

MIT Open Access Articles

A Fast L-p Spike Alignment Metric

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Dubbs, Alexander J., Brad A. Seiler, and Marcelo O. Magnasco. "A Fast L_p Spike Alignment Metric." *Neural Computation* 22.11 (2010): 2785-2808. © 2010 Massachusetts Institute of Technology

As Published: http://dx.doi.org/10.1162/NECO_a_00026

Publisher: MIT Press

Persistent URL: <http://hdl.handle.net/1721.1/60302>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



A Fast \mathcal{L}_p Spike Alignment Metric

Alexander J. Dubbs*

alex.dubbs@gmail.com

Brad A. Seiler

brad@bradseiler.com

Center for Studies in Physics and Biology, Rockefeller University, New York, New York 10065, U.S.A., and Harvard University, Faculty of Arts and Sciences, Cambridge, MA 02138, U.S.A.

Marcelo O. Magnasco

magnasco@rockefeller.edu

Center for Studies in Physics and Biology, Rockefeller University, New York, New York 10065, U.S.A.

The metrization of the space of neural responses is an ongoing research program seeking to find natural ways to describe, in geometrical terms, the sets of possible activities in the brain. One component of this program is spike metrics— notions of distance between two spike trains recorded from a neuron. Alignment spike metrics work by identifying “equivalent” spikes in both trains. We present an alignment spike metric having \mathcal{L}_p underlying geometrical structure; the \mathcal{L}_2 version is Euclidean and is suitable for further embedding in Euclidean spaces by multidimensional scaling methods or related procedures. We show how to implement a fast algorithm for the computation of this metric based on bipartite graph matching theory.

1 Introduction

The analysis of neural signals seeks to “translate” any set of neural impulses into a language we understand (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997). But because the neural signals elicited by the same stimulus are never exactly alike, we need a quantitative means of determining when two neural signals serve the same purpose. In what follows, we restrict our attention to signals from isolated cell electrophysiology and represent a spike train neural signal as a marked point process—an ordered list of spike times labeled by their respective neurons of origin (Aronov & Victor, 2004). To compare spike trains, Victor and Purpura (1996, 1997) introduced the notion of a spike metric, a distance function on the set of

*A. Dubbs is currently at the Mathematics Department, MIT.

spike trains endowing it with the topological properties of a metric space. Spike metrics have been used successfully, to quantify variability in data and characterize neural coding in the visual, auditory, olfactory, taste, and electric sensory systems (Victor, 2005). They have been proposed with the goals of clustering neural signals by stimulus and embedding them in Euclidean space (Victor & Purpura, 1997) via multidimensional scaling (MDS) (Kruskal & Wish, 1978; Bengio, Paiement, & Vincent, 2004). MDS allows one to visualize the geometry of spike data, especially after the use of sophisticated dimensionality-reduction procedures such as local linear embedding (Roweis & Saul, 2000) and Isomap (Tenenbaum, de Silva, & Langford, 2000). It also permits us to use more sophisticated clustering techniques designed to work only in vector spaces, such as PCA-guided K -means (Ding & He, 2004).

One approach to spike metric design for spike trains generated by single neurons uses the rate-coding hypothesis of spike generation—the idea that spike trains induced by the same stimulus are instances of the same variable-rate Poisson process (Rieke et al., 1997). The implication is that to compare two spike trains, one must compare their estimated underlying rate functions. To estimate the rate function of a spike train, represent it as a sum of Kronecker delta functions at the spike times and stream it through a low-pass filter or, equivalently, convolve it with a nonnegative function, or kernel, that integrates to 1. Common choices of kernels include the one-sided exponential tail (van Rossum, 2001), the two-sided exponential tail, the tent function (0 above and below a certain range, and linearly increasing and then decreasing at the same rate within that range), and the normal distribution function (Schrauwen & Campenhout, 2007). The value of the metric between two spike trains is the \mathcal{L}_p norm of the difference between their estimated rate functions ($p \geq 1$). For $p = 2$, such metrics resemble Euclidean distances, and MDS can embed them in (mostly) Euclidean finite-dimensional vector spaces.

However, there is substantial evidence that the time coding of neural signals contains meaning beyond that conveyed by the firing rate (Chase & Young, 2006; Gerstner, Kreiter, Markram, & Herz, 1997; Rieke et al., 1997; Samonds & Bonds, 2004). Based on this fact, Victor and Purpura (1996, 1997) introduced a spike metric equal to the minimum cost of aligning the spikes in two spike trains. Their approach was to generalize methods used for comparing strands of DNA. In particular, they reengineered Sellers's (1974) dynamic programming algorithm for aligning and calculating the evolutionary distance between pairs of DNA sequences to compute a distance between pairs of spike trains. This metric preserves the integrity of individual spikes instead of viewing them as contributions to a rate function. It has the added advantage of being able to compute distances between spike trains that may have contributions from multiple neurons. In the future, it may be generalized to solve the multiple-sequence alignment problem—the question of how to draw correspondences between analogous spikes in

multiple trains at once, as has been done with nucleotides in sequences of DNA (Notredame, 2002).

An issue with the Victor-Purpura metric is that it resembles an \mathcal{L}_1 norm on a vector space. If one uses it to compute all the distances among a group of spike trains elicited from a common stimulus and then embeds those spike trains using the table of distances in a procedure such as MDS, the resulting picture may become complicated. The embedded set of spike trains may have hyperbolic structure that is not present in the stimulus space. We discuss such issues at length in section 7.

Victor, Goldberg, and Gardner (2007) generalized the algorithm used by Victor and Purpura to work for a larger class of cost functions that do not necessarily meet the topological definition of *metric* (e.g., they do not satisfy the triangle inequality). They can be modified to emulate \mathcal{L}_p metrics on the space of spike trains.

We propose a competing, fast algorithm to calculate these \mathcal{L}_p spike metrics, well suited to embedding in \mathbb{R}^n when $p = 2$. Furthermore, we prove that these metrics meet the topological definition, endowing the set of spike trains with important properties of Euclidean spaces. When $p = 2$, embeddings of spike trains using this metric fit in Euclidean space with substantially less difficulty than they do for any other value of $p \geq 1$, so advanced dimensionality-reduction and clustering techniques may be used. We illustrate this point numerically. The procedure we use to calculate our metric is the Hungarian algorithm (Schrijver, 2003), which can be made substantially faster for the specific task of comparing spike trains. The crucial insight behind the Hungarian algorithm was independently discovered by Kuhn (1955) and Munkres (1957). Its function is to solve the minimum-weight matching problem on weighted bipartite graphs. The Hungarian algorithm is a special case of algorithms to solve general matching problems (Papadimitriou & Steiglitz, 1998) and assignment problems (Burkard, 2002). Our version of the algorithm is a special case of algorithms to solve transportation problems on Monge arrays, described in Burkard, Klinz, and Rudolf (1996). The original transportation problem was solved by Hoffman (1963) and formulated by Monge (1781).

2 The Metric

We desire to match spikes in one spike train to corresponding spikes in another spike train to compare the two signals. Two trains fired in response to the same stimulus may have different numbers of spikes. So an alignment of the spikes in the two trains will be inherently imperfect; some spikes may have to be deleted from the trains to make the alignment exact. Victor and Purpura's idea was to break the process of aligning two spike trains into steps, each with its own cost. The metric is equal to the sum of the costs incurred by the most efficient alignment (Victor & Purpura, 1996, 1997). Let $q > 0$. The cost of aligning two spikes on different trains is $q \Delta t$, where Δt

is the distance between them. The cost of deleting a spike on either train is 1. If in two trains the differences between spike times in aligned pairs of spikes are Δt_i , for $i \in \{1, \dots, k\}$, the number of spikes deleted from the first train is D_1 , and the number of spikes deleted from the second train is D_2 , the distance between the two spike trains is

$$\sum_{i=1}^k q \Delta t_i + D_1 + D_2.$$

The metric is defined to be the minimum of this quantity over all ways of aligning the trains. Our generalization is as follows. Let $p \geq 1$; the metric is

$$\min \left[\sum_{i=1}^k q^p \Delta t_i^p + D_1 + D_2 \right]^{1/p}.$$

When $p = 1$, this is the standard Victor-Purpura metric.

