Proceedings of the 2nd Joint Symposium on Neural Computation, University of California, San Diego and California Institute of Technology, Vol. 5, 191-202, Institute for Neural Computation, La Jolla, CA. (1995)

A unifying measure for neighbourhood preservation in topographic mappings

Geoffrey J. Goodhill, Steven Finch*, & Terrence J. Sejnowski

The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037.

* Human Communication Research Center, University of Edinburgh 2 Buccleuch Place Edinburgh EH8 9LW GREAT BRITAIN.

Abstract

In this paper, the abstract computational principles underlying topographic maps are discussed. We give a definition of a "perfectly neighbourhood preserving" map, which we call a topographic homeomorphism, and we prove that this has certain desirable properties. It is argued that when a topographic homeomorphism does not exist (the usual case), many equally valid choices are available for quantifying the quality of a map. We introduce a particular measure that encompasses several previous proposals, and discuss its relation to other work. This formulation of the problem sets it within the well-known class of quadratic assignment problems.

1 Introduction

Problems of mapping occur frequently both in understanding biological processes and in formulating abstract methods of data analysis. It is important to be as clear as possible regarding what computational principles are being addressed by such mappings. The particular types of mappings of concern in this paper are those sometimes referred to as topographic, topological, topology-preserving, or neighbourhood preserving. The intuitive notion here is that nearby points map to nearby points.[1] In neurobiology, many different mappings have this general form, for instance between the retina and more central structures [Udin & Fawcett 1988]. In data analysis a common goal is to map data from a high dimensional space to a low dimensional space so as to preserve as far as possible the "internal structure" of the data in the high dimensional space (see e.g. [Krzanowski 1988]), for example using PCA, multidimensional scaling [Shepard 1980], or neural network algorithms such as the self-organizing feature map [Kohonen 1982, Kohonen 1988] or the elastic net [Durbin & Willshaw 1987]. Again the intuition is to preserve the structure of the data by mapping nearby points to nearby points. However, it is vital to ask what this intuitive idea might mean more precisely. Without a clear set of principles it is impossible to address whether a particular mapping has achieved "neighbourhood preservation", whether one mapping algorithm has performed better than another on a particular problem, or what computational goals might be being pursued by mappings in the brain.

An important point to understand in addressing these issues is that there is no one correct answer. As discussed below, a large number of choices have to be made to reach a precise mathematical measure of neighbourhood preservation. Different combinations of choices will in general give different answers for the same mapping problem, and the combination of choices that is most appropriate will vary from

[1] In some contexts it is also desirable to specify that far-away points should map to far-away points.

problem to problem. Several measures of neighbourhood preservation have recently been proposed which implement particular sets of choices (e.g. [Durbin & Mitchison 1990, Bauer & Pawelzik 1992, Villmann et al 1994]). From a biological perspective, an interesting question is to investigate which combinations of choices yield mappings closest to those seen experimentally in various contexts, and how such choices could be implemented in the brain. From a practical perspective, it is desirable to understand more about what are the most appropriate methods for different problems.

We adopt the distinction made in [Marr 1982] between the "computational" and "algorithmic" levels of analysis. The former concerns the computational goals being addressed, while the latter concerns how these computational goals might be achieved. These two levels can sometimes be hard, or inappropriate, to disentangle when addressing biological problems [Sejnowski et al 1988]. However, for discussing topographic mappings from an abstract perspective it is important to be clear about this distinction. As an example, minimal wiring and minimal path length (discussed later) are clear computational level principles, while Kohonen's algorithm [Kohonen 1982] exists only at the algorithmic level since it is not following the gradient of an objective function [Erwin et al 1992].

The first principal aim of this paper is to clearly explain some of the choices possible at the computational level, in order to sharpen understanding of these issues and provide a framework in which existing and future measures can be interpreted. The second is to explore in more detail one set of choices, that still retain a fair degree of generality. The paper is in three parts. We first propose a rather general qualitative concept of neighbourhood-preservation, and prove a theorem showing that it has a useful property. We then present one particular quantitative measure, and prove that it is indeed an implementation of the general concept. In the third part we show how several recent measures can be fitted into this overall scheme, and discuss some other approaches that do not fit into this scheme.

The following assumptions are made throughout:

- All mappings are bijective (i.e. one-to-one).
- In each space there exists a fixed "similarity structure", which specifies for every pair of points their degree of similarity. A simple case is that this similarity is just a function of the euclidean distance between points in a geometric space. However, the similarity structure need not have a geometric interpretation. For purely "nearest neighbour" structure for example, for each space the set of possible similarities contains only two elements, where each pair of points is either similar or not similar.

