Pattern Recognition Letters 11 (1990) 313-319
North-Holland

May 1990

Kohonen's algorithm for the numerical parametrisation of manifolds

Michael ALDER, Roberto TOGNERI, Edmund LAI and Yianni ATTIKIOUZEL
Department of Electrical and Electronic Engineering, University of Western Australia, Nedlands, W.A. 6009, Australia

Received 22 June 1989

Abstract: T. Kohonen has described an algorithm for fitting a k-dimensional grid of points to a set of points taken from a k-manifold in $\mathbb{R}^n$, for $k \leqslant n$. The algorithm is inspired by a neural model and bears some of the marks of its ancestry. In this paper we show that if the process converges, it converges to a locally 1-1 mapping of the grid onto the manifold. Hitherto this result has only been proved for the case where $k=1$.

0. Introduction

The analytic process of parametrising manifolds is well known to Mathematicians and Physicists, [1] and is becoming increasingly familiar to Engineers concerned with, for example, robot manipulator trajectories [2]. The numerical equivalent does not appear to have been much studied, except in special cases.

Suppose we have a finite set of points in $\mathbb{R}^n$ and we wish to fit a manifold to them. There are two special cases of note, the first is the case of fitting a 1-manifold, a curve, to a set of points in $\mathbb{R}^2$. We may be concerned with choosing a curve that passes through all the points, as is commonly done with low order splines, or we may suppose there is some random variation and we wish to satisfy some goodness of fit criterion such as least squares, together with some simplicity criterion such as a bound on the degree of a polynomial. The second special case is where we have a set of points in $\mathbb{R}^n$ and we wish to find the hyperplane of dimension k which contains most of the variation of the data. The well-known Karhunen–Loève expansion accomplishes this last [3], and there are algorithms for the one-dimensional curve-fitting case, and also for the two-dimensional case. In all cases, the manifold is constrained globally by all the data points.

The problem of finding the 'best fit manifold', and assessing the goodness of fit, is then a generalisation (to the non-linear and to higher dimensions) of a standard problem. It is reasonable to want to (a) smooth data, (b) interpolate data and (c) reduce the effective dimension of data, without the constraints on dimension on the one hand or linearity on the other. It is necessary to tackle this problem locally, unlike the case of least squares best fitting of graphs of functions, since the global topology is unknown. We cannot therefore expect to emerge with one function, but must finish with at best a set of local 'charts', maps from the unit k-cube into $\mathbb{R}^n$. For obvious reasons, we shall call this process the *numerical parametrisation* of the underlying manifold from which the points are drawn.

The direct approach is to choose some 'small' neighbourhood and try to fit a graph to it. This is essentially what is done in some existing routines which use higher-dimensional splines. Another possibility, which we have not seen discussed in the literature, would be to find local co-ordinate charts for the manifold by trying to minimise a least squares criterion and also a 'stiffness' or curvature measure. An algorithm which seems to have some affinities with this last approach but is still different, is that of Kohonen [4]. In this paper we shall consider the Kohonen algorithm; we shall note its features and show that, when it converges, it can accomplish such a numerical parametrisation. The algorithm implements a philosophy of turning the data into a space of attractors, representing an essentially traditional approach to numerical solutions going back to Newton's method, but the details are new.

