Quantifying neighbourhood preservation in topographic mappings

Geoffrey J. Goodhill[1], Steven Finch[2], & Terrence J. Sejnowski[3]

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

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

3. The Howard Hughes Medical Institute
The Salk Institute for Biological Studies
10010 North Torrey Pines Road
La Jolla, CA 92037, USA
&
Department of Biology
University of California San Diego
La Jolla, CA 92037, USA

Institute for Neural Computation Technical Report Series

No. INC-9505

November 1995

Abstract

Mappings that preserve neighbourhood relationships are relevant in both practical and biological contexts. It is important to be clear about precisely what preserving neighbourhoods could mean. We give a definition of a "perfectly neighbourhood preserving" map, which we call a topographic homeomorphism, and prove that this has certain desirable properties. When a topographic homeomorphism does not exist (the usual case), many choices are available for quantifying the quality of a map. We introduce a particular measure, C, which has the form of a quadratic assignment problem. We also discuss other measures that have been proposed, some of which are related to C. A comparison of seven measures applied to the same simple mapping problem reveals interesting similarities and differences between the measures, and challenges common intuitions as to what constitutes a "good" map.

1 Introduction

Problems of mapping occur frequently both in understanding biological processes and in formulating abstract methods of data analysis. An important concept in both domains is that of a "neighbourhood preserving" map, also sometimes referred to as a topographic, topological, or topology-preserving map. Intuitively speaking, such maps take points in one space to points in another space such that nearby points map to nearby points (and sometimes in addition far-away points map to far-away points). Such maps are useful in data analysis and data visualization, where a common goal is to represent data from a high dimensional space in 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]). Just a few of the algorithms that have found application in this context are principal components analysis (PCA), multidimensional scaling [Torgerson 1952, Shepard 1962a, Shepard 1962b, Kruskal 1964a, Kruskal 1964b], Sammon mappings [Sammon 1969], and neural network algorithms such as the self-organizing feature map (SOFM) [Kohonen 1982, Kohonen 1988] or the elastic net [Durbin & Willshaw 1987]. One hope is that by preserving neighbourhoods in the mapping it will be possible to more clearly see structure in the high-dimensional data such as clusters, or that this type of dimension-reduction will reveal that the data occupies a lower-dimensional subspace than was originally apparent. In neurobiology there are many examples of neighbourhood-preserving mappings, for instance between the retina and more central structures [Udin & Fawcett 1988]. Another type of neighbourhood-preserving mapping in the brain is that for instance from the visual world to cells in the primary visual cortex which represent a small line segment at a particular position and orientation in the visual scene [Hubel & Wiesel 1977]. A possible goal of such biological maps is to represent nearby points in some sensory "feature space" by nearby points in the cortex [Durbin & Mitchison 1990]. This could be desirable since sensory inputs are often locally redundant: for instance in a visual scene pixel intensities are highly predictable from those of their neighbours. In order to perform "redundancy reduction" (e.g. [Barlow 1989]), it is necessary to make comparisons between the output of cells in the cortex that represent redundant inputs. Two ways this could be achieved are either by making a direct connection between these cells, or by constructing a suitable higher-order receptive field at the next level of processing. In both cases, the total length of wire required can be made short when nearby points in the feature space map to nearby points in the cortex (see [Cowey 1979, Durbin & Mitchison 1990, Nelson & Bower 1990, Mitchison 1991, Mitchison 1992] for further discussion).

So far we have discussed neighbourhood preservation in intuitive terms. However, it is vital to ask what this intuitive idea might mean more precisely, i.e. exactly what computational principles are being addressed by such mappings. 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 mappings in the brain might be pursuing. A large number of choices have to be made to reach a precise mathematical measure of neighbourhood preservation, as we discuss below. 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 problem to problem. Several measures of neighbourhood preservation have recently been proposed which implement particular sets of choices. In many cases there are few *a priori* grounds for choosing between different formulations: rather each may be useful for different types of problem. 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 the choices available and their appropriateness for different types of 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 difficult, 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. However, the SOFM [Kohonen 1982] exists only at the algorithmic level as it is not following the gradient of an objective function [Erwin et al 1992]. In particular, given a map, it does not provide a number measuring its quality.