2 Preserving similarities

We desire to capture more precisely the intuitive notion that a mapping preserves the "internal structure" of the data in one space in another space. There are at least two obvious choices for defining a mapping that is "perfect" in this sense. The first is to say that the mapping must *preserve similarities*; that is, for each pair of points in one space, the similarity between them should be equal to (or more weakly, linearly related to) the similarity between their images in the other space. The second is to say that the mapping must only preserve similarity *orderings*. That is, rather than comparing the absolute values of the similarities between pairs of points in one space and the distance between their images in the other, one is concerned only that their relative orderings within the two sets of similarities are the same. Although the latter appears rather weaker, preserving only similarity ordering does in fact impose very strong constraints on the mapping (as been extensively discussed in the context of non-metric multidimensional scaling: see e.g. [Shepard 1980]). It is this choice with which we will primarily be concerned. We first explore mathematically what sort of mapping is produced by preserving distance orderings between two continuous metric spaces, when such a mapping exists. In particular we prove a theorem that says that a mapping that preserves distances on such a space is a *homeomorphism*, i.e. a continuous mapping. It is useful to first investigate continuous spaces since most topological concepts, including homeomorphism, have few implications for discrete spaces, but strong implications for continuous ones. Later we discuss discrete spaces.

Proposition 2.1 *Let* $\langle X, f \rangle$, $\langle Y, g \rangle$ *be identical metric spaces with countable dense subsets. Let* $\mathcal{M} : X \to Y$ *be a bijection such that:*

$$\begin{array}{ll} \forall x_1, x_2, x_3, x_4 \in X & f(x_1, x_2) < f(x_3, x_4) \Rightarrow g(\mathcal{M}(x_1), \mathcal{M}(x_2)) \leq g(\mathcal{M}(x_3), \mathcal{M}(x_4)) \\ \forall y_1, y_2, y_3, y_4 \in Y & g(y_1, y_2) < g(y_3, y_4) \Rightarrow f(\mathcal{M}^{-1}(y_1), \mathcal{M}^{-1}(y_2)) \leq f(\mathcal{M}^{-1}(y_3), \mathcal{M}^{-1}(y_4)) \end{array}$$

Then $\mathcal{M}$ *is a homeomorphism, and* X *and* Y *are topologically equivalent.*

Proof:

Since $\mathcal{M}$ *is given as a bijection, it remains to prove it is continuous. Since* X *and* Y *have dense subsets, we know that the* ε*-balls,* $B(x, \varepsilon)$, $B(y, \varepsilon)$ *are non-singleton sets for any* $x \in X$ *and* $y \in Y$. *We first show that for any* $\varepsilon > 0$ *and* $x \in X$, *there exists* $w \neq x \in X$ *such that* $f(x, w) < \varepsilon$ *and* $g(\mathcal{M}(x), \mathcal{M}(w)) < \varepsilon$.

Suppose this is false. Then there exists $x \in X$ *such that the* ε*-ball,* $B(x, \varepsilon)$, *has an image under* $\mathcal{M}$ *which contains only the point* $\mathcal{M}(x)$ *in* $B(\mathcal{M}(x), \varepsilon)$. *But then consider a point* $v \neq x \in B(\mathcal{M}(x), \varepsilon)$. *Then* $f(\mathcal{M}^{-1}(v), x) > \varepsilon$ *(otherwise* $\mathcal{M}^{-1}(v)$ *would be a point in the* ε*-ball of* x *also in the* ε*-ball of* $\mathcal{M}(x)$*). Also consider a point* $w \neq x \in B(x, \varepsilon)$. *By assumption,* $f(x, w) < \varepsilon < f(x, \mathcal{M}^{-1}(v))$.

Consequently, by the assumption of the theorem, $g(\mathcal{M}(x), \mathcal{M}(w)) \leq g(\mathcal{M}(x), v)$. *But* $\mathcal{M}(w)$ *cannot be in* $B(\mathcal{M}(x), \varepsilon)$ *(for the same reason that* $\mathcal{M}^{-1}(v)$ *cannot be in* $B(x, \varepsilon)$*). Thus we have* $g(\mathcal{M}(x), \mathcal{M}(w)) > g(\mathcal{M}(x), v)$, *which yields a contradiction.*

To prove that $\mathcal{M}$ *is continuous, then for any* $\varepsilon > 0$ *and any point* $x \in X$, *we choose a point* $w \in B(x, \varepsilon)$ *such that* $g(\mathcal{M}(x), \mathcal{M}(w)) < \varepsilon$ *(as we have just demonstrated possible). We now choose* $0 < \delta < f(x, w) < \varepsilon$. *Choose a point* $u \in B(x, \delta)$. *Since* $f(x, u) < f(x, w)$, *we have by assumption that* $g(\mathcal{M}(x), \mathcal{M}(u)) < g(\mathcal{M}(x), \mathcal{M}(w)) < \varepsilon$. *Since this applies for any point* $u \in B(x, \delta)$, *this proves that* $\mathcal{M}$ *is continuous.*

In fact, the criterion that ordering of distances between points in the spaces X and Y is essentially monotonic under $\mathcal{M}$ is a far stronger constraint than is necessary for X and Y to be homeomorphic, and we call $\mathcal{M}$ a *topographic homeomorphism*.