1. The Kohonen algorithm: overview

The Kohonen algorithm [4, 5], takes a k-dimensional grid of points in $\mathbb{R}^n$ and moves them towards the points on the set as they are selected in some sequence. It allows the points to be selected non-uniformly from the manifold which is assumed to have dimension k. We shall refer to the points selected from the (putative) manifold as *attracting points* or *attractors*, and the points of the k-dimensional grid as *grid points*. Then we may think of the points of the manifold as attracting, or calling, the grid points from some initial position which may be randomly selected and pulling the grid points until they are on the manifold. The Kohonen algorithm gives a 'law of motion' which tells how to move the grid points in response to the choice of a particular attractor. It is claimed that this law of motion leads, after some presentation sequence of attractors, to a convergent state in which the grid is wrapped around the manifold, thus giving what we have called a 'numerical parametrisation' of the manifold. There is no procedure for testing goodness of fit, something with which we shall be concerned subsequently. The procedure for moving the grid towards the attracting points on the manifold is as follows: the grid point closest to the attracting point is found. A grid neighbourhood of points is chosen; this grid neighbourhood is defined by the topology of the grid, usually a rectangular array, and a size parameter, W. Thus if the grid dimension were 2 and the size parameter were 3 and the selected point was specified as $G(7,5)$, then all the points $G(x,y)$ with $4 \leqslant x \leqslant 10$ and $2 \leqslant y \leqslant 8$ would be in the grid neighbourhood (regardless of their location in $\mathbb{R}^n$). Each member inside the grid neighbourhood is moved along the straight line joining the grid point to the attracting point; the fraction moved is determined by a move parameter A. Both A and W are reduced, according to some schedule, from some initial value to zero at termination time T, the time at which all the points have been presented to the grid, possibly with repetitions.

The application described in [5] and based on the so-called tonotopic map of [4], is to the case of speech data. Speech is sampled and the spectrum measured, either by giving energy levels in a bank of filters or by LPC coefficients, or by some other procedure. Thus an utterance becomes a trajectory (discretised) in $\mathbb{R}^n$, for some n (the number of filters or LPC coefficients). It is asserted that there is reason to suppose that no matter what n is, and no matter what the set of measurements made (provided only that they resolve the data), the trajectories will lie on some two-dimensional surface embedded in $\mathbb{R}^n$. Changing from LPC coefficients to filter bank values will change the embedding but not the dimension of the 'speech space' which contains all the trajectories. Thus by presenting time samples of speech to a 2-grid and applying the Kohonen algorithm, the speech space itself will be found. Moreover, the fact that the points will be selected from the speech space non-uniformly will not affect convergence, it will indeed assign grid points in greatest numbers to the regions of highest density of the attracting points. This allows the grid points to be selected as centres for so-called *vector quantisation*, meaning that the trajectory can now be discretised into a symbol string, where the symbols are simply the grid point names, and we assign a region of a trajectory to the nearest grid point. The result of the Kohonen algorithm, it is claimed, is to have highest resolution in the regions of the space where the highest density of attracting points occurs. This, it is further claimed, is desirable.

The arguments given in [5] for believing that the speech trajectories lie on a manifold and that the dimension of the manifold is 2 are of uncertain force. It would clearly be desirable to have some experimental confirmation of the rather striking theory that the dimension has to be 2.

It is also clearly desirable that the terminal converged state which assigns the grid points to the manifold, and which hence allows us to interpolate points and to smooth them, should constitute a discrete approximation to a locally 1–1 map. It would be acceptable if the grid were wrapped around the manifold with some overlap, as for example when the points are on a circle or a sphere or a torus, but the mapping should not contain folds or wrinkles.

Kohonen, in [4], claims that the grid points 'tend to be ordered according to their mutual similarity' and that the asymptotic local point density of the grid points is a continuous monotone function of the probability density function of the attracting points (Proposition 5.1). The former phrase appears to mean that when the attracting points are selected from a manifold, the terminal converged state assigns the grid points to points of the manifold in such a way that they give a discrete approximation to a 1–1 map. In the case when the grid and the manifold are one-dimensional, both the grid and any set of attracting points from the manifold can be ordered, and if the grid points actually converge to attracting points (and not to some intermediate points), the meaning is simply that the map defined at convergence should be monotone. In this one-dimensional case, if the grid points are allowed to converge to the manifold (but not necessarily to the attracting points), then the simplest way to express the 'ordering according to mutual similarity' is to point out that a map defined on a totally ordered set of $r+1$ elements into $\mathbb{R}^n$ determines a unique piecewise affine map from the interval $[0, r]$ in $\mathbb{R}$ which takes the integer j to the $(j+1)$th point of the grid after convergence. Then we require this map to be monotone. For higher dimensions, not treated in [4], there are some problems; there are also problems when the manifold is compact. What the proof in [4] actually does is to show that for an attracting manifold consisting of a line segment, the algorithm operating on a linear array of points converges to a monotone map. Thus it is not valid for circles, for example, and the generalisation is not immediately apparent. For this reason a proof of the general result, that we can obtain numerical parametrisations of any manifold would have some interest. Experiments running simulations are also of interest. The main result of this paper is to give the generalisation for any dimension: we prove that the map which represents convergence and puts a k-cube onto the manifold in $\mathbb{R}^n$ is an *immersion*, that is to say, it is locally 1–1.