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 various measures can be interpreted. The second is to explore in more detail one set of choices, which are still at a fairly general level. In section 2 we first propose a rather general concept of neighbourhood-preservation, and prove a theorem showing that it has a useful property. Section 3 introduces a particular quantitative measure, and proves that it implements the general concept. In section 4 we show how some other measures can be fitted into this overall scheme, and discuss some alternative approaches. In section 5 we consider some solutions to a very simple mapping problem, and compare how various measures rate the quality of the different solutions.

All mappings under discussion in this paper are assumed to be bijective (i.e. one-to-one). In addition, we assume that in each space there exists a "similarity structure", which specifies for every pair of points its degree of similarity. In a simple case, this similarity is just euclidean distance between points in a geometric space. However, the similarity structure need not have a geometric interpretation. For example, for purely "nearest neighbour" structure, similarities are binary.

2 Preserving similarities

There are at least two obvious choices for defining a mapping that "perfectly" preserves the neighbourhood structure of the data in one space in another space. The first is to say that the mapping must *preserve similarities;* that is, for each pair of points in one space, its similarity should be equal to the similarity of the images of those points in the other space. This imposes the strongest possible constraint on the representation. The second is to say that the mapping must only preserve similarity *orderings*. That is, rather than comparing the absolute values of the similarity between pairs of points in one space and the similarity between their images in the other, one is concerned only that their relative orderings within the two sets of similarities are the same. Although this appears rather weaker, it in fact imposes strong constraints on the mapping (as been extensively discussed in the context of non-metric multidimensional scaling: see e.g. [Shepard 1980, Krzanowski 1988]). We choose this second type of constraint to investigate further.

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 as 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:*

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

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*. For discrete spaces 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. Particular examples of this are metric multidimensional scaling [Torgerson 1952] and Sammon mappings [Sammon 1969], 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 that attempts to quantify purely ordering violations are the "stress" functions of non-metric multidimensional scaling [Kruskal 1964a, Kruskal 1964b, Takane et al 1977]. Most of the other measures of discrepancy that have been proposed use absolute similarity values. The choices they make will be discussed further later. We next introduce one particular measure and demonstrate that it has some useful properties.

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 (or dissimilar) they are. Call these functions F for V_{in} and G for V_{out}. Then we define a cost functional C as follows

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

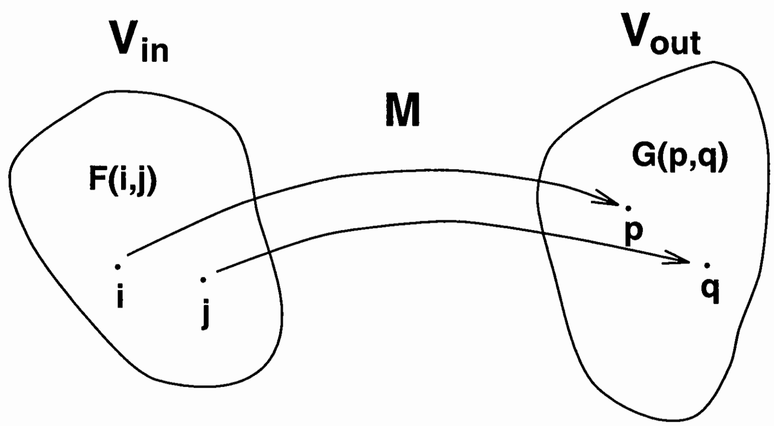

Figure 1: The mapping framework.

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 not be possible to interpret the points as lying in some geometric space. There are several 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 maximizing C will find it. Secondly, we show that some other proposals for discrepancy measures are in fact equivalent to C for particular choices of F and G. Thirdly, C is in the form of a quadratic assignment problem (see below), a class of problems for which a great deal of theory exists.

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 represents 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 some of the concepts and results of this field can be applied 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.[1]

[1] A related result can be found in [Hardy et al 1934, page 261].