What about discrete spaces? For these continuity is not defined. However, by analogy with the continuous case we say that a mapping between discrete spaces that has the property of preserving similarity orderings is a discrete approximation to a topographic homeomorphism. We will take this as our definition of a "perfect" mapping.

3 Measuring discrepancies

Our discussion has so far dealt only with the case where a topographic homeomorphism exists. The more practically relevant case is how to do the best one can when such a mapping does not exist. Some *measure of discrepancy* from perfect matching of similarity orderings is required, and here a multitude of choices are possible. If similarities are to be matched exactly one might choose a monotonic function of the difference in similarities and sum over all pairs of similarities. A particular example of this is metric multidimensional scaling [Torgerson 1952], discussed further below. If only the orderings are to be matched, one has the choice of whether or not to take into account absolute similarity values when calculating violations of similarity ordering. One of the few established methods for quantifying ordering violations without reference to the actual similarities are the "stress" functions of non-metric multidimensional scaling [Kruskal 1964(a), Kruskal 1964(b), Takane et al 1977]. However, we note that these measures are somewhat ad hoc.

Virtually all the other measures of discrepancy that have been proposed use absolute similarity values. The choices they make will be discussed further later. First we introduce one particular measure and demonstrate that it has some useful properties.

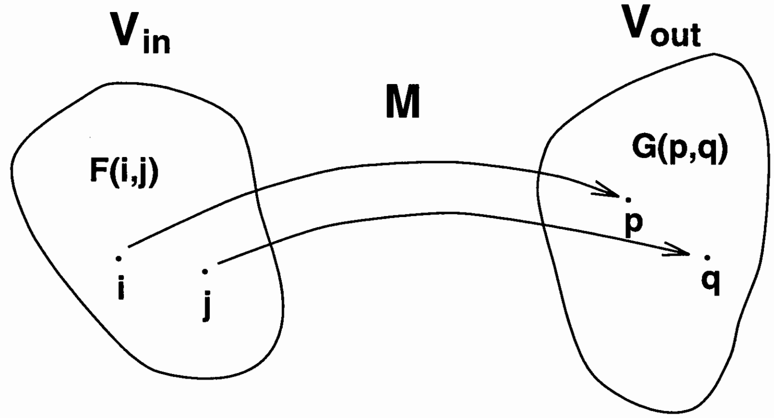

Figure 1: The mapping framework.

3.1 The C measure

Consider an input space V_{in} and an output space V_{out}, each of which contains N points (see figure 1). Let M be the mapping from points in V_{in} to points in V_{out}. We use the word "space" in a general sense: either or both of V_{in} and V_{out} may not have a geometric interpretation. Assume that for each space there is a symmetric "similarity" function which, for any given pair of points in the space, specifies how similar they are. Call these functions F for V_{in} and G for V_{out}. Then we define a cost functional C, given by

$$C = \sum_{i=1}^{N} \sum_{j<i} F(i,j)G(M(i), M(j)) \tag{1}$$

where i and j label points in V_{in}, and M(i) and M(j) are their respective images in V_{out}. The sum is over all possible pairs of points in V_{in}. Since M is a bijection it is invertible, and C can equivalently be written

$$C = \sum_{i=1}^{N} \sum_{j<i} F(M^{-1}(i), M^{-1}(j))G(i,j) \tag{2}$$

where now i and j label points in V_{out}, and M^{-1} is the inverse map.

A good mapping is one with a high value of C. However, if one of F or G were given as a *dissimilarity* function (i.e. increasing with decreasing similarity) then a good mapping would be one with a low value of C. How F and G are defined is problem-specific. They could be euclidean distances in a geometric space, or some (possibly non-monotonic) function of those distances. Or F and G could just be given, in which case it may or may not be possible to interpret the points as lying in some geometric space. There are two main reasons why C is a useful measure to consider in detail. Firstly, we prove that if a topographic homeomorphism exists between V_{in} and V_{out}, then maximising C will find it. Secondly, we show that several other proposals for discrepancy measures are in fact equivalent to C for particular choices of F and G.

Note that although C finds a mapping that preserves similarity orderings if one exists, it does not do this by matching the absolute values of the similarities. However when a topographic homeomorphism does not exist, C does take absolute similarity values into account for calculating the discrepancy. C is one particular choice of how to trade off different ordering violations. Whether it is more appropriate to for instance attempt to keep all ordering violations reasonably small, or to keep most very small at the expense of a few large violations (cf [Durbin & Mitchison 1990]), is problem specific.

3.2 Theorems

We now prove that if there is a mapping between two spaces which preserves the relative ordering of similarities, then maximizing C will find it. This work is presented within the framework of topology in the hope that we can apply some of the concepts and results of this field to the study of "topographic" mappings. We start with a lemma which shows that for finite vectors of real numbers, the inner product is maximized over all permutations within the two vectors if the elements of the vector are identically ordered, and that maximizing C leads to orderings which do not violate monotonicity.