We can formulate the metric in a way that is more mathematically specific as follows. Let $\mathbf{x} = \{x_i\}_{i=1}^m$ and $\mathbf{y} = \{y_j\}_{j=1}^n$ be spike trains, or strictly increasing finite sequences of real numbers. Let M be a matching between the two spike trains, a set of ordered pairs $\{(x_{i_1}, y_{j_1}), \dots, (x_{i_k}, y_{j_k})\}$ with no element of \mathbf{x} or \mathbf{y} repeated. Let \mathcal{M} be the set of all matchings. Let the cardinality of M , $|M|$, be the number of ordered pairs contained in M . Define

$$d_{p,q}[M](\mathbf{x}, \mathbf{y}) = \left[\sum_{(x_i, y_j) \in M} q^p |x_i - y_j|^p + (m - |M|) + (n - |M|) \right]^{1/p}. \tag{2.1}$$

Our metric is

$$d_{p,q}(\mathbf{x}, \mathbf{y}) = \min_{m \in \mathcal{M}} d_{p,q}[M](\mathbf{x}, \mathbf{y}).$$

We shall show that we can compute this expression through the Hungarian algorithm (Schrijver, 2003), with additional speed-ups, from $o(mn(m + n)^2)$ time to $o(mn)$ time.

3 Outline of the Algorithm

Our metric is a minimization over all possible matchings. We can map this minimization to a classical problem: weighted bipartite matching. In this problem, we have a graph whose nodes belong to two different classes with no edge connecting members of the same class (bipartite), and each edge of the graph has a weight, or cost. A matching or pairing is a subgraph in

which each node has at most one connection, and the matching with the smallest total cost is sought.

A priori, the number of distinct matchings is enormous. There is a profound result, called the Hungarian algorithm (Schrijver, 2003), that shows that only a small subset of these matchings needs to be examined. This result works by showing that the optimal match with precisely $k + 1$ pairings (cardinality $k + 1$) is related in a rather specific way to the optimal match of cardinality k : the optimal $k + 1$ match differs from the optimal k match by an M -augmenting path. Since there are many fewer such paths than matchings of cardinality $k + 1$, the search is considerably faster.

We shall show that the convexity of our metric translates into a specific property of the graph weights called the Monge property (Burkard et al., 1996). Using the Monge property, we can discard large fractions of the M -augmenting paths as candidates: the paths have to be monotonic and cannot “jump.” We call the resulting, much smaller, set of “ M -augmenting incompressible monotonic paths” the *shifts*. There are far fewer shifts to explore in connecting level k with level $k + 1$ —at most, $m + n - 2k - 1$.

4 The Shift Algorithm

For certain matchings M , we can construct a unique two-row matrix representation, or matching matrix, as follows. The first row contains the elements of \mathbf{x} in increasing order (from left to right), and the second row contains the elements of \mathbf{y} in increasing order. If $(x_i, y_j) \in M$, then x_i and y_j are in the same column of the matrix. If x_i is not matched in M , then it is in the same column as a $*$ symbol. The same goes for unmatched elements of \mathbf{y} . Any unmatched elements of \mathbf{x} and \mathbf{y} together appear in increasing order from left to right in M . Making this matrix takes $o(m + n)$ time. For instance, let $\mathbf{x} = \{b, e, f, h\}$, $\mathbf{y} = \{a, c, d, g, i\}$, where $a < b < \dots < h < i$, with $M = \emptyset$. The matching matrix looks like

$$\begin{bmatrix} * & b & * & * & e & f & * & h & * \\ a & * & c & d & * & * & g & * & i \end{bmatrix}.$$

Now assume b and h from \mathbf{x} are matched with c and g from \mathbf{y} . The new matrix looks like

$$\begin{bmatrix} * & b & * & e & f & h & * \\ a & c & d & * & * & g & i \end{bmatrix}.$$

Define the shift operation on a matching matrix as follows. Delete two $*$'s—one on the top and one on the bottom and such that there exist no other $*$'s

on either row between them. Then shorten the rows to compensate. Below are both possible shifts applied to the matrix above:

$$\begin{bmatrix} * & b & e & f & h & * \\ a & c & d & * & g & i \end{bmatrix}, \begin{bmatrix} * & b & * & e & f & h \\ a & c & d & * & g & i \end{bmatrix}.$$

Main theorem. *If M minimizes $d_{p,q}[M](\mathbf{x}, \mathbf{y})$ over matchings of cardinality $k < \min(m, n)$, then at least one matching that minimizes $d_{p,q}[M](\mathbf{x}, \mathbf{y})$ over matchings of cardinality $k + 1$ is some shift of the matching matrix of M .*

The proof is outlined in section 6 and completed in the appendix. The empty matching always has a matching matrix, and any shift applied to a matching matrix returns a matching matrix. Therefore, we can find the optimal matching of cardinality $k + 1$ by tabulating all shifts on the optimal matching of cardinality k and then picking the one that minimizes $d_{p,q}[M](\mathbf{x}, \mathbf{y})$. Even if there are two optimal matchings, M_1 and M_2 of cardinality $k + 1$, and only M_1 is a shift of the matrix corresponding to the optimal matching of cardinality k , at least one optimal matching of cardinality $k + 2$ is a shift of M_1 .

The time cost of the naive implementation of the algorithm is as follows. Making the matching matrix takes $o(m + n)$ time. Finding the distance-minimizing matching of order $k + 1$ by calculating the expected gain or loss to the distance function from every possible shift on the distance-minimizing matching of order k takes $o(m + n)$ time. There are $o(\min(m, n))$ rounds of searching, so we get $o((m + n) \min(m, n))$ time.

We have implemented two key speed-ups. $\min(m, n)$ rounds of shifting is the worst-case scenario; by theorem 3 the algorithm can terminate as soon as all possible shifts increase the value of $d_{p,q}[M](\mathbf{x}, \mathbf{y})$. After making the first matching matrix, cut it in two between any two spikes (on either train) separated by more than $2^{1/p}/q$. By corollary 4, the optimal matching is the union of the two optimal matchings found by repeatedly shifting these two submatrices. To prove the algorithm's correctness, we need a result from graph theory.

5 The Hungarian Algorithm

For completeness we condense chapter 3 of Schrijver (2003) below. Here we demonstrate one way in which the optimal matching of cardinality $k + 1$ is related to the optimal matching of cardinality k . This is the first step toward showing that the relationship is in fact a shift.

Definitions. *A graph $G = (V, E)$ is a set of vertices V and edges E , where an edge e is defined to be a pair of two vertices, $e = \{v, v'\}$, where $v, v' \in V$. G is complete*

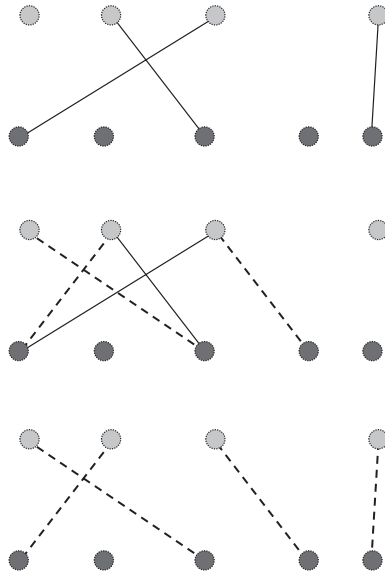


Figure 1: M -augmenting paths. Two sets of spikes (top, light gray and bottom, dark gray) are to be matched; the horizontal coordinate depicts time. (Top) A matching M of cardinality 3 between two sets is depicted as black edges joining members of the two sets. (Middle) An M -augmenting path P , with edges that already were in M in solid and new edges in dashed. The path should start and end in dashed (new) edges and alternate between dashed (new) and solid (old). It need not contain any solid edges at all. (Bottom) A new matching $M \Delta P$ is obtained as the symmetric difference between the first two graphs.

bipartite if $V = A \cup B$ where $A \cap B = \emptyset$ and every edge e contains exactly one element of each of A and B . G is edge-weighted if there exists a weight function $w: E \rightarrow \mathbb{R}$ on the edges of G . From now on, all graphs mentioned will be complete bipartite and edge-weighted. A matching M on G is a subset of E such that no two edges in M share a vertex. Let $w(M) = \sum_{e \in M} w(e)$ be the weight of matching M .