2. Features of the Kohonen algorithm

As a numerical parametrisation algorithm, there are some features which are marks of its ancestry. In particular, the serial presentation of the points taken from the manifold to the grid makes for an algorithm which does not need to store a large quantity of data. Since it is abstractly a model of a certain kind of learning, this is natural enough. Also, it is inherently parallel, with only limited non-local interactions. This makes it possible to run the algorithm on a network with an architecture which does not seem to have been much investigated, one with a small blackboard but otherwise similar to a systolic array. Finally, an aspect of the algorithm which has been mentioned above is that it will preserve the density of the frequency distribution of the points, or at least give a numerical approximation to it.

We shall not discuss here the origins of the algorithm in neural modelling nor the extent to which it is plausible as such a model.

The algorithm, as has been stated above, requires for its complete specification, a schedule of how to decrease the move size and the neighbourhood size. (This is strongly reminiscent of the annealaing schedule used in the simulated annealing algorithm for non-convex optimisation [6].) As was discussed above, while experiments suggest that for any low-dimensional manifold which satisfies certasin broad conditions there are schedules which work, they also suggest that for any schedule there are manifolds for which the schedule will fail to work. Indeed, as stated, the Kohonen algorithm need not have the grid converged to the manifold at all.

Experimentally, one finds that if the manifold is embedded as a non-convex set, one gets a kind of 'convex approximation' to the manifold. The point is illustrated in [4]. It is not clear whether this is a significant defect, nor on how it is changed as the number of grid points increases, nor how it depends on the shrinking schedule. For small numbers of points and low dimensions it is often the result which agrees with naive intuitions.

Kohonen gives no way of measuring the 'goodness of fit' of the terminal converged state of the grid to the manifold. In [4] some qualitative observations are made on the degree of folding which ensues in the case where one tries to fit a grid of dimension 1 or 2 to a manifold of dimension which is greater than that of the grid, but no measurement of this is proposed. Consequently, using Kohonen's algorithm to confirm the dimension of a set is impractical; one may come away with one's convictions intact, but one has no quantitative justification for so doing. There is no independent way of deciding on the dimension of the manifold from which the set has been taken, a notoriously difficult problem. See [7].

Finally, there is a serious absence of theory. The convergence theorem which Kohonen gives applies only to the one-dimensional case of a grid along a line and the attractors selected from an interval, and in any case is not argued very rigorously. The existence of very badly embedded manifolds is not considered. Since there are some monsters which are well known to topologists, the generalisation is far from trivial and the formal machinery for doing so is not present in [4] or [5]. We conclude from these observations that the algorithm is not yet in a fully developed form, but that it is important enough to merit further development. This is quite aside from any value it may have as a neural model, which suggests other, quite different, reasons for developing it.

3. An immersion theorem

We shall assume in what follows that we know k, the dimension of the manifold. To avoid some of the technical difficulties which arise from premature discretisation, we shall consider the case where the grid is replaced by a unit k-cube. Then a proof that the attraction of a sequence of points from the manifold induced a map from the k-cube to the manifold would certainly leave us with the result for the discrete approximation of a grid. It would also relieve us of the need to consider alternative grid geometries and allow us to use the machinery of analysis. The immediate problem then becomes to formulate Kohonen's law of attraction in such a way that it reduces to the standard case when we restrict back to a grid, because if we simply choose the points in a neighbourhood and move them alone, our mapping immediately becomes discontinuous. It is possible to give a rule for moving the points of the cube which is continuous and yield precisely the Kohonen rule under restriction to a grid, by putting a 'collar' around the neighbourhood; inside the neighbourhood points move according to the Kohonen rule, outside the collared neighbourhood they do not move at all, as with the Kohonen rule, and inside the collar, they move by an intermediate amount. The collar can be made arbitrarily narrow, and the map can be continuous and indeed smooth. This again yields the Kohonen rule when restricted to a grid. Our preference for smooth functions is based largely on a desire to be able to use traditional analytic machinery in describing the system, and partly on a prejudice that nature seems to be that way.

The smoothing of the Kohonen function can be expressed much more simply than by such artifacts as 'collars'. We merely need to require that points of the cube which are sufficiently close should be moved approximately the same distance, and that points close to an attracting point should be moved more than those further away. The point do dwell on is that there are two different senses of 'close', one is 'close in $\mathbb{R}^n$', and the other is 'close in the k-cube'. It is the function of the Kohonen rule to ensure that these finally come to be the same.

For our formal treatment, we observe that each attracting point operates in the same way as any other up to a translation, so we may consider the rule of motion to be specified completely by saying what happens when the attractor is at the origin, and adding the obvious translation to handle the case when it is not. We also note that the attractor takes one smooth mapping of a k-cube into $\mathbb{R}^n$ and turns it into another. Thus the sequence of manifold points determines a sequence of transformations of the space of smooth mappings from $\mathbb{I}^k$, the unit k-cube, to $\mathbb{R}^n$, usually written $\mathscr{C}^\infty(\mathbb{I}^k, \mathbb{R}^n)$. We shall call this space $S(k,n)$. Then the origin in

$\mathbb{R}^n$, regarded as a conveniently located manifold point, acts as a map

$$O: S(k,n) \to S(k,n).$$

We now define a *Kohonen Operator* as such a transformation having the properties:

(1) $\forall f \in S(k,n)\ \forall x \in \mathbb{I}^k\ \exists \alpha \in [-1,1]\ O(f)(x) = \alpha f(x)$, where $\alpha \neq 0$, and α is a smooth function of x for every f.

(2) $\forall f \in S(k,n)\ \forall u, v \in \mathbb{I}^k$

$$|O(f)(u)| < |O(f)(v)| \Leftrightarrow \forall x \in \mathbb{I}^k$$
$$(\forall y \in \mathbb{I}^k\ |f(x)| \leqslant |f(y)| \Rightarrow |x-u| \leqslant |x-v|).$$

The first of these requires that O should move curves no farther from the attracting point, and points move in directly along the radius, while the second says that if a point $f(u)$ finishes up closer to the attractor than a point $f(v)$, then it must be that the closest point of the curve to the origin (in the sense of close in $\mathbb{R}^n$) is closer to u than to v (in the sense of close in $\mathbb{I}^k$). The reasons for writing this formally are plain.

It is easy to see that the Kohonen rule satisfies these conditions when we restrict them to any grid.

Suppose now that a sequence of points, $a_1, a_2, a_3, \ldots$ is taken from M, a manifold in $\mathbb{R}^n$, and that $g: \mathbb{Z}^+ \to \mathbb{R}^+$ is a monotone function satisfying:

$$\lim_{r\to\infty} g(r) = 0.$$

Define a *Kohonen Sequence* as a sequence of operators, T_r by

$$\forall f \in S(k,n)\ \forall x \in \mathbb{I}^k,$$
$$T_r(f)(x) = a_r + g(r)\ O(f - A_r)(x),$$