Lemma 3.1 *Let $X = \{x_i\}$ and $Y = \{y_i\}$ be non-strictly monotonically decreasing countable sequences of non-negative real numbers, finitely many of which are non-zero. Let $\mathcal{P} : N \to Y$ be a permutation of $\{y_i\}$. Then (1) the value*

$$C(\mathcal{P}) = \sum_{i=1}^{\infty} x_i \mathcal{P}(i)$$

is globally maximized (over $\mathcal{P}$) if $\mathcal{P}(i) = y_i$ (i.e. $\langle \mathcal{P}(i) \rangle$ is non-strictly monotonically decreasing), and (2) Maximizing $C(\mathcal{P})$ finds a permutation of Y, $\mathcal{P}^*$ such that $x_i < x_j \Rightarrow \mathcal{P}^*(i) \leq \mathcal{P}^*(j)$ and $\mathcal{P}^*(i) < \mathcal{P}^*(j) \Rightarrow x_i \leq x_j$.

Proof:

(1) Proof follows from induction on the number of non zero elements in X, n.

Case $n = 2$:

We have two possibilities for $\mathcal{P}$, so consider difference between the two possible sums, d:

$$d = (x_1 y_1 + x_2 y_2) - (x_1 y_2 + x_2 y_1) = (y_1 - y_2)(x_1 - x_2)$$

But $d \geq 0$ since $y_1 \geq y_2$ and $x_1 \geq x_2$ by the assumption of monotonicity, so $x_1 y_1 + x_2 y_2 \geq x_1 y_2 + x_2 y_1$. Thus if $\mathcal{P}(i) = y_i$ the sum is maximized.

General case:

Assume that for $n = k$ the proposition is true. Consider the sequence $X = \{x_1, x_2, \ldots, x_{k+1}, 0, 0, \ldots\}$ where all the $\{x_i\}$ are non-zero for $i \leq k + 1$. Consider

$$C(\mathcal{P}) = \sum_{i=1}^{k+1} x_i \mathcal{P}(i) = x_1 \mathcal{P}(1) + \sum_{i=2}^{k+1} x_i \mathcal{P}(i)$$

By inductive assumption, this can only be maximized if $\mathcal{P}(i) \geq \mathcal{P}(i + 1)$ for $k \geq i \geq 2$, since the second term can only be maximized in this case. Similarly,

$$C(\mathcal{P}) = x_{k+1} \mathcal{P}(k + 1) + \sum_{i=1}^{k} x_i \mathcal{P}(i)$$

So this can only be maximized if $\mathcal{P}(i) \geq \mathcal{P}(i + 1)$ for $k - 1 \geq i \geq 1$. Consequently, $\mathcal{P}(i)$ is monotonically decreasing with i, and consequently $\mathcal{P}(i) = y_i$, since it is easy to show there is only one monotonically decreasing sequence of $\langle y_1, \ldots, y_{k+1} \rangle$. By induction if $\langle y_1, \ldots, y_N, 0, 0, \ldots \rangle$, where N is the number of non-zero elements in Y, is monotonically ordered, $C(\mathcal{P})$ is maximized.

(2) Let us suppose that $\mathcal{P}(i) < \mathcal{P}(j)$ but $x_i > x_j$. Then we define a new ordering, $\mathcal{P}'$ by $\mathcal{P}'(n) = \mathcal{P}(n), i \neq j$, but $\mathcal{P}'(i) = \mathcal{P}(j)$ and $\mathcal{P}'(j) = \mathcal{P}(i)$. Clearly $\mathcal{P}'$ is a valid permutation. Consider

$$d = C(\mathcal{P}) - C(\mathcal{P}')$$

$$= \sum_{i=1}^{\infty} x_i \mathcal{P}(i) - \sum_{i=1}^{\infty} x_i \mathcal{P}'(i)$$

$$= x_i\mathcal{P}(i) + x_j\mathcal{P}_j - x_i\mathcal{P}(j) - x_j\mathcal{P}(i)$$
$$= (x_i - x_j)(\mathcal{P}(i) - \mathcal{P}(j))$$

But $x_i - x_j > 0$, *and* $\mathcal{P}(i) - \mathcal{P}(j) < 0$, *so* $d < 0$, *so* $C(\mathcal{P}') > C(\mathcal{P})$. *Thus by modus tollens* $x_i > x_j \Rightarrow \mathcal{P}(i) \geq \mathcal{P}(j)$. *Similarly, it is easy to show that* $\mathcal{P}(i) > \mathcal{P}(j) \Rightarrow x_i \geq x_j$.

What does a mapping maximizing C give us between two continuous metric spaces X and Y in the case when it is possible to find a mapping which preserves distance orderings within X and Y? We show that the answer is, one of the mappings which does indeed preserve distance orderings between the points.