A path P on G is an ordered subset of distinct elements of E , (e_1, \dots, e_t) , such that e_i and e_{i+1} share a vertex for all $1 \leq i < t$ but e_1 and e_t do not share a vertex. A path P is M -augmenting for a matching M if (1) t is odd, (2) $e_2, e_4, \dots, e_{t-1} \in M$, and (3) e_1 and e_t each share only one vertex with an element of M . For an example of an M -augmenting path, see Figure 1. For an edge e , define its length function $l(e)$ to be $w(e)$ if $e \in M$ and $-w(e)$ if $e \notin M$. For a path P , let $l(P) = \sum_{e \in P} l(e)$.

If $C, D \subseteq E$, $C \Delta D$ contains all edges that are in exactly one of C and D . It follows that if P is M -augmenting, then $M \Delta P$ is a matching of cardinality $|M| + 1$.

Theorem 1 (Schrijver, 2003, section 3.5, proposition 1). Let P be an M -augmenting path of maximum length. If M is a matching of the minimum weight of

cardinality k , then $M' = M \Delta P$ is a matching of the minimum weight of cardinality $k + 1$. The proof is in the appendix.

6 Proof of Shift Algorithm

Here we outline the proof that augmenting paths of maximum length correspond to shifts. The details of the proof are in the appendix.

Lemma 1. *Computing $d_{p,q}(\mathbf{x}, \mathbf{y})$ is equivalent to finding a minimum weight matching on the complete bipartite weighted graph with vertex sets $A = \{a_1, \dots, a_m\}$ and $B = \{b_1, \dots, b_n\}$ and weight function $w(a_i, b_j) = q^p |x_i - y_j|^p - 2$.*

Proof. Every matching M on this graph is associated with a matching M (abusing notation) between spike trains x and y . To minimize $d_{p,q}[M](\mathbf{x}, \mathbf{y})$ over M , it is equivalent to minimize

$$\sum_{(x_i, y_j) \in M} (q^p |x_i - y_j|^p - 2) = \sum_{(a_i, b_j) \in M} w(a_i, b_j) = w(M).$$

Now we need to find an analog for shifting a matching matrix in terms of performing an operation on graphs. It turns out that the analog is taking the symmetric difference between M and a special type of M -augmenting path, which we will also call a shift. It is a path of the form

$$(\{a_i, b_j\}, \{a_i, b_{j+1}\}, \{a_{i+1}, b_{j+1}\}, \dots, \{a_{i+N}, b_{j+N}\})$$

or

$$(\{a_i, b_j\}, \{a_{i+1}, b_j\}, \{a_{i+1}, b_{j+1}\}, \dots, \{a_{i+N}, b_{j+N}\}),$$

with a couple of additional constraints. In the first case, there are no unmatched elements of \mathbf{x} between x_i and y_j and no unmatched elements of \mathbf{y} between x_{i+N} and y_{j+N} , and in the second case, there are no unmatched elements of \mathbf{y} between x_i and y_j and no unmatched elements of \mathbf{x} between x_{i+N} and y_{j+N} . Figure 2 shows an M -augmenting path that is a shift. In this and all subsequent figures, we place the sets of nodes corresponding to the spikes on trains \mathbf{x} and \mathbf{y} on parallel lines, and nodes farther to the right correspond to spikes occurring later in time than the spikes associated to nodes on their left. We can now restate the main theorem:

Main theorem, restated. *Let a bipartite graph have vertex sets $A = \{a_1, \dots, a_m\}$ and $B = \{b_1, \dots, b_n\}$ and weight function $w(a_i, b_j) = q^p |x_i - y_j|^p - 2$. Let M minimize $w(M)$ over matchings of cardinality $k < \min(m, n)$. Then there exists an M -augmenting path P that is a shift such that $M \Delta P$ minimizes $w(M)$ over matchings of cardinality $k + 1$.*

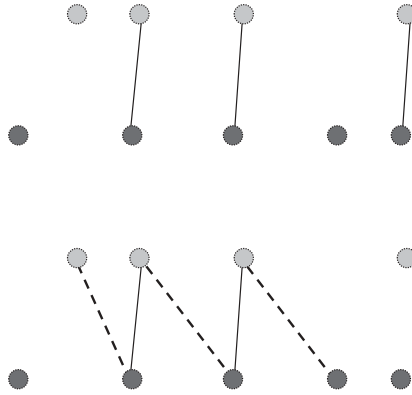


Figure 2: Shifts are a subset of M -augmenting paths. The matching on top has been drawn with the convention stated below lemma 1: light gray nodes corresponding to spikes in train x are on top, dark gray nodes corresponding to spikes in train y are on bottom, and nodes farther to the right correspond to spikes occurring later in time. Figures 1 through 6 follow this convention. The M -augmenting path P on the bottom contains, as in Figure 1—solid edges in M and new dashed edges. Because P moves monotonically forward in time and does not skip any node, it is a shift.

Lemma 2. *Strict Monge property: Let $x_{i_1} < x_{i_2}$ and $y_{j_1} < y_{j_2}$. For $p > 1$,*

$$|x_{i_1} - y_{j_1}|^p + |x_{i_2} - y_{j_2}|^p < |x_{i_1} - y_{j_2}|^p + |x_{i_2} - y_{j_1}|^p.$$

The proof uses convexity and is in the appendix.

Lemma 3. *Let $p > 1$. If M is a minimum weight matching of cardinality k containing edges $\{a_{i_1}, b_{j_1}\}$ and $\{a_{i_2}, b_{j_2}\}$, then either $i_1 < i_2$ and $j_1 < j_2$ or $i_1 > i_2$ and $j_1 > j_2$.*

Proof. Remove edges $\{a_{i_1}, b_{j_1}\}$ and $\{a_{i_2}, b_{j_2}\}$ and replace them with edges $\{a_{i_1}, b_{j_2}\}$ and $\{a_{i_2}, b_{j_1}\}$ to form the matching M' of cardinality k . By lemma 2,

$$\begin{aligned} w(M') - w(M) &= q^p |x_{i_1} - y_{j_2}|^p + q^p |x_{i_2} - y_{j_1}|^p - q^p |x_{i_1} - y_{j_1}|^p \\ &\quad - q^p |x_{i_2} - y_{j_2}|^p. \end{aligned}$$

Since M is of minimum weight, this expression must be positive. Lemma 2 implies that either $i_1 < i_2$ and $j_1 < j_2$ or $i_1 > i_2$ and $j_1 > j_2$. This lemma can be pictorially understood as meaning that if M is a minimum weight matching of cardinality k , it contains no edges that cross, using our convention for drawing figures stated below lemma 1. See Figure 3 for a demonstration of lemma 3.

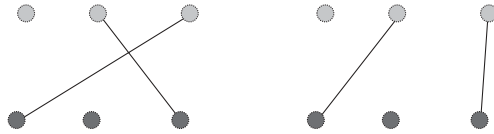


Figure 3: Convexity forbids edges from crossing. The matching depicted on the left cannot be a minimum weight matching of cardinality 2 because it violates lemma 3, since it has crossing edges. The matching depicted on the right could be a minimum weight matching of cardinality 2, as it does not violate lemma 3.