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*

$$\begin{aligned} d &= C(\mathcal{P}) - C(\mathcal{P}') \\ &= \sum_{i=1}^{\infty} x_i \mathcal{P}(i) - \sum_{i=1}^{\infty} x_i \mathcal{P}'(i) \\ &= x_i \mathcal{P}(i) + x_j \mathcal{P}_j - x_i \mathcal{P}(j) - x_j \mathcal{P}(i) \\ &= (x_i - x_j)(\mathcal{P}(i) - \mathcal{P}(j)) \end{aligned}$$

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*

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

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 (if one exists). Consequently, C satisfies an important criterion for being a reasonable measure for the quality of a "topographic" mapping.

3.3 Relation to quadratic assignment problems

Formulating neighbourhood preservation in terms of the C measure sets it within the well-studied class of quadratic assignment problems (QAPs). These occur in many different practical contexts, and take the form of finding the minimal or maximal value of an equation similar to C (see [Burkard 1984] for a general review, and [Lawler 1963, Finke et al 1987] for more technical discussions). An illustrative example is the optimal design of typewriter keyboards [Burkard 1984]. If $F(i, j)$ 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), then the keyboard that minimizes average typing time will be the one that minimizes 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. The substantial theory developed for QAPs is directly applicable to the C measure. As a concrete example, QAP theory provides several different ways of calculating bounds on the minimum and maximum values of C for each problem. This could be very useful for the problem of assessing the quality of a map relative to the unknown best possible (rather than simply making a comparison between two maps). One particular case is the eigenvalue bound [Finke et al 1987]. If the eigenvalues of symmetric matrices F and G are λ_i and μ_i respectively, such that $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ and $\mu_1 \ge \mu_2 \ge \ldots \ge \mu_n$, then it can be shown that $\sum_i \lambda_i \mu_i$ gives a lower bound on the value 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], the references in [Finke et al 1987], and [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.

4 Some other measures of neighbourhood preservation

We now discuss a number of other measures of neighbourhood preservation that have been proposed, in roughly chronological order.

4.1 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 reviews see [Shepard 1980, Young 1987].

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 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 operates by selecting the mapping M (by adjusting the positions of points in V_{out}) which minimizes

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

If the minimum value of this is zero the resulting map is clearly a topographic homeomorphism. However, there is a departure from our framework up to now since the G matrix is not fixed (though the underlying metric is). If it is assumed instead that points in the output space are at fixed locations, and the space of possible mapping consists of the space of possible permutations, then equation 3 is in fact the same as the C measure. Expanding out the square in equation 3 gives

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

The sums over the F^2's and G^2's are both independent of the mapping (under the assumption above), leaving twice the C measure. Thus the metric MDS minimization for fixed output space structure is identical to maximizing C.

4.2 Nonmetric Multidimensional Scaling

The aim of nonmetric MDS (NMDS) is to position points as in metric MDS, except that now it is attempted to match only the ordering of similarities between the input and output spaces, rather than the absolute values [Shepard 1962a, Shepard 1962b]. The mathematical measure used to quantify deviations from perfect ordering is somewhat ad hoc, and in fact often varies between different packaged software routines that implement NMDS. The first measure proposed was called STRESS [Kruskal 1964a, Kruskal 1964b], and is given in our notation by

$$\text{STRESS} = \sqrt{\frac{\sum_{i=1}^{N} \sum_{j<i} (G(i,j) - D(i,j))^2}{\sum_{i=1}^{N} \sum_{j<i} G(i,j)^2}} \quad (5)$$

where the $D(i,j)$'s are "disparities". These are target values for each $G(i,j)$ such that if the G's achieved these values, then ordering would be preserved. Thus STRESS is zero for a topographic homeomorphism. An algorithm for calculating the disparities is given in [Kruskal 1964b]. Another commonly used version of NMDS is ALSCAL [Takane et al 1977], which uses an objective function called SSTRESS where the differences between the G's and D's are raised to the power 4, and the normalization is by a sum of quartic D's rather than squared G's. A different algorithm is used for calculating disparities than the Kruskal method. Note that STRESS and SSTRESS do not purely measure ordinal discrepancies, since they are expressed in terms of absolute similarity values rather than orderings.