Corollary 3.1.1 *Let* $\langle X, f\rangle$ *and* $\langle Y, g\rangle$ *be finite metric spaces such that* $|X| = |Y|$. *Consider the set* H *of bijections* $\mathcal{B} : X \to Y$ *such that*

$$\forall x_1, x_2, x_3, x_4 \in X \quad f(x_1, x_2) < f(x_3, x_4) \Rightarrow g(\mathcal{B}(x_1), \mathcal{B}(x_2)) \leq g(\mathcal{B}(x_3), \mathcal{B}(x_4))$$
$$\forall y_1, y_2, y_3, y_4 \in Y \quad g(y_1, y_2) < g(y_3, y_4) \Rightarrow f(\mathcal{B}^{-1}(y_1), \mathcal{B}^{-1}(y_2)) \leq g(\mathcal{B}^{-1}(y_3), \mathcal{B}^{-1}(y_4))$$

Let $\mathcal{M}^* : X \to Y$ *be a bijection which maximizes*

$$C(\mathcal{M}) = \sum_{x,y \in X} f(x,y)g(\mathcal{M}(x), \mathcal{M}(y))$$

over all bijections $\mathcal{M}$. *Then* $\mathcal{M}^* \in H$

Proof:

Apply lemma 3.1 part (2) to $\langle f(x,y)\rangle$ *and* $\langle g(\mathcal{M}^*(x), \mathcal{M}^*(y))\rangle$, *and consider* $\mathcal{M}$ *as effecting permutations over these sequences. Since elements in* $\mathcal{H}$ *correspond to mappings satisfying the assumption of lemma 3.1 part (2), this shows that if* H *is non empty, then* $\langle f(x,y)\rangle$ *and* $\langle g(\mathcal{M}^*(x), \mathcal{M}^*(y))\rangle$ *satisfy the stated criterion of this corollary.*

To summarize these theorems, we have shown that maximizing C makes the order of distances between points the same in V_{in} as in V_{out} under M, if this is possible under any bijection. We also showed earlier that bijections between continuous spaces which have the property of preserving distance orderings are certainly homeomorphisms. Therefore, maximizing C provides a way of finding what is, in the continuous limit, a natural homeomorphism which preserves distance relations between the two spaces. Consequently, C satisfies an important criterion for being a reasonable measure for the quality of a "topographic" mapping.

4 Relation of C to other measures of neighbourhood preservation

We now illustrate how particular choices of F and G lead to the equivalence of optimizing C to some other approaches.

4.1 Minimal wiring

In minimal wiring [Mitchison & Durbin 1986, Durbin & Mitchison 1990], a good mapping is one that maps points that are nearest neighbours in V_{in} as close as possible in V_{out}, where closeness in V_{out} is measured by for instance euclidean distance raised to some power. The motivation here is the idea that it is often useful in processing e.g. sensory data to perform computations that are local in some space of input features V_{in}. To do this in V_{out} (e.g. the cortex) the images of neighbouring points in V_{in} need to be connected; the similarity function in V_{out} is intended to capture the cost of the wire (e.g. axons) required to do this. Minimal wiring is equivalent to equation 1 for

$$F(i,j) = \begin{cases} 1 & : \text{ i, j neighbouring} \\ 0 & : \text{ otherwise} \end{cases}$$

$$G(M(i), M(j)) = ||M(i) - M(j)||^p$$

For the cases of 1-D or 2-D square arrays investigated in [Mitchison & Durbin 1986, Durbin & Mitchison 1990], neighbours are taken to be just the 2 or 4 adjacent points in the array respectively.

4.2 Minimal path length

An alternative to minimal wiring is to say that a good map is one such that in moving between nearest neighbours in V_{out} one moves the least possible distance in V_{in}. This is for instance the mapping required to solve the Travelling Salesman Problem (TSP) where V_{in} is the distribution of cities and V_{out} is the one-dimensional tour. This goal is implemented by the elastic net algorithm [Durbin & Willshaw 1987, Durbin & Mitchison 1990, Goodhill & Willshaw 1990], which measures similarity in V_{in} by squared distances:

$$F(i,j) = ||v_i - v_j||^2$$

$$G(i,j) = \begin{cases} 1 & : \text{ i, j neighbouring} \\ 0 & : \text{ otherwise} \end{cases}$$

where v_k is the position of point k in V_{in} (we have only considered here the regularization term in the elastic net energy function). It can be seen that minimal wiring and minimal path length are virtually symmetrical cases under equation 1. Their relationship is discussed further in [Durbin & Mitchison 1990] (see also [Mitchison 1995]).