Theorem 2. *Let $p > 1$, let M be a minimum weight matching of cardinality k , and let P be an M -augmenting path of maximum length. Then P is of the form*

$$(\{a_{i_1}, b_{j_1}\}, \{a_{i_1}, b_{j_2}\}, \{a_{i_2}, b_{j_2}\}, \dots, \{a_{i_l}, b_{j_l}\})$$

or

$$(\{a_{i_1}, b_{j_1}\}, \{a_{i_2}, b_{j_1}\}, \{a_{i_2}, b_{j_2}\}, \dots, \{a_{i_l}, b_{j_l}\}),$$

where the sequences $\{a_i\}$ and $\{b_i\}$ are either strictly increasing or strictly decreasing.

The exact proof is in the appendix. It can be understood using diagrams. We want to show that P does not turn back on itself; it moves either forward or backward through the graph. If it should decide to change direction, one of two things happens. It might make use of two edges in M that cross, contradicting lemma 3 (see Figure 4, top). If not, it uses two edges that are both not in M that cross (if it uses an edge in M and an edge not in M that cross, one of the previous two problems necessarily occur). In the second case, both of those edges will be in $M \Delta P$. But if P is of maximum length, by theorem 1, $M \Delta P$ is a minimum matching of cardinality $k + 1$, which also cannot contain crossing edges by lemma 3 (see Figure 4, bottom). This is a contradiction, and it follows that P cannot turn around.

We are now ready to sketch the proof of the main theorem for $p > 1$. It follows for $p = 1$ by real analysis. Both proofs are in the appendix.

Let M be a minimum weight matching of cardinality k and P be an M -augmenting path of maximum length. P is of the form described by theorem 2. Assume without loss of generality that the vertex at which P starts is $a_{i'}$, and the vertex at which it ends is $b_{j'}$. Two things could go wrong to prevent transforming M into $M \Delta P$ from being a shift on M 's matching matrix. One is that there exists an unmatched spike on either spike train between $x_{i'}$ and $y_{j'}$. In that case, it can be shown that if M is the minimum matching of cardinality k , then $M \Delta P$ is not the minimum matching of cardinality $k + 1$; it can be improved on by replacing an edge (see Figure 5). The second possible problem is that there exists an edge in M that some edge in P crosses. Then,

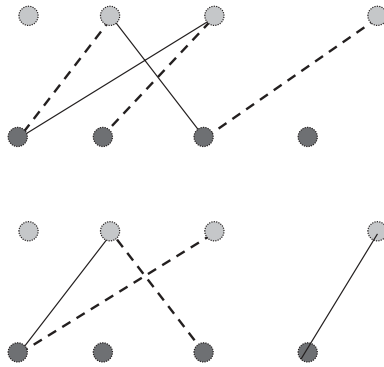


Figure 4: A shift cannot turn back on itself, because either the initial matching or the final matching would have crossing edges. (Top) M is the matching depicted by the two solid lines. The complete set of lines forms an M -augmenting path P . M is not a minimum weight matching of cardinality 2 by lemma 3, contradicting the hypothesis of theorem 2, so we do not need to consider this case in its proof. (Bottom) Here M is the matching of cardinality 2 depicted by the two solid black lines. The two dashed lines and the contiguous solid line form an M -augmenting path. $M \Delta P$ contains crossing edges, so by lemma 3, it is not the minimum weight matching of cardinality 3. Thus by theorem 1, P is not an M -augmenting path of maximum length.

it can be shown that M or $M \Delta P$ contains crossing edges, contradicting lemma 3 (see Figure 6). With these two possibilities eliminated, P could only be an M -augmenting path corresponding to a shift.

Theorem 3. *Let M be a minimum weight matching of cardinality k . If a minimum weight matching of cardinality $k + 1$ has greater weight than M , then M is a minimum weight matching over all matchings of any cardinality. The proof is in the appendix.*

Corollary 4. *Let M_k denote the minimum weight matching of cardinality k . If M_{k_0} is the minimum weight matching over matchings of all cardinalities, then no M_k contains a positive-weighted edge for $k \leq k_0$.*

Proof. The proposition is vacuous if $k_0 = 0$. So assume the statement is true for $k, 0 < k < k_0$; we will prove it for $k + 1$. Assume that M_{k+1} has an edge e of positive weight. Then $M_{k+1} \Delta \{e\}$ is a matching of cardinality k , so $w(M_{k+1} \Delta \{e\}) \geq w(M_k)$. But $w(M_{k+1} \Delta \{e\}) < w(M_{k+1}) \leq w(M_k)$ by theorem 3, a contradiction. The implication of this corollary is that if any two adjacent columns of a matching matrix contain entries that are separated by more than $2^{1/p}/q$, the matrix can be cut in two and the two halves optimized separately.

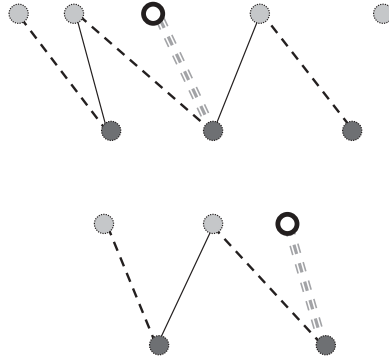


Figure 5: Shifts cannot skip previously unmatched vertices. (Top) M is the matching composed of the two solid lines, and the M -augmenting path P comprises those edges plus the three dashed ones. P cannot be maximum length because it skips the vertex highlighted as white with a solid outline. The skipped vertex can be joined to a unique bottom vertex without crossing lines, as shown by the triple dashed edge. This edge must be shorter than one of its two neighbors in the path; if it was shorter than the solid edge, the original matching would not be minimal. If it was shorter than the dashed edge, then replacing the triple-dashed edge for the dashed edge would give a new matching that is better than the old one, but the difference is not an M -augmenting path; by theorem 1, there must be an even better one. (Bottom) The last edge in a path cannot jump over a vertex. M is the single solid black edge, and the M -augmenting path P comprises that edge and the two dashed ones. $M\Delta P$ would have lower weight if the triple dashed edge replaced its nearest dotted edge in P , so P is not the M -augmenting path of maximum length, by theorem 1. (In this figure, the optimal shift would be just the triple dashed edge.)

7 Embeddability of the \mathcal{L}_1 Metrics

We shall be having the following procedure in mind. Data are gathered from an experiment in the form of spike trains; a table of distances between them is generated according to some spike metric, and this table of distances is analyzed to try to elucidate some underlying regularity in the neural responses. The simplest forms of analysis of distance tables are in the form of Euclidean embedding methods such as MDS. The central requirement we shall pose of this procedure is that it should not generate spurious complexity: neurophysiology is complex enough as it is, without making it look even more complicated.

Three main issues affect the embeddability of metrics other than \mathcal{L}_2 in Euclidean space, causing embeddings to become spuriously complex: degeneracy, noncompactness, and hyperbolicity.

Degeneracy refers to the fact that some metrics “lose” information, so that many configurations map to the same table of distances. Figure 7

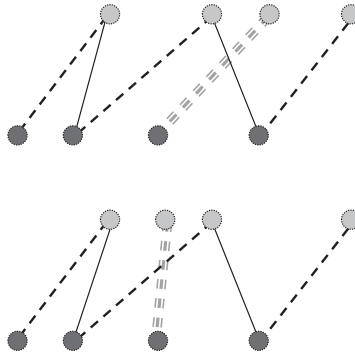


Figure 6: Shifts cannot skip previously matched vertices, because the edges from these previously matched vertices must cross either an edge in the old matching or in the new matching. (Top) M is the matching composed of the two solid lines plus the triple dashed line, which represents an old edge not in the path; an M -augmenting path is superimposed. The dashed line crosses a black line in the path; therefore the old matching was not minimal. (Bottom) Same convention. M is the matching composed of the two solid black lines and the triple dashed line. An M -augmenting path P is composed of the three dashed and two solid black lines. $M\Delta P$ is the matching composed of the triple dashed lines and the one dashed line. $M\Delta P$ has crossing edges.

illustrates the point using the \mathcal{L}_1 metric on the plane; a similar figure, rotated by 45 degrees, would pertain to the L_∞ metric. The \mathcal{L}_1 metric is highly degenerate in large dimensionality, because it generates many triangles that are collinear. As a result, it loses a considerable amount of information that would be preserved in the \mathcal{L}_2 metric.

