x \in M_l} x$
   (b) Update to the new mean S to $\bar{S} = \bar{\gamma}^{-1}(S)$.
4. **Pruning Step**: Remove spikes from the proposed mean $\bar{S}$, that are matched less then $N/2$ times.
   (a) For each spike s_k in $\bar{S}$, count the number of times s_k appears in $\{\gamma^i(\bar{S}^i)\}_{i=1}^N$.
   (b) Remove the spikes from $\bar{S}$ which appear at most $N/2$ times in $\{\gamma^i(\bar{S}^i)\}_{i=1}^N$. Denote the new set $\bar{S}^*$.
5. Safety check: To avoid being stuck in local minimum check if addition or/and deletion of specific single spike will improve the mean.
   (a) Check $\hat{S}^*$ as $\bar{S}^*$ except one spike with minimal number of appearances if improves the sum of squares
   (b) Check $\hat{\;}S^{*}$" as the current mean with one spike added at random if improves the sum of squares
6. Mean Upadate: Update S with proposition that minimizes the sum of squares among $\{\bar{S}^*;\ \hat{S}^*;\ \hat{\;}S^*\}$.
7. Go back to (2) if break condition not achieved.

Notation and definitions

[Mean:] For spike trains $S_1, S_2, \cdots, S_N \in \mathcal{S}$, Karcher mean under a metric d ($d_{elastic}$ or d_{GVP}) is:
$S^* = arg\min_{S\in\mathcal{S}} \sum_{k=1}^N d[\lambda](S_k, S)^2$.

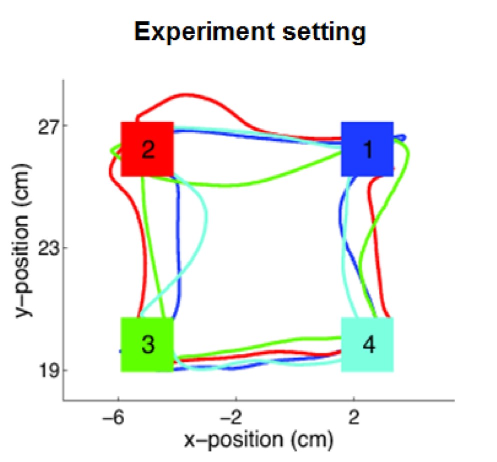
[Spike train:] S is a spike train with spike times $0 < s_1 < s_2 < \cdots < s_M < T$, where $[0, T]$ denotes the recording time domain. We denote this spike train as
$S = (s_j)_{j=1}^M = (s_1, s_2, \dots, s_M)$.

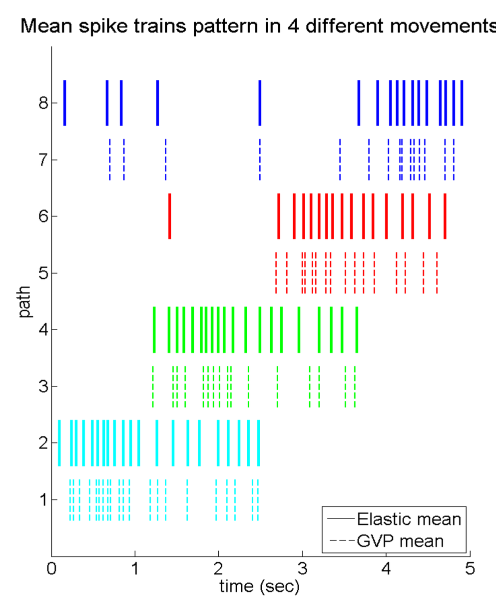
[Warping:] Let Γ be the set of all time warping functions, where a time warping is defined as an orientation-preserving diffeomorphism of the domain $[0, T]$. That is, $\gamma : [0, T] \to [0, T]$ is a time-warping if, in addition to being continuous and (piecewise) differentiable, it satisfies these 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. By applying $\gamma \in \Gamma$ on a spike train $S = (s_j)_{j=1}^M$, one can get a warped spike train $\gamma(S) = (\gamma(s_k))_{k=1}^M$.