4.3 The approach of Jones et al

[Jones et al 1991] investigated the effect of the shape of the cortex (V_{out}) relative to the lateral geniculate nuclei (V_{in}) on the overall pattern of ocular dominance columns in the cat and monkey, using an optimization approach. They desired to keep both neighbouring cells in each LGN (as defined by a hexagonal array), and anatomically corresponding cells between the two LGNs, nearby in the cortex (also a hexagonal array). Their formulation of this problem can be expressed as a maximization of C when

$$F(i,j) = \begin{cases} 1 & : \text{ i, j neighbouring, corresponding} \\ 0 & : \text{ otherwise} \end{cases}$$

and

$$G(i,j) = \begin{cases} 1 & : \text{ i, j first or second nearest neighbours} \\ 0 & : \text{ otherwise} \end{cases}$$

For 2-D V_{in} and V_{out} they found a solution such that if $F(i,j) = 1$ then $G(M(i), M(j)) = 1, \forall i,j$. Alternatively this problem could be expressed as a minimization of C when $G(i,j)$ is the stepping distance (see below) between positions in the V_{out} array.

4.4 The approach of Villmann et al

In [Villmann et al 1994] the primary concern is with the case of data in a geometric, continuous input space V_{in} being mapped onto a square array of points, where there are in general many more points in V_{in} than V_{out}. We consider the situation after some process of for instance vector quantization has occurred, so that there are now the same number of points in both spaces, the positions of points in the input space are fixed, and we can talk about the degree of neighbourhood preservation of the bijective mapping between these two sets of points. [Villmann et al 1994] give a way of defining neighbourhoods, in terms of "masked Voronoi polyhedra" (see also [Martinetz & Schulten 1994]). This defines a neighbourhood structure where for any two points (two centers of masked Voronoi polyhedra), there is an integer dissimilarity which defines the "stepping distance" between them (cf [Kendall 1971]). They define a series of measures $\Phi(k)$ which give the number of times points which are neighbours in one space are mapped stepping distance k apart in the other space (they consider all indices for both directions of the map). If all the $\Phi(k)$ are zero they call the mapping "perfectly topology preserving". It is intuitively clear that such a map is the discrete approximation to a topographic homeomorphism as defined above. However, the measurement of discrepancies when such a map does not exist is not equivalent to optimizing C.

Their formulation has the useful property that the distribution of non-zero $\Phi(k)$ gives information about the scale of "discontinuities" in the map. However, no rule is specified in [Villmann et al 1994] for combining the $\Phi(k)$ into a single number that specifies the overall quality of a particular mapping, and thus allows different mappings to be directly compared. A simple way to do this would be to take a sum of the $\Phi(k)$ weighted by some function of k. If this function were increasing with k (and good mappings were defined to be the minima of the product), this would express a desire to minimize large scale discontinuities at the expense of small scale ones. A function that was decreasing with k would be equally plausible: which function is most appropriate depends on the problem.

4.5 The Topographic Product

The "topographic product" was introduced in [Bauer & Pawelzik 1992], based on ideas first discussed in the context of nonlinear dynamics. It is somewhat similar to the approach of [Villmann et al 1994],[2] in that they define a series of measures $Q(i,j)$ which give information about the preservation of neighbourhood relations at all possible scales. Briefly, $Q_1(i,j)$ is the distance between point i in the input space and its jth nearest neighbour as measured by distance orderings of their images in the output space, divided by the distance between point i in the input space and its jth nearest neighbour as measured by distance orderings in the input space. $Q_2(i,j)$ gives analogous information where i and j are points in the output space. [Bauer & Pawelzik 1992] specify a way of combining the Q's to yield a single number P, the "topographic product", which defines the quality of a particular mapping. Although originally expressed in terms of geometric spaces, the distance orderings in this definition could equally well be replaced by abstract similarity orderings that do not have a geometric interpretation. Again the concern is with orderings: the perfect map ($P = 1$) is clearly a topographic homeomorphism.

4.6 Metric Multidimensional Scaling

Metric multidimensional scaling (metric MDS) is a technique originally developed in the context of psychology for representing a set of N "entities" (e.g. subjects in an experiment) by N points in a low (usually two) dimensional space. For these entities one has a matrix which gives the numerical dissimilarity (or similarity) between each pair of entities. The aim of metric MDS is to position points representing entities in the low dimensional space so that the set of distances between each pair of points matches as closely as possible the given set of dissimilarities. The particular objective function optimized is the summed squared deviations of distances from dissimilarities. The original method was presented in [Torgerson 1952]; for a recent review see [Goodhill, Simmen & Willshaw 1995].

In terms of the framework presented above, F is a dissimilarity matrix. Note that there may not be a geometric space of any dimensionality for which these dissimilarities can be represented by distances (e.g. if the dissimilarities do not satisfy the triangle inequality), in which case V_{in} does not have a geometric

[2] The topographic product was introduced first; however it is more convenient for exposition purposes to explain it second.

interpretation. V_{out} is the low dimensional (continuous) space in which the points representing entities are positioned, and G is euclidean distance in V_{out}. Metric MDS says to choose the mapping M (by adjusting the positions of points in V_{out}) which minimises

$$\sum_{i=1}^{N} \sum_{j<i} (F(i,j) - G(M(i), M(j)))^2 \tag{3}$$

If the minimum value of this is zero the resulting map is clearly a topographic homeomorphism. However, this arrangement violates our assumption that G is fixed, since the G matrix varies with the mapping. We can though see the relationship to the C measure as follows. Expanding out the square gives

$$\sum_{i=1}^{N} \sum_{j<i} (F(i,j)^2 + G(M(i), M(j))^2 - 2F(i,j)G(M(i), M(j))) \tag{4}$$

The first term in equation 4 is given by the input dissimilarity matrix and is therefore fixed. The last term is just twice the C measure. This could be made arbitrarily large by moving the points in the output space far apart. However, the middle term acts to keep points in V_{out} close together. Thus the optimum is given by an appropriate balance between these two terms, and is in general not the same as optimizing C.

5 Discussion

5.1 Relation of C to quadratic assignment problems

In formulating neighbourhood preservation in terms of the C measure we have in fact set it within the much-studied class of quadratic assignment problems (QAPs).[3] These occur in many different practical contexts (see [Burkard 1984] for review), and take the form of finding the minimal or maximal value of an equation similar to C. An illustrative example is the optimal design of typewriter keyboards. Now $F(i, j)$ (say) is the average time it takes a typist to sequentially press locations i and j on the keyboard, while $G(p, q)$ is the average frequency with which letters p and q appear sequentially in text of a given language (note that in this example F and G are not necessarily symmetrical). The keyboard that minimises average typing time will be the one that minimises the product

$$\sum_{i=1}^{N} \sum_{j=1}^{N} F(i,j)G(M(i), M(j))$$

(cf equation 1), where $M(i)$ is the letter that maps to location i. A large amount of theory has been developed for QAPs, which is directly applicable to the C measure.

5.2 Finding a good optimum of C

QAPs are in general known to be of complexity NP-hard. A large number of algorithms for both exact and heuristic solution have been studied (see e.g. [Burkard 1984, Simić 1991]). However, particular instantiations of F and G may make possible very efficient algorithms for finding good local optima, or alternatively may beset C with many bad local optima. Such considerations provide an additional practical constraint on what choices of F and G are most appropriate. For instance, the elastic net algorithm for the minimal path length problem is efficient when similarity in the input space is given by squared distance, but not by linear distance (see e.g. [Dayan 1993]).

[3] We thank Eric Mjolsness for drawing our attention to this link.

5.3 Biological considerations

A biological problem that has been much discussed in the context of mappings is to understand the pattern of interdigitation in primary visual cortex in cat and monkey of attributes of the visual image such as spatial position, orientation, ocular dominance, and so on. It has been proposed several times that the form of this map is a result of a desire to preserve neighbourhoods between a high dimensional space of features and the two dimensional cortex (see e.g. [Durbin & Mitchison 1990, Goodhill & Willshaw 1990, Obermayer et al 1992]). Usually geometric distance in the high dimensional space is taken to represent dissimilarity (we note briefly that this is a somewhat constraining assumption). The approach outlined in the present paper suggests that one way to proceed in addressing this and related biological mapping problems is to (1) formulate some reasonable way of specifying the similarity functions for the space of input features and the cortex (this could for instance be in terms of the correlations between different input variables, and intrinsic cortical connections, respectively (see e.g. [Goodhill & Löwel 1995])), then (2) explore which sets of mapping choices give maps resembling those seen experimentally, taking account of biological constraints.

Finally we suggest that a consideration of how to most closely achieve a topographic homeomorphism by, in addition to adapting the mapping, allowing the similarity function in the cortex to change (within some constrained range), could provide insight into the development of lateral connections in the cortex (cf [Katz & Callaway 1992]).

5.4 Limitations

We have explored only the case of one-to-one mappings. In many practical contexts there are many more points in V_{in} than V_{out}, and it is necessary to also specifiy a many-to-one mapping from points in V_{in} to N "exemplar" points in V_{in}, where N = number of points in V_{out}. It may be desirable to do this adaptively while simultaneously optimizing the form of the map from V_{in} to V_{out} (see e.g. [Yuille 1990]).

Additionally, one-to-one mappings are rare in biological contexts. Rather more frequently axonal arbors form many-to-many connections with dendritic trees. There are several conceivable ways in which for instance the C measure could be generalized to allow for a weighted match of each point in V_{in} to many in V_{out}, and vice-versa.

6 Conclusions

We have attempted to bring out into the open a number of issues regarding topographic or neighbourhood preserving mappings that have sometimes been obscure or confused in the past. By enumerating choices and making connections between different methods for measuring neighbourhood preservation, we hope to bind different perspectives on this problem into a more coherent whole. In particular, we have tried to establish the following points.

- At a very general level, we may talk about neighbourhood relations within a space in terms of a similarity function, independent of any geometric considerations.
- A mapping which preserves similarity orderings between two such spaces, called here a topographic homeomorphism, is worthy of the name "neighbourhood preserving" or "topographic".
- Many existing mathematical formulations of neighbourhood preservation (for instance versions of the C measure defined above) specify a topographic homeomorphism (if one exists) as the optimal map (though this remains to be rigorously proved in some cases).

- When a topographic homeomorphism does not exist, there are a multitude of different approaches for quantifying the degree to which similarity orderings are not preserved. Several of these can be expressed as special cases of the C measure.
- There are no obvious a priori criteria for choosing between these different sets of choices. However, once a particular mapping problem is specified it may become clear which approach is most appropriate (for instance, a minimal path length formulation for solving TSP-type problems).
- The explicit link to quadratic assignment problems may provide a useful source of both algorithms and inspiration for further addressing topographic maps, and particularly how they can be efficiently found.

Acknowledgements

GJG is very grateful to Christopher Longuet-Higgins for early inspiration, to Martin Simmen, Klaus Pawelzik and Hans-Ulrich Bauer for stimulating discussions of these ideas, and to Eric Mjolsness and Dimitris Tsioutsias for helpful pointers to the quadratic assignment problem literature.

Bibliography

Bauer, H.U. & Pawelzik, K.R. (1992). Quantifying the neighbourhood preservation of self-organizing feature maps. *IEEE Trans. Neural Networks*, **3**, 570-579.

Burkard, R.E. (1984). Quadratic assignment problems. *Europ. Journ. Oper. Res.*, **15**, 283-289.

Dayan, P.S. (1993). Arbitrary elastic topologies and ocular dominance. *Neural Computation*, **5**, 392-401.

Durbin, R. & Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, **343**, 644-647.

Durbin, R. & Willshaw, D.J. (1987). An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, **326**, 689-691.

Erwin, E., Obermayer, K. & Schulten, K. (1992). Self-organizing maps: convergence properties and energy functions. *Biol. Cybern.*, **67**, 47-55.

Goodhill, G.J. & Löwel, S. (1995). Theory meets experiment: correlated neural activity helps determine ocular dominance column periodicity. *Trends in Neurosciences*, in press.

Goodhill, G.J., Simmen, M., & Willshaw, D.J. (1995). An evaluation of the use of Multidimensional Scaling for understanding brain connectivity. *Phil. Trans. Roy. Soc. B*, **348**, 265-280.

Goodhill, G.J. & Willshaw, D.J. (1990). Application of the elastic net algorithm to the formation of ocular dominance stripes. *Network*, **1**, 41-59.

Jones, D.G., Van Sluyters, R.C. & Murphy, K.M. (1991). A computational model for the overall pattern of ocular dominance. *J. Neurosci.*, **11**, 3794-3808.

Katz, L.C. & Callaway, E.M. (1992). Development of local circuits in mammalian cortex. *Ann. Rev. Neurosci.*, **15**, 31-56.

Kendall, D.G. (1971). Construction of maps from "odd bits of information". *Nature*, **231**, 158-159.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, **43**, 59-69.

Kohonen, T. (1988). Self-organization and associative memory. Springer, Berlin.

Kruskal, J.B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.

Kruskal, J.B. (1964). Non-metric multidimensional scaling: a numerical method. *Psychometrika*, **29**, 115-129.

Krzanowski, W.J. (1988). Principles of multivariate analysis: a user's perspective. Oxford statistical science series; v. 3. Oxford University Press.

Marr, D. (1982). Vision. W.H. Freeman and Company, New York.

Martinetz, T. & Schulten, K. (1994). Topology representing networks. *Neural Networks*, **7**, 507-522.

Mitchison, G. (1995). A type of duality between self-organizing maps and minimal wiring. *Neural Computation.*, **7**, 25-35.

Mitchison, G. & Durbin, R. (1986). Optimal numberings of an N × N array. *SIAM J. Alg. Disc. Meth.*, **7**, 571-581.

Obermayer, K., Blasdel, G.G. & Schulten, K. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Phys. Rev. A*, **45**, 7568-7589.

Sejnowski, T.J., Koch, C. & Churchland, P.S. (1988). Computational Neuroscience. *Science*, **241**, 1299-1306.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting and clustering. *Science*, **210**, 390-398.

Simić, P.D. (1991). Constrained nets for graph matching and other quadratic assignment problems. *Neural Computation*, **3**, 268-281.

Takane, Y., Young, F.W. & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.

Torgerson, W.S. (1952). Multidimensional Scaling, I: theory and method. *Psychometrika*, **17**, 401-419.

Udin, S.B. & Fawcett, J.W. (1988). Formation of topographic maps. *Ann. Rev. Neurosci.*, **11**, 289-327.

Villmann, T., Der, R., Herrmann, M. & Martinetz, T. (1994). Topology preservation in SOFMs: general definition and efficient measurement. In: *Informatik Aktuell - Fuzzy-Logik*, Ed. B. Reusch, Springer-Verlag, 159-166.

Yuille, A.L. (1990). Generalized deformable models, statistical physics, and matching problems. *Neural Computation*, **2**, 1-24.