Noncompactness refers to the fact that even if the underlying points lie on a low-dimensional space, the tables generated by some metrics can only be embedded in many dimensions. As an example, we illustrate with a proper metric between points on the line, such that the table of distances generated can be embedded only in a space with as many dimensions as points. In a compact space, every infinite sequence contains a convergent subsequence. Noncompact spaces are infinite in extent, and therefore sequences can run off to infinity without accumulating anywhere. Consider the sequence of vectors

$$\begin{aligned}
 v_1 &= (1, 0, 0, 0, 0, \dots) \\
 v_2 &= (1, 1, 0, 0, 0, \dots) \\
 v_3 &= (1, 1, 1, 0, 0, \dots) \\
 v_4 &= (1, 1, 1, 1, 0, \dots) \\
 &\vdots
 \end{aligned}$$

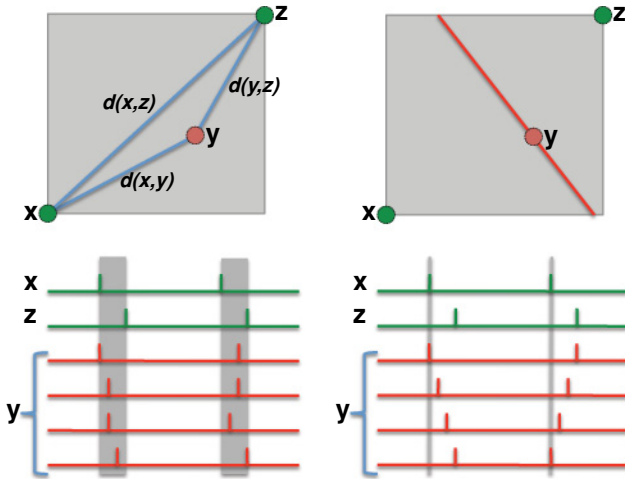


Figure 7: Geodetic degeneracy of the \mathcal{L}_1 metric on the plane (Manhattan metric). The bottom row shows the points on the plane from the top row embedded as spike trains; the x coordinate from the point is embedded as the left spike and the y coordinate as the right spike, with the two spikes being separated enough that the metric cannot confuse them). (Left) Any point z within the shaded rectangle satisfies $d(x, y) + d(y, z) = d(x, z)$ and therefore defines a degenerate, collinear triangle. Such triangles must be embedded as collinear in any Euclidean space, with y lying on the line joining x and z . The bottom left shows this diagram realized as spikes; any spike train y with two spikes, one in each shaded area, will embed as a collinear triangle according to an \mathcal{L}_1 spike metric. (Right) For any y on the red line (any line of slope $= -1$ intersecting the gray rectangle), both $d(x, y)$ and $d(y, z)$ are constant, independent of y . Thus every y on the red line defines the same table of distances and the metric is degenerate. (Bottom right) This diagram is embedded as spike trains. Any of the spike trains labeled y will give rise to exactly the same geodetic triangle, with y collinear and equidistant between x and z ; therefore, an \mathcal{L}_1 spike metric obliterates any difference between the lower four spike trains.

This sequence is the textbook example of noncompactness of the infinite-dimensional hypercube. Even if we bound the spatial extent of the individual coordinates, the fact that there are infinitely many coordinates permits this sequence to “escape.” Notice that the \mathcal{L}_2 distances between points in this sequence satisfy

$$d(v_i, v_j) = \sqrt{|i - j|} \quad \forall i, j. \tag{7.1}$$

Any metric that generates distances so distributed will, on embedding using MDS or a similar Euclidean method, become noncompact. For

instance, define a distance between points on the line as

$$d(x, y) = \sqrt{|x - y|}.$$

Then, although the points are on a line, MDS would need as many dimensions as points to embed them, and they would be embedded in a curled-up hypercube rather than a straight line. The embedding process, rather than illuminating the underlying simple structure of the space, obscures it.

This problem affects all van Rossum metrics when the underlying kernel has discontinuities, such as the half-exponential or square kernels. Consider an \mathcal{L}_2 kernel metric with a “square” kernel of unit width. The distance between two spike trains A and B containing a single spike each, at times t_A and t_B , respectively, is given by

$$d(A, B) = \left\{ \int (k(t - t_A) - k(t - t_B))^2 dt \right\}^{\frac{1}{2}},$$

where $k(t) = 1$ if $t \in (0, 1)$ and zero otherwise. Without loss of generality, set $t_A = 0$ and $t_B = \Delta$. Then this expression becomes $d(A, B) = \sqrt{2\Delta}$ when $\Delta < 1$.

Taking the square root of an \mathcal{L}_1 metric has been proposed (Aronov & Victor, 2004) as a means to alleviate certain problems with the metric. This will immediately cause noncompactness, as the resulting metrics will contain subspaces behaving like equation (7.1).

Hyperbolicity refers to the fact that Euclidean space cannot accommodate many simple metrical structures that come from negatively curved (hyperbolic) spaces. When a metric, either correctly or accidentally, generates a distance table containing such structures, MDS will develop imaginary eigenvalues and will embed using complex-valued coordinates (Aronov & Victor, 2004). The simplest such structure is a tetrahedron having three sides that are metrically valid yet too small for Euclidean space, as shown in Figure 8.

To summarize this section, we give a specific example as to how these considerations generate spurious complexity, as discussed above. Consider a set of N spike trains S_i with two spikes each, at times

$$S_i = (\eta_i, 20 + \xi_i),$$

with $i \in 1, \dots, N$, and where the η and ξ are independent gaussian random numbers of unit variance. It is clear that the underlying space of responses is two-dimensional, regardless of how large N may be; we shall use $N = 1000$. This could correspond to a situation in which the neuron always responds with the same response, two spikes are separated by 20 milliseconds, the spikes have a jitter of 1 millisecond, much smaller than the separation between spikes, and 1000 measurements were made. In Figure 9, we depict

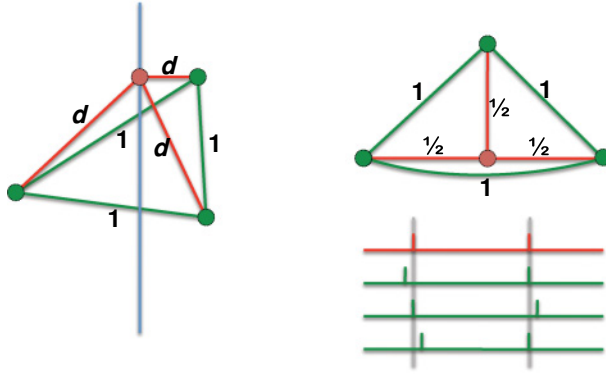


Figure 8: Hyperbolicity of \mathcal{L}_1 . (Left) Consider a tetrahedron formed by three points on an equilateral triangle of unit side, plus an additional point on the line perpendicular to the center of the triangle. The smallest distance d between the latter and former points in Euclidean space is $\sqrt{3}/2$ and is achieved when the fourth point is coplanar with the triangle. However, any distance $d \geq 1/2$ satisfies all triangle inequalities and defines a geodetically acceptable tetrahedron. Distances $1/2 \leq d < \sqrt{3}/2$ can be embedded in negatively curved, hyperbolic manifolds such as the Poincaré disk but not in Euclidean space, and hence cause algorithms like MDS to find complex coordinates for them. We refer to such structures as hyperbolic tetrahedra. (Right, top) Four points on the plane under the \mathcal{L}_1 metric forming a hyperbolic tetrahedron; the center point is at $(0, 0)$, and the equilateral triangle is defined by the points $(\pm\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$. (Bottom right) Spike embedding of the diagram above. The distance between the top spike train and any of the three lower trains is twice the distance among the lower three in the \mathcal{L}_1 metric.