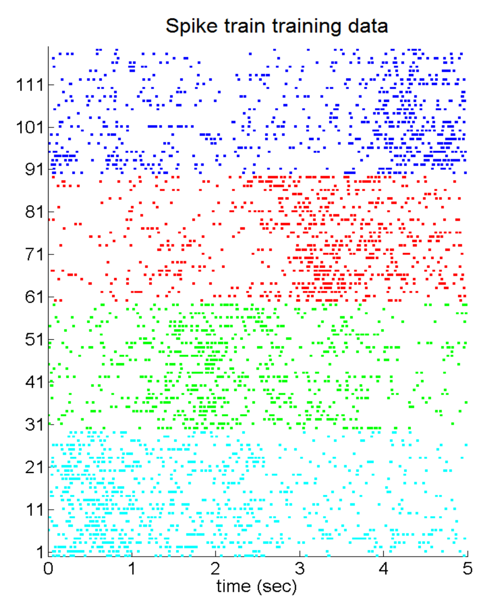
[GVP:] GVP distance between $X = (x_i)_{i=1}^M$ and $Y = (y_j)_{j=1}^N$ is given in the following form:
$d_{GVP}[\lambda](X, Y) = \min_{\gamma\in\Gamma}\left(E_{OR}(X, \gamma(Y)) + \lambda\sum_{\{i,j:x_i=\gamma(y_j)\}}(x_i - y_j)^2\right)^{1/2}$

[Elastic:] the elastic Euclidean metric between $X = (x_i)_{i=1}^M$ and $Y = (y_j)_{j=1}^N$ is given in the following form:
$d_{elastic}[\lambda](X, Y) = \min_{\gamma\in\Gamma}\left(E_{OR}(X, \gamma(Y)) + \lambda\int_0^T(1 - \sqrt{\dot{\gamma}(t)})^2 dt\right)^{1/2}$,
where $E_{OR}(\cdot, \cdot)$ is the cardinality of the Exclusive OR - number of unmatched spike times.

Comparisons on real data







Spike train from MI neuron recorded while performing 4 types of hand movements.

- The dataset consists of 240 spike trains which correspond to four different trajectories of a hand movement along a rectangle (60 spikes for each) as given on figure above. The trajectories follow the contour of the rectangle, but differ in starting point: (Blue, Red, Green, Cyan).
- Half of the dataset is used to learn the mean spike train patterns for 4 types of trajectories.
- The class for a new spike train X is assigned to one of four classes (Blue, Red, Green, Cyan) based on normalized distance to corresponding mean $\mu \in \{\mu_{Blue}, \mu_{Red}, \mu_{Green}, \mu_{Cyan}\}$.

$$arg\min \frac{d(\mu, X)}{\sigma}| \quad \mu \in \{\mu_{Blue}, \mu_{Red}, \mu_{Green}, \mu_{Cyan}\}$$

Table 1: **Classification efficiency**

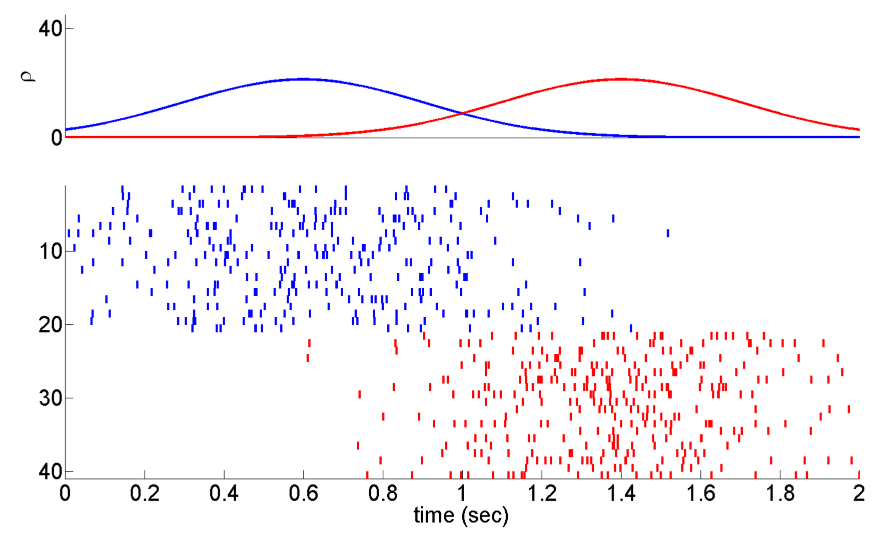
% of succesfull classification for classes:	Blue	Red	Green	Cyan	Overall
$\lambda = 65$					
using Elastic metric	66,7	90	80	56,7	73,3
usign GVP metric	60	83,3	96,7	100	**85**
$\lambda = 40$					
using Elastic metric	83,3	90	90	90	**88,3**
usign GVP metric	70	83,3	90	90	83,3

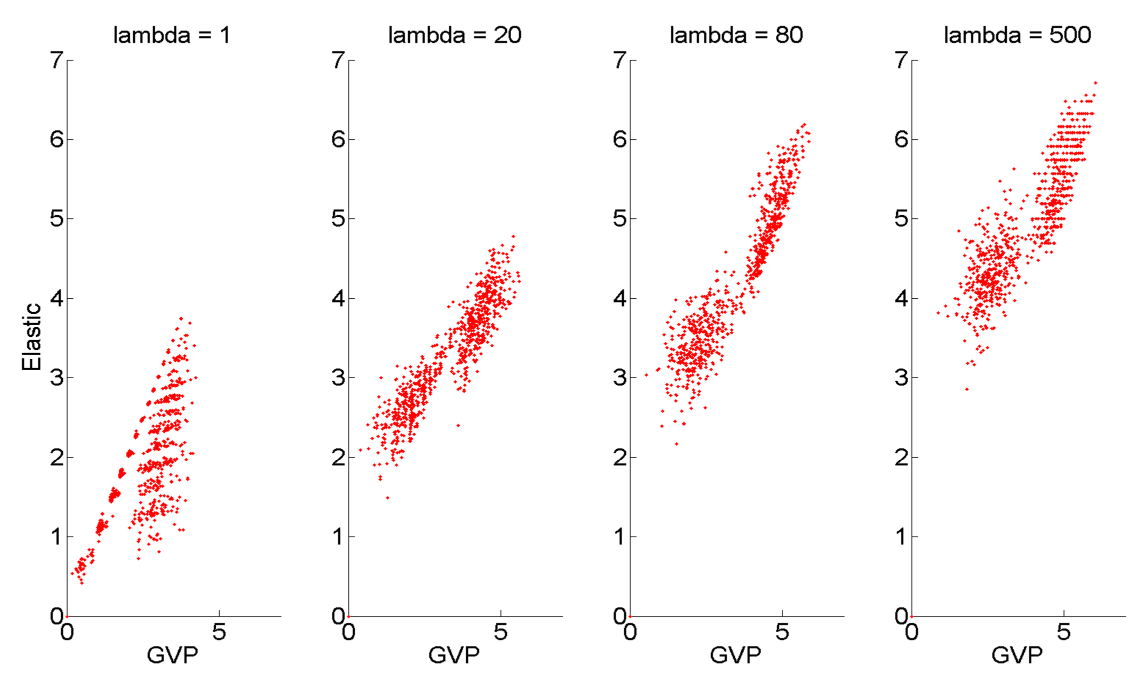
Comparisons on simulated data

- Simulated 40 spike trains from a Poisson point process on interval $[0, 2]$ with Gaussian intensity function, as seen on picture to the right.

$$\rho(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-m)^2}{2\sigma^2}} \qquad (1)$$

- 20 spikes were generated with $\mu = 0.6$ and 20 with $\mu = 1.4$, σ^2 was set equal in both cases to 0.3.





- Each red dot represents a pair of spike trains for which GVP and Elastics distance are the coordinates,
- In all cases $\lambda \in \{1, 20, 80, 500\}$ - almost linear relation between metrics,
- In all cases the metrics form two clusters corresponding to spike train pairs that have small distances (are generated within a group) and large distances (between the two groups),
- Discretisation of Elastic metric occurs when $\lambda = 1$, due to the fact that the $[m + n - 2\sum_{j=1}^n\sum_{i=1}^m \mathbb{I}_{x_i=\gamma(y_j)}]$ component is dominating, when $\lambda = 500$ the penalty for moving spikes is so big that it's not worth matching any spikes, again the "matching" term sets up the discretization, however it is no longer dominating.

Classification efficiency w.r.t metric parameter λ

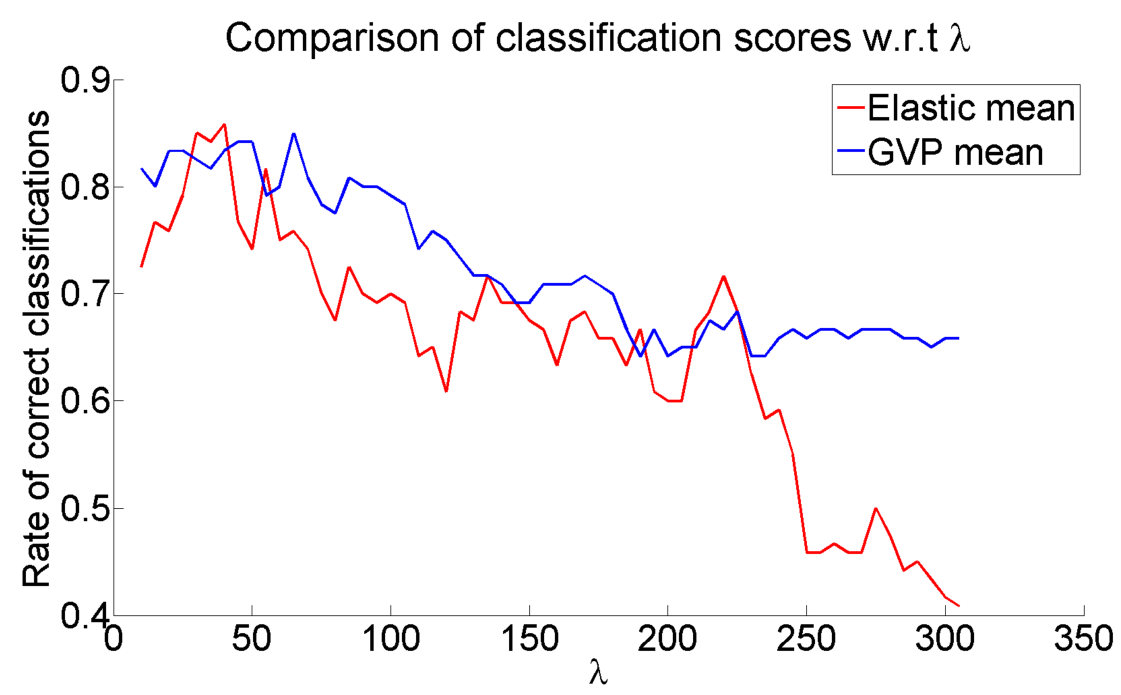


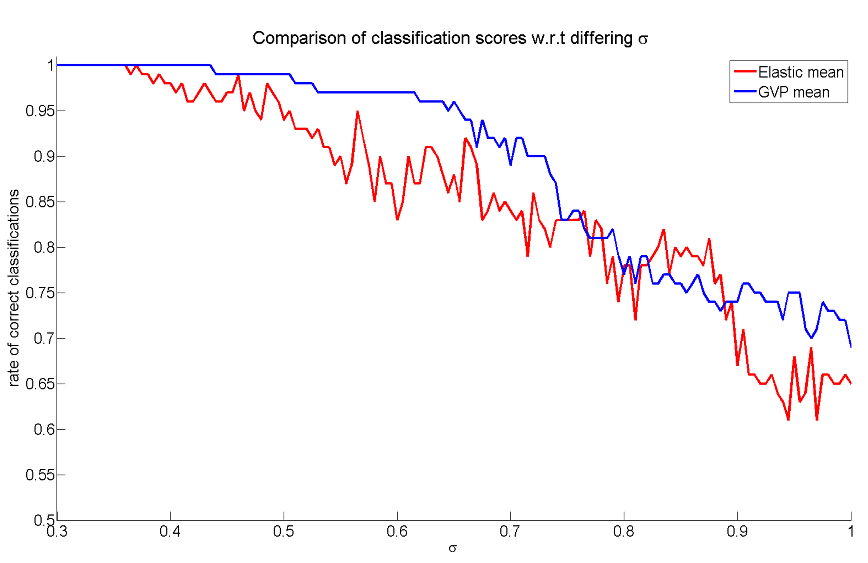
Figure 1: Classification scores for a range of parameter $\lambda \in (10, 200)$.

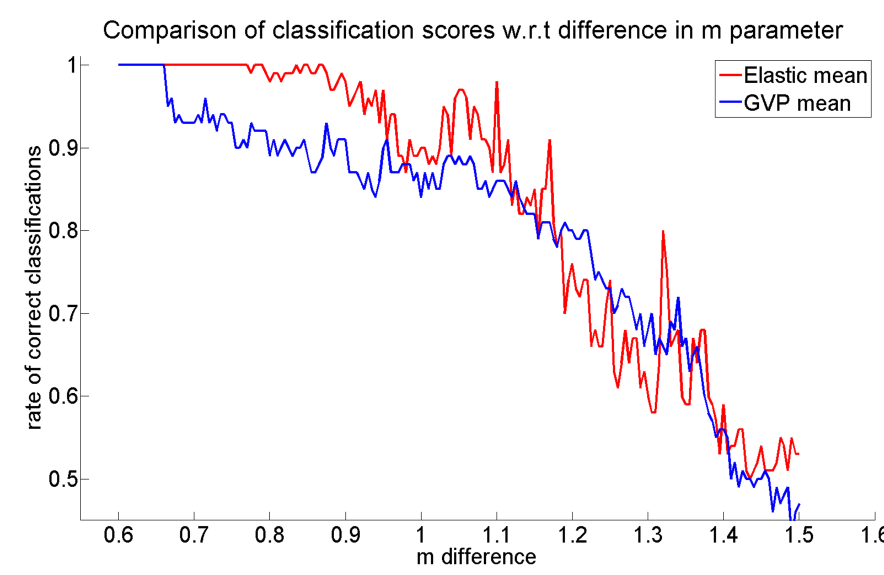
For this dataset we have

- For most parameter values λ GVP performs better then Elastic,
- However, maximum correct classification score is held by Elastic metric.
- Towards large values of lambda classification efficiency drops. Elastic metric assignes random values, but GVP maintains the level of around 70%.

Choice of λ may depend on data type that is investigated.

Classification efficiency w.r.t model parameters: $m,\ \sigma$





80 spikes generated according to Poisson point process model as training and test datasets. Two settings investigated:

[1] Groups hold the same $\sigma = 0.15$, but differ in m: from 0.5 to 1.5.

[2] Groups have fixed, but different m values: $m_1 = 0.8,\ m_2 = 1.2$, σ is varied in range $[0.1, 0.5]$, σ is same for both groups.

- The class for a new spike train X is assigned based on normalized distance to corresponding mean $\{\mu_1, \mu_2\}$: $\min \frac{d(\mu, X)}{\sigma}| \quad \mu \in \{\mu_1, \mu_2\}$

Results:

- When shift between intensity functions is fixed GVP outperforms Elastic even for large variance
- When the variance is fixed and intensity functions are being shifted Elastic outperforms GVP, until classification differences become irrelevant.