(where A_r is the constant element of $S(k,n)$ sending everything to a_r) which satisfies the condition:

$$\forall f \in S(k,n)\ \forall x \in \mathbb{I}^k\ \forall \varepsilon \in \mathbb{R}^+$$
$$\exists N \in \mathbb{Z}^+\ \forall r \in \mathbb{Z}^+,$$
$$r > N \Rightarrow (|f(x) - T_r(f)(x)| > 0 \Leftrightarrow$$
$$\exists y \in \mathbb{I}^k\ |x-y| < \varepsilon\ \&$$
$$\forall z \in \mathbb{I}^k\ |f(y) - a_r| < |f(z) - a_r|).$$

This, when translated into English, is the neighbourhood shrinking procedure. It ensures that outside some ε-neighbourhood of the closest point, the operation does nothing, and ensures that ε gets smaller as we go down the sequence.

It is easy to see that the Kohonen rule is a finite approximation to this case, alternatively that we can approximate the finite case arbitrarily closely by a Kohonen sequence of operators. The only problem is in the choice of g, where a finite approximation might go to zero after finitely many terms.

Proposition. *Suppose a sequence of points, $a_1, a_2, a_3, \ldots$ is given and that these are dense and uniformly distributed in a smooth k-manifold M, $k > 0$, embedded in $\mathbb{R}^n$. Suppose that the Kohonen sequence of operations converges for some initial state f_0, an embedding, to some non-constant final state f, both smooth maps from $\mathbb{I}^k$ to M. Then f is, with probability* 1, *an immersion from $\mathbb{I}^k$ to M, i.e. it is locally* 1–1.

Proof. We have to show that the derivate of f has rank k everywhere. It is clear that f_0 has full rank and that O preserves rank (and of course smoothness), so we need examine only the limiting case.

We take $r \in \mathbb{N}$ and $\varepsilon \in \mathbb{R}^+$ and two points b and c in $\mathbb{I}^k$ which are both mapped to within ε of the manifold such that both f_r and the manifold may be taken to be affine in some ball.

Consider now the effect on the line segment $[f_r(b), f_r(c)]$, approximately the image of the line segment $[b,c]$ in $\mathbb{I}^k$, of an attracting point a on the manifold. Each point moves radially if at all towards a. Let the angle at a be called α and let the distance $|a - f_r(b)|$ be called u, and the distance $|a - f_r(c)|$ be called v. Suppose if w is the original distance, then the size of the jump is written $j(w)$. Then the condition that the separation of the points $f_r(b), f_r(c)$ be increased by the action of the attraction towards a is easily seen to be:

$$2\cos(\alpha) > \frac{2u\,j(u) + 2v\,j(v) - (j(u))^2 - (j(v))^2}{u\,j(v) + v\,j(u) - j(u)\,j(v)}.$$

If, without loss of generality, $j(v)$ is very small compared with $j(u)$, either by virtue of the shrinking of neighbourhoods or by virtue of the details of the dynamic, this reduces to

$$\cos(\alpha) > \frac{2u - j(u)}{2v}.$$

We note that there is an advantage then in having $j(u)$ tend to $2u$ as u tends to zero, when the point moves almost as far to the opposite side of the attractor, as this maximises the angle for which the separation is increased. In this case, we have that α is at its maximum of a right angle. We take then a hypersphere with centre the midpoint of $f_r(b)$ and $f_r(c)$, and radius half the separation. Then any point of the manifold which lies inside the hypersphere contracts the line joining the two points when it becomes an attractor, and all other points will not. The case where we have $j(u)$ samller than $2u$ is slightly more complicated but in essence the same and we shall not consider it here. $j(u) > 2u$ is evidently imcompatible with convergence.

Let B denote the ball of radius δ with centre $f_r(b)$ and C that with centre $f_r(c)$. Let Q denote the ball centred at the midpoint which has the points $f_r(c), f_r(b)$ on its boundary. Then the set of points which lie on the manifold and inside B or C and also Q are those attractors which will shrink the line segment, and those in B or C but not Q will expand it. Since the distribution is uniform, by hypothesis, the probability of these events is given by the relative measures of the two regions. It is easy to see that these tend to a ratio of contracting to expanding which is less than or equal to one. This holds for all subsequent iterations, so the probability of the line segment contracting to a point subsequently is zero. This guarantees that f is locally 1-1. The nonsingularity of the derivative is marginally stronger, and follows from the observation that the ratio of the directional derivatives of consecutive operations on f tends to a non-zero constant.

Remark. There is a probability density function which can be obtained from the immersion f by taking the k-measure of a neighbourhood of a point in $\mathbb{I}^k$, the k-measure of its image under f and taking the 'compression ratio'. It is easy to see that this is simply the absolute value of the Jacobian determinant at each point in the case where $k = n$, and is more generally given via the Kronecker tensor. It is also clear from the above argument that when the pdf of the points on the manifold is uniform we will get the induced pdf of f tending to a non-zero constant, the value of which depends on the law of attraction and the shrinking law.

It then follows, via approximations which are locally constant, that if the points on the manifold come from a smooth density function then the pdf derived from f is a monotone function of it, provided only that the laws are invariant under the Euclidean Group on $\mathbb{R}^n$.

Remark. Note that there is nothing to prevent us wrapping $\mathbb{I}^k$ around the manifold in such a way that the final map is not 1-1. We have not, of course, shown that convergence must occur, merely that if it does we must have an immersion, and we have supposed that the dimension is known. It is easy to see that when the dimension of the manifold is less than that of the disk being attracted, that the result above still holds; what happens is that convergence cannot now occur. In the case where the dimension of the disk is *less* than that of the manifold, we can get both convergence and the immersion.

Remark. The application to the case of a finite set of points, is fraught with difficulties. It is plain that a finite set of points could have come from a lot of different manifolds, and that if there is a large amount of variation in the pdf for the points, we may find our manifold can be approximated by something topologically very different. It is hardly surprising that topological invariants do not readily survive numerical approximation, what is surprising, as Dr. Johnson might have remarked, is that they can survive at all: yet the human eye can and does pick out non-simply-connected manifolds from point sets.

4. Summary and conclusions

We have discussed the Kohonen algorithm for numerical parametrisation of manifolds. We have shown that in the idealised case of an infinite set of attracting points, dense in a k-manifold, the result of an abstract Kohonen operation on any reasonable initial state is to produce an immersion onto the manifold of a k-cube. By starting from several different initial locations we can therefore hope to cover the manifold with k-disks. This is done in a way which is neither wholly local nor wholly global.

Associated issues are (1) the degree of goodness of fit and measures of it, (2) degrees of goodness of agreement on any overlap and (3) the question of whether the dimension of the manifold really is the same as the dimension of the k-disks fitted. These issues are evidently related, and will be addressed elsewhere. Also pertinent is the stability of finite approximations, in which probability zero events are an everday occurrence.

References

[1] Arnold, V.I. (1978). *Mathematical Methods of Classical Mechanics.* Springer, Berlin.

[2] Noakes, J.L. (1987). Trajectory generation and obstacle avoidance. *IMA J. Math. Control & Information* 4, 293-300.

[3] Karhunen, K. (1947). *Ann. Acad. Sci. Fenn., Ser. A*, 37 (1).

[4] Kohonen, T. (1988). *Self-Organization and Associative Memory.* Springer Series in Info. Sci 8. Springer, Berlin.

[5] Tattersal, G.D., P.W. Linford and R. Lingaard (1988). Neural arrays for speech recognition. *Br. Telecom J.* 66 (2).

[6] Lundy, M. and A. Mees (1986). Convergence of an annealing algorithm. *Math. Programming* 34, 111-124.

[7] Farmer, J.D., E. Ott and J.A. Yorke (1983). The dimension of chaotic attractors. *Physica* 7D, 153-180.