the result of generating one realization of these random variables and then computing three tables of distances between the spike trains: one according to the van Rossum \mathcal{L}_2 metric using a square kernel of width 5, a second one according to the original Victor-Purpura \mathcal{L}_1 metric with $Q = 5$, and a third one using our own procedure. (Using the square root of the distances in the second table, as suggested in Aronov and Victor (2004), would in this case almost surely generate the same table as the van Rossum metric, since the spikes are so separated that misattribution is unlikely.) We then subjected the tables to MDS to verify the distribution of eigenvalues, which is shown in Figure 9, where we depict the absolute value of eigenvalues normalized by the sum of absolute values (i.e., fraction of variance explained). All three procedures have two leading eigenvalues, but while our own has no further significant eigenvalues, the other two procedures have a significant amount of the variance into higher-order modes, spuriously suggesting a higher dimensionality.

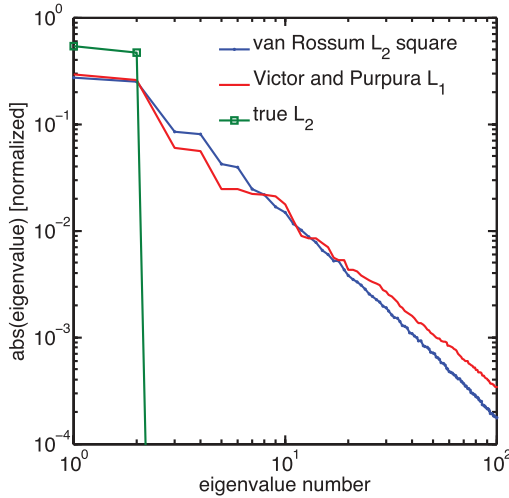


Figure 9: Distribution of MDS eigenvalues for the embedding procedure described in the text, for the van Rossum \mathcal{L}_2 metric with a square kernel, the Victor-Purpura \mathcal{L}_1 metric, and the \mathcal{L}_2 metric described in this letter. Only the latter recovers precisely two nonzero eigenvalues. Note the power-law behavior in the eigenvalue distribution. A substantial fraction (over 30%) of eigenvalues in the Victor-Purpura metric was negative (i.e., hyperbolic).

8 Discussion

We have presented a spike metric satisfying two important desiderata: that it be grounded in the time-coding hypothesis of spike generation and that it be closely related to the Euclidean \mathcal{L}_2 norm. We demonstrated that the latter property is important to subsequent stages of analysis of spike data, especially if MDS is used. Houghton (2009) has recently shown that this metric slightly outperforms the standard \mathcal{L}_1 metric for clustering spiking responses to zebra finch songs, recorded from field L of an anesthetized zebra finch, when using metric clustering.

We have also presented a fast algorithm to calculate our metric and proved its correctness and amortized complexity. Our proof used the toolboxes of graph theory and combinatorial optimization and demonstrates that they can be usefully brought to bear on problems in computational biology.

We conjecture that related procedures to the one we have presented could be used to align both spike trains with inputs from multiple neurons (as done in Aronov, Reich, Mechler, & Victor, 2003) and strands of DNA.

Future work will include applications to multiple alignment of neural signals and DNA. Our metric suggests new measures of the quality of multiple alignments, and our algorithm suggests new ways of computing

them. One approach to aligning multiple spike trains would be to generalize the CLUSTAL procedures for progressive alignment, that is, building a multiple alignment from a collection of aligned pairs (Higgins & Sharp, 1988; Thompson, Higgins, & Gibson, 1994).

Appendix

Definitions. G is connected if one can travel from any vertex to any other vertex via the edges. A component of G is a maximal connected subgraph. A path in which the first and last edges share a vertex is called a circuit.

Lemma 4 (Schrijver, 2003, theorem 1). *If M is a matching in G , then either there exists an M -augmenting path P or no matching of greater cardinality exists. If N is a matching of greater cardinality, we can pick $P \subseteq M \cup N$.*

Proof. If M is a matching of maximum cardinality and P is an M -augmenting path, then $M \Delta P$ would be a matching of greater cardinality, a contradiction. If M is not a matching of maximum cardinality, there exists a larger matching N . Let H be the graph with vertices V and edges $M \cup N$. Every connected component of H is a path or a circuit. As $|N| > |M|$, one of these components C contains more edges of N than M . Its edges must alternate between those in M and those in N , since no two edges of either one can be connected to the same vertex. Thus, in order to have more edges in N than in M , it must start and end with nonidentical edges in N . Therefore, C is an M -augmenting path.

Proof of theorem 1 (from Schrijver, 2003, section 3.5, proposition 1). Let N be an arbitrary matching of size $k + 1$. By lemma 4, we can pick an M -augmenting path Q in $M \cup N$. By definition, $l(Q) \leq l(P)$. $|N \Delta Q| = k$, and since M is of minimum weight for cardinality k , $w(N \Delta Q) \geq w(M)$. Therefore,

$$w(N) = w(N \Delta Q) - l(Q) \geq w(M) - l(P) = w(M').$$

Therefore M' is of minimum weight over matchings of cardinality $k + 1$.

Proof of lemma 2. Let $\alpha = x_{i_1} - y_{j_2}$, $\beta = x_{i_2} - y_{j_1}$, and $\gamma = y_{j_2} - y_{j_1}$. Now, $0 < \gamma < \beta - \alpha$. We want

$$f(\gamma) = |\alpha|^p + |\beta|^p - |\alpha + \gamma|^p - |\beta - \gamma|^p > 0.$$

The left-hand side has roots for γ at 0 and $\beta - \alpha$:

$$f'(\gamma) = -p \cdot \text{sign}(\alpha + \gamma)|\alpha + \gamma|^{p-1} + p \cdot \text{sign}(\beta - \gamma)|\beta - \gamma|^{p-1}.$$

This expression is zero if and only if $\gamma = (\beta - \alpha)/2$. By Jensen's inequality,

$$f\left(\frac{\beta - \alpha}{2}\right) = |\alpha|^p + |\beta|^p - 2\left|\frac{\alpha + \beta}{2}\right|^p > 0.$$

The lemma follows by Rolle's theorem.

Proof of theorem 2. We assume without loss of generality that P is of the form

$$(\{a_{i_1}, b_{j_1}\}, \{a_{i_1}, b_{j_2}\}, \dots, \{a_{i_l}, b_{j_l}\}).$$

Assume for the sake of contradiction that for some t , either $i_t < i_{t+1}$ and $i_{t+1} > i_{t+2}$ or $i_t > i_{t+1}$ and $i_{t+1} < i_{t+2}$ (The cases where for some t , either $j_t < j_{t+1}$ and $j_{t+1} > j_{t+2}$ or $j_t > j_{t+1}$ and $j_{t+1} < j_{t+2}$ are analogous.) The cases are similar, so we treat only the first one. It has two subcases: $j_{t+1} > j_{t+2}$ and $j_{t+1} < j_{t+2}$. In the first subcase, the edges $\{a_{i_t}, b_{j_{t+1}}\}$ and $\{a_{i_{t+1}}, b_{j_{t+2}}\}$ are both in M , contradicting lemma 3. In the second subcase case, the edges $\{a_{i_{t+1}}, b_{j_{t+1}}\}$ and $\{a_{i_{t+2}}, b_{j_{t+2}}\}$ are both in $M\Delta P$, which by theorem 1, is a minimum weight matching of cardinality $k + 1$, contradicting lemma 3.