4.3 Sammon mappings

The Sammon approach [Sammon 1969] is similar to metric MDS in that it tries to match similarities exactly by moving points around in V_{out}, but different in that it uses normalization terms which make the mapping nonlinear. The objective function is

$$\frac{1}{\sum_{i=1}^{N}\sum_{j<i} F(i,j)} \sum_{i=1}^{N}\sum_{j<i} \frac{(F(i,j) - G(M(i),M(j)))^2}{F(i,j)}. \tag{6}$$

It has recently been argued that this approach produces better maps than for instance the SOFM [Bezdek & Pal 1995, Lowe & Tipping 1995]. The Sammon formula is not symmetric to interchange of F and G: the quality of the map depends on which space is considered the input space and which the output space.

4.4 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 the C measure for

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

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

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.

4.5 Minimal path length

A counterpart 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 Traveling 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:

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

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

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], where the abilities of minimal wiring and minimal path length are compared with regard to reproducing the structure of the map of orientation selectivity in primary visual cortex (see also [Mitchison 1995]).

4.6 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

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

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 $G(p,q)$ is the stepping distance (see below) between positions in the V_{out} array. They found this gave appropriate behaviour for the problem addressed.

4.7 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 seems 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 decreases with k could also be plausible: which function is most appropriate depends on the problem.

4.8 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 ($P = 0$) is clearly a topographic homeomorphism. [Bauer & Pawelzik 1992] show the application of the topographic product to dimension-reducing mappings of speech recognition data.

4.9 The approach of Bezdek and Pal

[Bezdek & Pal 1995] also argue for a criterion that preserves similarity orderings rather than actual similarities. They call such a transformation "metric topology preserving" or MTP. The method they propose for calculating the discrepancy from an MTP map is to use a rank correlation coefficient between similarities in the two spaces, in particular Spearman's ρ. This is defined as the linear correlation coefficient of the ranks (see e.g. [Press et al 1988]):

$$\rho = \frac{\sum_i (R_i - \overline{R})(S_i - \overline{S})}{\sqrt{\sum_i (R_i - \overline{R})^2}\sqrt{\sum_i (S_i - \overline{S})^2}}$$

where R_i and S_i are the corresponding rankings in the ordered lists of the F's and and G's. ρ has the useful property that it is bounded in the range $[-1, 1]$, and [Bezdek & Pal 1995] prove that $\rho = 1$ corresponds to an MTP map. Our C measure is not so bounded, and is useful for comparing different solutions to the same mapping problem rather than solutions to different problems. However, the same is to some extent true also for ρ, since $\rho = 0.5$ for one problem (e.g. a mapping from three dimensions to two dimensions) may indicate a very different quality of solution from a $\rho = 0.5$ for another problem (e.g. a mapping from fifty dimensions to two dimensions). Perhaps the most important difference between the C and ρ measures is that ρ measures discrepancies in terms of rank order violations, whereas C uses similarity values. Interesting comparisons are made in [Bezdek & Pal 1995] between the performance as measured by ρ of PCA, Sammon mappings and the SOFM applied to various mapping problems. They conclude that the SOFM generally performs significantly worse.

5 A comparison of measures

Comparisons between the performance of various mapping algorithms on particular problems have been illuminating [Durbin & Mitchison 1990, Bezdek & Pal 1995, Lowe & Tipping 1995]. Here we ask a different set of questions. How do different measures compare in rating the same maps? Do the measures generally give a consistent ordering for different maps? How well does this ordering compare with intuitive assessments? Answers to these questions reveal more about the relationships between different measures, and aid the choice of an appropriate measure for particular problems.

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

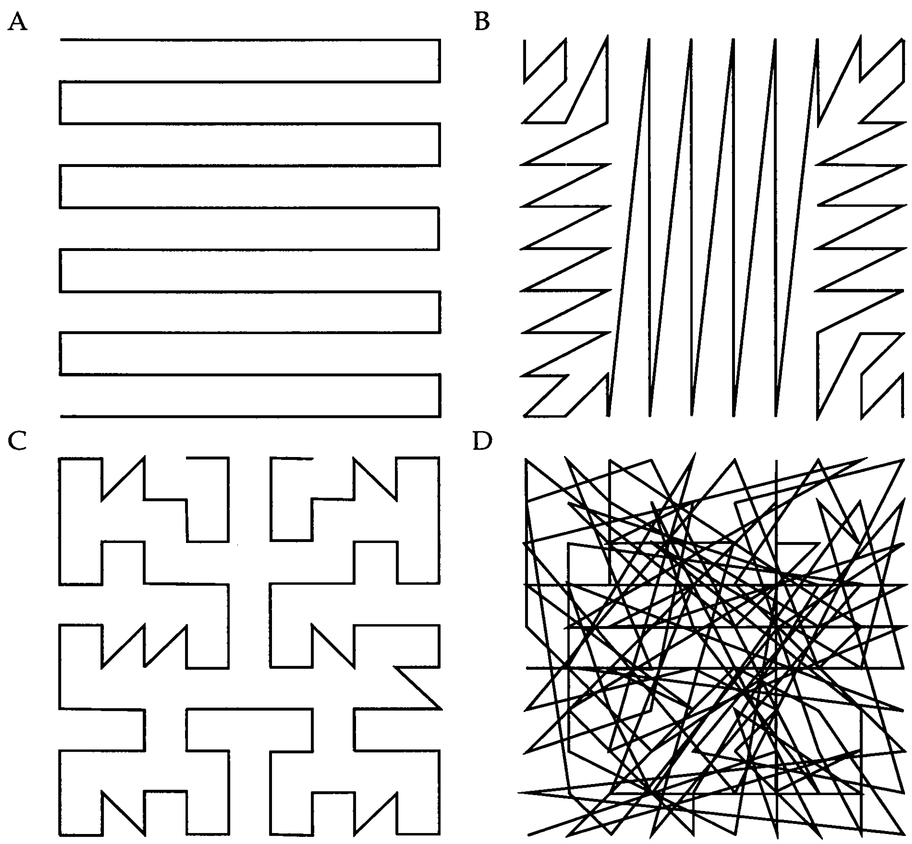

Figure 2: Some alternative mappings between a square array and a line.

We consider the very simple case of mapping between a regular 10×10 square array of points (V_{in}) and a regular 1×100 linear array of points (V_{out}). This is a very simple example of dimension reduction, and has the virtue that solutions are easily represented. Figure 2 shows the four alternative maps considered (labelled A-D). The F and G dissimilarity functions are taken to be euclidean distances in each array, except for appropriate modifications for the minimal path and minimal wiring measures as described earlier. Map A is an optimal minimum path solution. Maps B and C are taken from [Durbin & Mitchison 1990]: map B is an optimal minimal wiring solution [Mitchison & Durbin 1986], and map C is that found by the elastic net algorithm (for parameters see [Durbin & Mitchison 1990]).[3] Map D is a random mapping. Visual inspection suggests that this corpus should provide a range of different quality values.

The actual numbers yielded by some of the measures discussed above are shown in table 1, and the relative ordering assigned to the different maps for all measures is shown in table 2. There are several points to note. As expected, all of the measures rate the random map D the worst, and $\rho \approx 0$ for this map. The Sammon, C, ρ and STRESS measures show remarkable agreement in their orderings (rating map A the best) despite being each formulated on quite different mathematical principles. It is a common intuition in the mapping literature that for a case such as shown in figure 2 the best map is one that resembles a Peano curve, i.e. map C. However, only the topographic product agrees with this assessment. As a control, we repeated the calculations using dissimilarities in both spaces that were euclidean distances raised to various powers between 0.5 and 2.0. In all cases the ordering for each measure was unchanged, except for minimum wiring, where the orderings obtained for powers 0.5 and 0.6 (in addition to power 1.0) are shown in table 2 (see also [Durbin & Mitchison 1990]). With these additions, every possible ordering of the maps is represented in the table, given that D is always last. Spearman's ρ has the perhaps

[3]Qualitatively similar maps would be found by suitable variants of the SOFM, such as [Angeniol et al 1988].

	Min wiring	Min path	TP	Sammon	C	ρ	STRESS
Type	Decrease	Decrease	See caption	Decrease	Increase	Increase	Decrease
Map A	990	99.0	−0.04608	38.94	1066000	0.6498	0.423
Map B	914	159.9	−0.04843	39.98	1057000	0.6413	0.442
Map C	1140	102.3	−0.03244	49.93	1021000	0.6332	0.519
Map D	5458	484.9	−0.1638	70.02	890300	0.0502	0.594

Table 1: Costs (to four significant figures) for the four maps given by measures discussed in the text. "Type" tells how the measure changes with increasing mapping quality, so that for type "decrease" small numbers mean better maps. The topographic product is decreasing in absolute value, and the negative values reflect the fact the dimension of the output space (the line) is "too small". Spearman's ρ was calculated using the routine `spear` of [Press et al 1988], and STRESS by the program ALSCAL [Young & Harris 1990]. Metric MDS is not included since it clearly gives the same ordering of maps as the C measure.

Ranking	MW 1.0	MW 0.5	MW 0.6	Min path	TP	Sammon	C	ρ	STRESS
1	B	C	B	A	C	A	A	A	A
2	A	B	C	C	A	B	B	B	B
3	C	A	A	B	B	C	C	C	C
4	D	D	D	D	D	D	D	D	D

Table 2: Relative ordering of the 4 maps given by each measure. First three columns show minimum wiring for different powers (see text).

attractive feature that it is unchanged by any such monotonic transformation of the F's or G's. It also has the advantage of having a predictable value for random maps. More generally, it appears that the output of most measures is best treated as being at the ordinal level of measurement [Stevens 1951].

In terms of absolute costs, there is little discrimination between the three non-random maps given by the Sammon measure. This is because the range of F is from 1 to $\sqrt{200}$, whereas the range of G is from 1 to 100, and this inherent, map-independent mismatch dominates. Greater discrimination can be obtained by multiplying all G values by some number $\alpha < 1$. For instance, $\alpha = \frac{\sqrt{200}}{100}$ gives the values shown in table 3 for the Sammon measure.

For the minimum path measure the cost for map C is greater than that for map A by $8(\sqrt{2} - 1)$, since there are 8 diagonal segments in map C. Although the minimum of the elastic net energy function is at map C, in practice (for unknown reasons) the algorithm tends to find slightly longer, more "folded" maps. This is analogous to the effect observed for a different mapping problem, where striped solutions are obtained

	Sammon
Map A	0.1984
Map B	0.2192
Map C	0.4180
Map D	0.8197

Table 3: Values of the Sammon measure when values of G are scaled so that the range of G matches the range of F. Note that there is now greater discrimination between the values for different maps.

even though these are not the global optimum [Goodhill & Willshaw 1990, Dayan 1993]. A general issue we have not considered is how ease of implementation and efficient optimization might bias a choice of measure.

Why is the "intuitively appealing" map C so rarely rated the best? In general, the answer must lie in the way ordering violations at different scales are assessed (for further discussion see [Bauer & Pawelzik 1992]). For map C, the majority of points in the square map close on the line to their 4 neighbours in the square. However, for some points near the middle of the square, one neighbour is mapped to a point on the line a very large distance away (of the order of half the length of the line). This is far greater than the maximum equivalent distance in maps A or B, and perhaps accounts for the poor rating given to map C by several of the measures. Of course, it could be that a different map of the same qualitative form as map C would be rated better than maps A or B by some of the other measures. A more extensive analysis is clearly required: here our aim is simply to draw attention to these issues.

6 Discussion

6.1 Degeneracy

We have implicitly only considered the case where is it possible to arrange the F's and the G's in *strictly* monotonic sequences. If $F(i,j) = F(p,q)$ for $i,j \neq p,q$, or similarly for the G matrix (i.e. degeneracy), it is necessary for measures that aim to preserve orderings to specify how these cases will be treated. This issue has been extensively discussed for NMDS, where two approaches have been defined. The "tied" approach attempts to maintain degeneracy in the output space if it exists in the input space, whereas the "untied approach" does not [Kruskal 1964a]. This issue remains to be explicitly addressed for the other measures which use orderings.

6.2 Many-to-one mappings

We have discussed 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 specify 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}. For instance, shifting a point from one cluster to another may increase the "clustering cost", but by moving the positions of the cluster centers decrease the sum of this and the "continuity cost". The elastic net for instance [Durbin & Willshaw 1987] trades off these two contributions explicitly with a ratio that changes during the minimization, so that finally each cluster contains only one point and the continuity cost dominates (for discussions see [Simic 1990, Yuille 1990]). The SOFM trades off these two contributions implicitly.

6.3 Biological considerations

A biological problem that has been much discussed in the context of mappings is to understand the pattern of interdigitation in the primary visual cortex of cat and monkey of attributes of the visual image such as spatial position, orientation, ocular dominance, and disparity. 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 (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. 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), then (2) explore which sets of mapping choices give maps resembling those seen experimentally, taking account of biological constraints. 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]).

However, 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.

7 Conclusions

We have attempted to clarify a number of issues regarding topographic or neighbourhood preserving mappings. By enumerating choices and making connections between different methods for measuring neighbourhood preservation, we hope to achieve a coherent perspective on this problem from disparate approaches. 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".
- Several 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.
- 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, each of which has its own advantages and disadvantages.
- Applying several different measures to a set of solutions to the same mapping problem reveals interesting similarities and differences between the measures, suggesting numerous avenues for future work.

Acknowledgements

We are very grateful to Christopher Longuet-Higgins for early inspiration, to Martin Simmen, Klaus Pawelzik, Hans-Ulrich Bauer, Peter Dayan and Paul Viola for stimulating discussions of these ideas, to Hans-Ulrich Bauer for very kindly supplying us with code for calculating topographic products, and to Eric Mjolsness and Dimitris Tsioutsias for helpful pointers to the quadratic assignment problem literature.

Bibliography

Angeniol, B., de la Croix Vaubois, G. & Le Texier, J. (1988). Self-organizing feature maps and the traveling salesman problem. *Neural Networks*, **1**, 289-293.

Barlow, H.B. (1989). Unsupervised learning. *Neural Computation*, **1**, 295-311.

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

Bezdek, J.C. & Pal, N.R. (1995). An index of topological preservation for feature extraction. *Pattern Recognition*, **28**, 381-391.

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

Cowey, A. (1979). Cortical maps and visual perception. *Qua. Jou. Exper. Psychol.*, **31**, 1-17.

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 traveling 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.

Finke, G., Burkard, R.E. & Rendl, F. (1987). Quadratic assignment problems. *Annals of Discrete Mathematics*, **31**, 61-82.

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

Hardy, G.H., Littlewood, J.E. & Pólya, G. (1934). Inequalities. Cambridge University Press.

Hubel, D.H. & Wiesel, T.N. (1977). Functional architecture of the macaque monkey visual cortex. *Proc. R. Soc. Lond. B*, **198**, 1-59.

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

Kruskal, J.B. (1964b). 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.

Lawler, E.L. (1963). The quadratic assignment problem. *Management Science*, **9**, 586-599.

Lowe, D. & Tipping, M. (1995). Nonlinear topographic feature extraction using neural networks: a comparison. Manuscript submitted for publication.

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. (1991). Neuronal branching patterns and the economy of cortical wiring. *Proc. Roy. Soc. B.*, **245**, 151-158.

Mitchison, G. (1992). Axonal trees and cortical architecture. *Trends Neurosci.*, **15**, No. 4, 122-126.

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 $N \times N$ array. *SIAM J. Alg. Disc. Meth.*, **7**, 571-581.

Nelson, M.E. & Bower, J.M. (1990). Brain maps and parallel computers. *Trends Neurosci.*, **13**, 403-408.

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.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1988). Numerical recipes in C: the art of scientific computing. Cambridge University Press: Cambridge.

Sammon, J.W. (1969). A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.*, **18**, 401-409.

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

Shepard, R.N. (1962a). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, **27**, 125-140.

Shepard, R.N. (1962b). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, **27**, 219-246.

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

Simic, P.D. (1990). Statistical mechanics as the underlying theory of 'elastic' and 'neural' optimizations. *Network*, **1**, 89-103.

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

Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In *Handbook of experimental psychology*, ed S.S. Stevens, 1-49. New York: Wiley.

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.

Young, F. W. (1987). Multidimensional scaling: history, theory, and applications. Hillsdale, New Jersey: Lawrence Erlbaum.

Young, F. W. & Harris, D.F. (1990). Multidimensional scaling: procedure ALSCAL. In *SPSS base system user's guide* (ed. M. Norusis), 397-461. Chicago: SPSS.

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