Proof of main theorem. We show that if M is a minimum weight matching of cardinality k , then an M -augmenting path of maximum length, P , is of the form

$$(\{a_i, b_j\}, \{a_i, b_{j+1}\}, \dots, \{a_{i+N}, b_{j+N}\}) \text{ or} \\ (\{a_i, b_j\}, \{a_{i+1}, b_j\}, \dots, \{a_{i+N}, b_{j+N}\}),$$

where in the first case, there are no unmatched elements of x between x_i and y_j and there are no unmatched elements of y between x_{i+N} and y_{i+N} , and in the second case, there are no unmatched elements of y between x_i and y_j and no unmatched elements of x between x_{i+N} and y_{i+N} . Therefore, transforming M into $M\Delta P$ is equivalent to the shift operation described in section 4.

We first prove the case where $p > 1$. Using theorem 2, assume without loss of generality that P is of the form

$$(\{a_{i_1}, b_{j_1}\}, \{a_{i_1}, b_{j_2}\}, \dots, \{a_{i_l}, b_{j_l}\}),$$

where $\{a_{i_t}\}$ and $\{b_{i_t}\}$ are strictly increasing. The theorem is clear when $l = 1$. So let $l > 1$, and assume that there exists an integer r such that $i_t < r < i_{t+1}$ (we can assume that there exists an integer between j_t and j_{t+1} ; this case is similar). Now there are two cases: either a_r is connected to an edge in M , or it is not. If it is connected to an edge, call that edge $\{a_r, b_s\}$. $s > j_{t+1}$; otherwise

the fact that $\{a_r, b_s\}$ and $\{a_{i_t}, b_{j_{t+1}}\}$ are both in M contradicts lemma 3. But in that case, since $\{a_r, b_s\} \notin P$, $\{a_r, b_s\}$ and $\{a_{i_{t+1}}, b_{j_{t+1}}\}$ are both in $M\Delta P$, which by theorem 1 is a minimum weight matching of cardinality $k + 1$, contradicting lemma 3.

Now consider the case when a_r is not connected to an edge. Replace the edge $\{a_{i_{t+1}}, b_{j_{t+1}}\}$ with $\{a_r, b_{j_{t+1}}\}$ in P to form P' . P' may not be an M -augmenting path, but $M\Delta P'$ is still a matching of cardinality $k + 1$. We know that $q^p |x_{i_t} - y_{j_{t+1}}|^p \leq q^p |x_r - y_{j_{t+1}}|^p$, since otherwise, the edge $\{a_r, b_{j_{t+1}}\}$ would be in M instead of $\{a_{i_t}, b_{j_{t+1}}\}$. Therefore, $|x_{i_t} - y_{j_{t+1}}| \leq |x_r - y_{j_{t+1}}|$. If $y_{j_{t+1}} \geq x_r$, this implies $-x_{i_t} + y_{j_{t+1}} \leq -x_r + y_{j_{t+1}}$ making $x_{i_t} \geq x_r$, a contradiction. So $x_r > y_{j_{t+1}}$. Since $x_{i_{t+1}} > x_r$, $x_r - y_{j_{t+1}} < x_{i_{t+1}} - y_{j_{t+1}}$, making

$$w(M\Delta P') - w(M\Delta P) = q^p |x_r - y_{j_{t+1}}|^p - q^p |x_{i_{t+1}} - y_{j_{t+1}}|^p < 0.$$

$M\Delta P'$ is a matching of cardinality $k + 1$ of lower weight than that of $M\Delta P$, but since P is of maximum length, theorem 1 implies that $M\Delta P$ has minimum weight. This is a contradiction, and the theorem follows for $p > 1$.

Finally, assume that there is an unmatched element of \mathbf{x} between x_{i_1} and y_{j_1} ; call it x_u , with corresponding vertex a_u (the case when there is an unmatched element of \mathbf{y} between x_{i_1} and y_{j_1} is analogous). Since $x_u < x_{i_1}$ by the previous paragraph, $y_{j_1} \leq x_u < x_{i_1}$. Replace the edge $\{a_{i_1}, b_{j_1}\}$ with $\{a_u, b_{j_1}\}$ in P to form P' . P' may be an M -augmenting path, but as before, $M\Delta P'$ is still a matching of cardinality $k + 1$. $x_u < x_{i_1}$ implies $x_u - y_{j_1} < x_{i_1} - y_{j_1}$, and since these quantities are nonnegative,

$$w(M\Delta P') - w(M\Delta P) = q^p |x_u - y_{j_1}|^p - q^p |x_{i_1} - y_{j_1}|^p < 0.$$

Using theorem 1, P could not have been an M -augmenting path of maximum length, a contradiction. The theorem follows for $p > 1$.

Now we address the $p = 1$ case. For some constants $C \in \mathbb{Z}, \alpha_1, \dots, \alpha_J > 0$, and $u_1, \dots, u_J \in \{\pm 1\}$, the difference in length between two M -augmenting paths P and P' is

$$g(p) = l(P') - l(P) = q^p \sum_{i=1}^J u_i \alpha_i^p + 2C.$$

$g(p)$ is real analytic, and its Taylor series $T(p)$ converges everywhere. Either $g(p)$ is identically 0, or there exists some least integral exponent e such that the coefficient of p^e in $T(p)$ is nonzero. If it is positive, then there exists some $\epsilon_1 > 0$ such that $f(p)$ is strictly positive on $(1, 1 + \epsilon_1)$; likewise, if it is negative, we can pick ϵ_1 such that $g(p)$ is strictly negative on $(1, 1 + \epsilon_1)$ by Nash (1959), theorem 1. We can determine such an ϵ_1 for every pair of M -augmenting paths P and P' that do not have the same length for all p . Let ϵ_2 be the minimum of all of these ϵ_1 's. There exists a set of M -augmenting

paths Q_1, \dots, Q_K such that for $p \in (1, \epsilon_2)$, they all have equal length, and they are all longer than all other M -augmenting paths. They all must be shifts. By continuity, when $p = 1$, they still have equal length and are at least as long as all other M -augmenting paths. Therefore, $M \Delta Q_1$ is a matching of minimum weight of cardinality $k + 1$ when $p = 1$.

Definition. Let M be a matching of minimum weight of cardinality k and P be an M -augmenting path. We call P an M -shift if it corresponds to a shift in the matching matrix of M .

Proof of theorem 3. Let M_k be the minimum weight matching of cardinality k . It suffices to show that if $w(M_k) < w(M_{k+1})$, then if M_{k+2} exists, $w(M_{k+1}) < w(M_{k+2})$. It follows by induction that $w(M_k)$ has lower weight than all minimum weight matchings of cardinality greater than k . Therefore, the sequence $w(M_1), w(M_2), \dots, w(M_{\min(m,n)})$ is decreasing, then increasing, and the point at which the switch occurs is its minimum.

All M_k -shifts have negative length, since the one with the greatest length, P , has negative length, as $w(M_k) < w(M_{k+1})$ (using theorem 1). Therefore, all M_{k+1} -shifts have negative length since they are also M_k shifts, except for possibly one, Q , that is not. If Q does not exist or is not the M_{k+1} -shift of greatest length, then the shift of greatest length has negative length, and $w(M_{k+1}) < w(M_{k+2})$, so we are done. So assume that Q is the M_{k+1} -shift of greatest length. We shall prove that $l(Q) \leq 0$. Sending M_k to $M_k \Delta P = M_{k+1}$ and then sending M_{k+1} to $M_{k+1} \Delta Q = M_{k+2}$ corresponds to deleting two pairs of stars in one of four classes of matching matrices. We consider only two of them, since the others are analogous:

$$\begin{bmatrix} \cdots & x_i & x_{i+1} & \cdots & * & \cdots & x_{i+s} & \cdots & x_{i+t} & * & \cdots \\ \cdots & * & y_j & \cdots & y_{j+r} & \cdots & * & \cdots & y_{j+t-1} & y_{j+t} & \cdots \end{bmatrix}$$

$$\begin{bmatrix} \cdots & x_i & x_{i+1} & \cdots & x_{i+r} & \cdots & * & \cdots & x_{i+t} & * & \cdots \\ \cdots & * & y_j & \cdots & * & \cdots & y_{j+s} & \cdots & y_{j+t-1} & y_{j+t} & \cdots \end{bmatrix}.$$

To get the other two cases, switch the top stars to the bottom and the bottom stars to the top in each of the above examples, and reindex accordingly. In both matrices, deleting the inner pair of stars corresponds to transforming M_k into M_{k+1} and deleting the outer pair of stars corresponds to transforming M_{k+1} into M_{k+2} . In the first matching matrix, there are three possible M_k -shifts. The middle one is P . Call the left-most one P_- and the right-most one P_+ . $l(P_+) \leq l(P) < 0$, and $l(P_-) \leq l(P)$. $l(P) + l(Q) = l(P_-) + l(P_+)$, since performing the shift P followed by Q is

the same operation as performing the shift P_- followed by P_+ on the matrix. Therefore, $l(Q) = l(P_+) - l(P) + l(P_-) < 0$.

Now consider the second matching matrix. There are M_k -augmenting paths P_- and P_+ such that changing M_k to $M_k \Delta P_-$ or $M_k \Delta P_+$ is equivalent to deleting both the first and third stars in the matrix (from the left) or both the second and fourth stars from the matrix. As before, $l(P_+) \leq l(P) < 0$, $l(P_-) \leq l(P)$, and $l(P) + l(Q) = l(P_-) + l(P_+)$. Again, $l(Q) = l(P_+) - l(P) + l(P_-) < 0$.

A.1 Discontinuous Kernels in the Convolution Metrics. The convolution metrics introduced in van Rossum (2001) convert spike trains S into real-valued functions f_S and then compute the distance between the functions using the standard \mathcal{L}^p metric in function space. The spike trains are transformed into functions by considering each spike to define a Dirac delta function at time t_i , and then convolving the deltas with a given kernel function $k(t)$:

$$S = \{t_i, i : 1 \rightarrow N\} \rightarrow g(t) \equiv \sum_{i=1}^N \delta(t - t_i) \rightarrow f_S \equiv \int g(t)k(t - s) ds,$$

so exchanging the integral and summation sign,

$$f_S(t) = \sum_{i=1}^N k(t - t_i).$$

These functions are then meant to be compared using the \mathcal{L}^p metric in function space:

$$d_p(f, g) = \left[\int (f(t) - g(t))^p dt \right]^{1/p}.$$

Given that spike timings are affected by jitter and other sources of noise, it is reasonable to expect that the final spike metric will be a smooth function of the times t_i in the original spike trains. While the function is evidently continuous for appropriately integrable kernels k , if the kernel is discontinuous and $p > 1$, there is a smoothness problem. Common choices of discontinuous kernels k include a square function (box count) and the half-exponential function $k(t > 0) = \exp(-t)$, $k(t < 0) = 0$.

To see this, consider two spike trains S and U with one spike each, at $t = 0$ and $t = \Delta$, respectively, and $p = 2$. We shall assume Δ is suitably small. Then

$$d(S, U) = \left[\int (k(t) - k(t - \Delta))^2 dt \right]^{1/2}.$$

If k were smooth, then $k(t) - k(t - \Delta)$ is $\mathcal{O}(\Delta)$ with a support of $\mathcal{O}(1)$; the integral is then $\mathcal{O}(\Delta^2)$, and the distance, its square root, becomes again $\mathcal{O}(\Delta)$. But if k is discontinuous, as in the boxcar kernel $k(0 < t < 1) = 1, k(t < 0 | t > 1) = 0$, then $k(t) - k(t - \Delta)$ is $\mathcal{O}(1)$ with a support $\mathcal{O}(\Delta)$. The integral is $\mathcal{O}(\Delta)$ and the distance $\mathcal{O}(\sqrt{\Delta})$. The distance then shows a singularity at small differences Δ , which causes the noncompactness issues addressed in section 7.

Acknowledgments

A.D. and B.S. thank B. Kaiser and the RU Outreach program for support and encouragement. This work was supported in part by the National Science Foundation under grant EF-0928723.

References

- Aronov, D., Reich, D., 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.
- Bengio, Y., Paiement, J., & Vincent, P. (2004). Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps and spectral clustering. In S. Thrun, L. K. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems*, *16*. Cambridge, MA: MIT Press.
- Burkard, R. (2002). Selected topics on assignment problems. *Discrete Applied Mathematics*, *123*, 257–302.
- Burkard, R., Klinz, B., & Rudolf, R. (1996). Perspectives of Monge properties in optimization. *Discrete Applied Mathematics*, *70*, 95–161.
- Chase, S., & Young, E. (2006). Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus. *Journal of Neuroscience*, *26*, 3889–3898.
- Ding, C., & He, X. (2004). K-means clustering via principal component analysis. In *Proceedings of the International Conference on Machine Learning* (pp. 225–232). San Francisco: Morgan Kaufmann.
- Gerstner, W., Kreiter, A., Markram, H., & Herz, A. (1997). Neural codes: Firing rates and beyond. *Proceedings of the National Academy of Sciences*, *94*, 12740–12741.
- Higgins, D., & Sharp, P. (1988). CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene*, *73*, 233–244.
- Hoffman, A. (1963). On simple linear programming problems. In *Proceedings of Symposia in Pure Mathematics* (Vol. 7, pp. 317–327).

- Houghton, C. (2009). A comment on "A fast L_p spike alignment metric." arXiv:0907.3137.
- Kruskal, J. J., & Wish, M. (1978). *Multidimensional scaling*. Thousand Oaks, CA: Sage.
- Kuhn, H. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2, 83–97.
- Monge, G. (1781). Déblai et remblai. *Mémoires de l'Académie des Sciences*.
- Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society of Industrial and Applied Mathematics*, 5, 32–38.
- Nash, S. (1959). The higher derivative test for extrema and points of inflection. *American Mathematical Monthly*, 66, 709–713.
- Notredame, C. (2002). Recent progress in multiple sequence alignment: A survey. *Pharmacogenomics*, 3, 131–144.
- Papadimitriou, C., & Steiglitz, K. (1998). *Combinatorial optimization: Algorithms and complexity*. Mineola, NY: Dover.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
- Roweis, S., & Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323–2326.
- Samonds, J., & Bonds, A. (2004). From another angle: Differences in cortical coding between fine and coarse discrimination of orientation. *Journal of Neurophysiology*, 91, 1193–1202.
- Schrauwen, B., & Campenhout, J. (2007). Linking non-binned spike train kernels to several existing spike train metrics. *Neurocomputing*, 70, 1247–1253.
- Schrijver, A. (2003). *A course in combinatorial optimization*. Available online at <http://www.math.ku.edu/~jmartin/old-courses/math996-F06/Schrijver.pdf>.
- Sellers, P. (1974). On the theory and computation of evolutionary distances. *SIAM Journal on Applied Mathematics*, 26, 787–793.
- Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2319–2323.
- Thompson, J., Higgins, D., & Gibson, T. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673–4680.
- van Rossum, M. (2001). A novel spike distance. *Neural Computation*, 13, 751–763.
- Victor, J. (2005). Spike train metrics. *Current Opinion in Neurobiology*, 15, 585–592.
- Victor, J., Goldberg, D., & 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., & Purpura, K. (1996). Nature and precision of temporal coding in visual cortex: A metric-space analysis. *Journal of Neurophysiology*, 76, 1310–1326.
- Victor, J., & Purpura, K. (1997). Metric-space analysis of spike trains: Theory, algorithms, and applications. *Network*, 8, 127–164.

This article has been cited by: