PERGAMON

Neural Networks 15 (2002) 1185–1196

Neural Networks

www.elsevier.com/locate/neunet

Straight monotonic embedding of data sets in Euclidean spaces

Pierre Courrieu*

Laboratoire de Psychologie Cognitive, CNRS-UMR 6146, Université de Provence, 29 avenue Robert Schuman, 13621 Aix-en-Provence Cedex 1, France

Received 12 July 2001; accepted 15 May 2002

Abstract

This paper presents a fast incremental algorithm for embedding data sets belonging to various topological spaces in Euclidean spaces. This is useful for networks whose input consists of non-Euclidean (possibly non-numerical) data, for the on-line computation of spatial maps in autonomous agent navigation problems, and for building internal representations from empirical similarity data. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Non-Euclidean data sets; Monotonic embedding; Multidimensional scaling; Singular Cholesky factor; Spatial maps

1. Introduction

The main concern of this paper is the spatial representation in Euclidean spaces of data sets belonging to metric or non-metric topological spaces. In a previous study (Courrieu, 2001), the particular case of spaces of clusters was treated, a cluster being defined as a finite unordered set of points. Two methods were proposed for encoding clusters in an Euclidean code space, while preserving various topological properties of data spaces. The present work concerns other types of data spaces, and is related to a well known field in data analysis, i.e. 'multidimensional scaling' (Cox & Cox, 1994; Kruskal, 1964a,b; Shepard, 1962a,b; Young & Torgerson, 1967). Usual multidimensional scaling algorithms search for a set of points of R^n whose mutual distances (for a given metric) approximate a given 'dissimilarity' measure between the corresponding elements of the data space, while the dissimilarity measure is not necessarily a metric. Most of these algorithms attempt to minimize an error criterion using a gradient descent type procedure, which gives rise to quite slow computation and frequently causes the search process to be trapped into local minima. A recent variant of gradient descent allows for improving the performance of these methods (Demartines & Hérault, 1997). From a mathematical point of view, multidimensional scaling is closely related to the well known 'isometric embedding' problem (Blumenthal, 1936; Fréchet, 1910; Micchelli, 1986; Schoenberg, 1937, 1938). This connection will be widely exploited in this paper.

The Neural Networks community is concerned with this problematic in several ways. First, neural computation research has developped powerful methods for approximating continuous mappings on compact subsets of Euclidean spaces, from finite sets of data points. The Euclidean nature of the support space is particularly clear for Radial Basis Function Networks and Radial Spline systems (Girosi & Poggio, 1990; Poggio & Girosi, 1990), since Radial Basis Functions are functions of an Euclidean distance on the input space. Moreover, fundamental properties of networks, such as their approximation and regularization capabilities, critically depend on the hypothesis concerning the support space. However, as concerns practical applications, one can note that available data spaces are frequently not Euclidean, and even that they are frequently not real vectorial spaces. This results in difficulties for engineers faced with implementing artificial neural network applications, since they must transform given data spaces into real vectorial spaces empirically, and in general without any guarantee concerning the relevance of this operation. Certain numerical data sets can be artificially considered as Euclidean spaces, however, using an Euclidean metric on such data sets frequently leads to very disappointing results. A well known example of this is the case of numerical time series: one can always consider time series as vectors and compute an Euclidean distance between two vectors, however, this is rarely relevant. Dissimilarity measures provided by Dynamic Programming methods ('elastic template matching') are in general much more appropriate, however, the

* Corresponding author. Tel.: +33-4-42-95-37-28; fax: +33-4-42-20-59-05.

E-mail address: courrieu@up.univ-mrs.fr (P. Courrieu).

PII: S0893-6080(02)00091-6

resulting space is not Euclidean, a priori, and in fact it is not necessarily metric if the measure used is not a true metric (Okochi & Sakai, 1982; Vinstuk, 1968). There are also data spaces which are not numerical, such as spaces of symbol strings, for example. However, there are well-known methods for defining dissimilarity measures on symbol string spaces, these measures being called 'edition distances' (Lowrance & Wagner, 1975; Wagner & Fischer, 1974). One can find many other examples of data spaces which are not Euclidean, or even not numerical, but on which one can define a dissimilarity measure which has at least some properies of a metric. In general, triangle inequality is the most difficult property to obtain. Now, assume that one can define a simple continuous strictly increasing transform of a dissimilarity measure, and the useful part of the data space with this monotonically transformed dissimilarity measure is a metric set isometrically embeddable in an Euclidean space (this is a 'monotonic embedding'). It is clear that this would help to solve a number of practical problems of data encoding for neural network applications. Another field of interest concerns autonomous agents (alive or artificial) which must locate objects and themselves in a given environment, using approximative evaluations of distances for building a 'spatial cognitive map'. The interest of multidimensional scaling in this navigational context is obvious, however, this requires on line computation, that is simple and fast algorithms. The problem is clear for robotic applications, while the principle of an on line computation of spatial cognitive maps also seems relevant from a neurobiological point of view if one consider the behavior of hippocampal place cells (Cressant, Muller, & Poucet, 1997, 1999; Muller, 1996; O'Keefe & Nadel, 1978; Poucet, Save, & Lenck-Santini, 2000). Finally, multidimensional scaling methods were widely used in the area of psychological science for approximating the so-called 'psychological spaces' from empirical subjective similarity data (Nosofsky, 1992; Shepard, 1987). This can be viewed as a generalization of spatial cognitive maps to more abstract spaces.

In this paper, we present a particular approach of multidimensional scaling which, in a sense, is less general than usual multidimensional scaling methods, since it does not allow for embedding data sets in any real metric space. However, the method developed here is particularly appropriate to neural network applications since it allows for an exact monotonic embedding of data sets in Euclidean spaces, using a straight, fast and incremental algorithm. The incremental character of the algorithm implies that one can embed any new item without recomputing or modifying the embedding of previously embedded items. This is absolutely necessary for embedding the current input of a network, or of a navigation system, without modifying the embedding of the learning set (and hence the network itself), or of the landmarks. A straight fast procedure is required for on-line computation, while the exactitude of the embedding is not the most relevant characteristic in the case of noisy data, however, this is a way of obtaining fast computation. The present approach lies on mathematical fundations which extend the earlier mentioned classical results concerning the isometric embedding problem. Surprisingly, these classical results were much more widely used in the study of positive definite functions than in the field of multidimensional scaling, despite the fact that multidimensional scaling is basically an isometric (or monotonic) embedding problem.

2. Conventions and background

2.1. Notations

The identity matrix is denoted I. The transposed row vector of a collumn vector v is denoted v', while the transposed matrix of a matrix Q is denoted Q'. For specifying the content and size of a function vector or a function matrix, it will be convenient to use notations of the form:

$$v = [v_i]_{i=1,\dots,n},$$

where v_i can be an expression,

$$D = [d_{ij}]_{i,j=0,\dots,n},$$

where d_{ij} can be an expression.

For example, let A, B and C be three $(n+1)\times(n+1)$ matrices, then

$$A + B - C = [a_{ij} + b_{ij} - c_{ij}]_{i,j=0,\dots,n}.$$

2.2. Usual definitions

2.2.1. Topological and metric spaces

A 'metric' or 'distance' associated to a set S, is a real valued function d on $S \times S$ such that, for any $a, b, c \in S$, one has the four following properties: (1) $d(a,b) \geq 0$ and $d(a,a) = 0$, (2) $d(a,b) = d(b,a)$, (3) $d(a,b) \leq d(a,c) + d(b,c)$, (4) if $a \neq b$ then $d(a,b) > 0$. The triangle inequality (3) can also be writen as $d(a,b) \geq |d(a,c) - d(b,c)|$.

A function which satisfies only requirements (1)–(3) is called a 'semi-metric'. A function d' which satisfies only requirements (1) and (2) is sufficient for inducing a topology on S. An open (resp. closed) 'ball' of center $x \in S$, and of radius $r \geq 0$, is the set of points $\{y \in S; d'(x,y) < r(\text{resp.} \leq r)\}$. The set of all balls is a neighbourhood system and then, (S, d') is a topological space. Now, if d is a true metric, then (S, d) is called a 'metric space', or 'metric set'. If S is a finite set of $n+1$ elements, then one can associate to $(S,d)a$ 'distance matrix' of the form $D = [d_{ij}]_{i,j=0,\dots,n}$, and one can do the same even if d is not a true metric (while avoiding the term 'distance matrix' in this case).

2.2.2. Minkowskian metrics

A Minkowskian metric is a metric associated to R^n of the form

$$M_q(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^q\right)^{1/q}, \qquad q \geq 1.$$

The most usual Minkowskian metrics are the 'city-block' metric M_1, the Euclidean metric M_2, and the 'dominance' metric $M_\infty(X,Y) = \max_i |x_i - y_i|$. While all Minkowskian metrics are invariant by changing the origin or the sign of coordinates, the Euclidean metric is the only one which is also invariant by any orthogonal transform of the coordinates (e.g. rotation).

2.2.3. Embeddings

A metric space (S', d') is said to be 'isometrically embeddable' in a metric space (S, d) if there is a mapping f from S' to S such that, for any $a, b \in S'$, one has $d(f(a), f(b)) = d'(a, b)$.

A topological space (S', d') will be said to be 'monotonically embeddable' in a metric space (S, d) iff there is a continuous strictly increasing function g on $[0, \infty)$ such that $(S', g(d'))$ is a metric space isometrically embeddable in (S, d).

2.2.4. Properties of symmetric matrices

The 'Rayleigh's ratio' of a symmetric matrix B by a non-zero vector v is the ratio $R_B(v) = v'Bv/(v'v)$. The lowest eigenvalue of B is equal to $\inf_v R_B(v)$, while the greatest eigenvalue of B is equal to $\sup_v R_B(v)$.

A real symmetric matrix B of order $n \times n$ is said to be 'positive definite' iff for any vector $v \in R^n$, one has: $v'Bv \geq 0$. It is equivalent to say that none of the eigenvalues of B is negative.

A real symmetric matrix B of order $n \times n$ is said to be 'strictly positive definite' iff for any non-zero vector $v \in R^n$, one has: $v'Bv > 0$. It is equivalent to say that all eigenvalues of B are strictly positive.

A real symmetric matrix B of order $n \times n$ is said to be 'almost negative definite' (Donoghue, 1974; Micchelli, 1986) iff for any vector $v \in R^n$ such that $\sum_{i=1}^{n} v_i = 0$, one has: $v'Bv \leq 0$.

A real symmetric matrix B of order $n \times n$ will be said to be 'almost strictly negative definite' iff for any non-zero vector $v \in R^n$ such that $\sum_{i=1}^{n} v_i = 0$, one has: $v'Bv < 0$.

2.3. Useful theorems

Bessel–Parseval's inequality. Let H be a pre-Hilbert space, and $(e_i)_{i=1,\ldots,m}$ be an orthonormal family of H. Then, for any vector $x \in H$, the family $(|\langle x, e_i\rangle|^2)_{i=1,\ldots,m}$ is summable and $\sum_{i=1}^{m} |\langle x, e_i\rangle|^2 \leq \|x\|^2$, where $\langle\cdot,\cdot\rangle$ denotes the scalar product of H, and $\|\cdot\|$ is the corresponding norm. The orthonormal family $(e_i)_{i=1,\ldots,m}$ is complete for x iff $\sum_{i=1}^{m} |\langle x, e_i\rangle|^2 = \|x\|^2$ (Parseval's identity).

Theorem 1 (Gerschgorin). *Let B be a square $n \times n$ matrix. Then all eigenvalues of B belong to the union of the intervals defined by*

$$|x - b_{ii}| \leq \sum_{j \neq i} |b_{ij}|, \qquad 1 \leq i \leq n.$$

Lemma 1 (Fréchet, 1910; Schoenberg, 1938). Any finite metric set (S, d) of $n+1$ points may be embedded isometrically in (R^n, M_∞).Solution:

$$[x_{ij}]_{i=1,\ldots,n,\, j=0,\ldots,n} = [d_{ij}]_{i=1,\ldots,n,\, j=0,\ldots,n}.$$

Theorem 2 (Schoenberg, 1937, 1938; Micchelli, 1986). *A necessary and sufficient condition for the isometric embeddability of a finite metric set (S, d) of $n+1$ elements in an Euclidean space is that one of the two following statements be true:*

(i) *the matrix $[d_{0i}^2 + d_{0j}^2 - d_{ij}^2]_{i,j=1,\ldots,n}$ is positive definite,*
(ii) *the matrix $[d_{ij}^2]_{i,j=0,\ldots,n}$ is almost negative definite.*

Moreover one has (i) $\Leftrightarrow$ (ii).

Corollary 1. *The two following statements are equivalent:*

(iii) *the matrix $[d_{0i}^2 + d_{0j}^2 - d_{ij}^2]_{i,j=1,\ldots,n}$ is strictly positive definite,*
(iv) *the matrix $[d_{ij}^2]_{i,j=0,\ldots,n}$ is almost strictly negative definite.*

Proof. It suffices to replace '$\geq$' and '$\leq$' by '$>$' and '$<$', respectively, in the inequalities stated by Schoenberg (1938, pp. 525–526). □

Corollary 2. *If the finite metric set (S,d) is isometrically embeddable in an Euclidean space, then the required minimum dimension of this space is equal to the rank of the matrix $[d_{0i}^2 + d_{0j}^2 - d_{ij}^2]_{i,j=1,\ldots,n}$.*

Proof. Considering d as an Euclidean distance and taking the element of S with the index 0 as the origin (that is $X_0 = 0$), one has:

$$d_{0i}^2 + d_{0j}^2 - d_{ij}^2 = \|X_i\|^2 + \|X_j\|^2 - \|X_i - X_j\|^2 = 2X_i \cdot X_j,$$

that is

$$[d_{0i}^2 + d_{0j}^2 - d_{ij}^2]_{i,j=1,\ldots,n} = 2X'X.$$

Since $X'X$ is symmetric, one has $X'X = Q\Delta Q'$, where Q is the orthogonal matrix of eigenvectors, and Δ is the diagonal matrix of eigenvalues of $X'X$. Since these eigenvalues are not negative (after Theorem 2), on can take $X = \Delta^{1/2}Q'$ as an embedding solution. Then the number of dimensions is equal to the number of non-zero eigenvalues, and this completes the proof. □

3. Monotonic embeddability in metric spaces

In this section, one examines general conditions of monotonic embeddability of data sets in a real metric space, whenever it is not required that this space be Euclidean.

Lemma 2. *Let S be any finite set comprised of $n+1$ distinct elements, and let μ be a real valued function on $S \times S$ such that, for any $a, b \in S$, one has $\mu(a,a) = 0$, and if $a \neq b$ then $\mu(a,b) = \mu(b,a) > 0$, $\mu(a,b) < \infty$. Then there is a real $\gamma(S) > 0$ such that for any stricly positive real $p \leq \gamma(S)$, the space (S, μ^P) is a metric set isometrically embeddable in (R^n, M_∞).*

Proof. Consider the matrix $[\mu_{ij}]_{i,j=0,\dots,n}$ associated to the (non-metric) space (S, μ), let $\inf \mu = \min_{i \neq j} \mu_{ij} > 0$, and $\sup \mu = \max_{i \neq j} \mu_{ij}$. Then, for a real $p > 0$, the space (S, μ^p) is a metric set provided that for any i, j, k, one has the triangle inequality $\mu_{ij}^p \leq \mu_{ik}^p + \mu_{kj}^p$, which is guaranteed if p is such that $(\sup \mu)^p \leq 2(\inf \mu)^p$, that is $p \leq \ln(2)/\ln(\sup \mu/\inf \mu) \leq \gamma(S)$. Then the isometric embeddability of (S, μ^p) in (R^n, M_∞) results from Fréchet–Schoenberg's lemma (i.e. Lemma 1), which proves Lemma 2. □

Given Lemma 1, it is quite natural to ask whether any finite metric set is isometrically embeddable in an Euclidean space. The answer to this question is easy, since one can find many decisive examples.

Lemma 3. *There are finite metric sets which are not isometrically embeddable in an Euclidean space.*

Proof. Consider finite metric sets of $n+1$ elements, $n \geq 4$, whose distance matrices have the following structure.
Chose a real vector $z \in R^n$, $n \geq 4$, such that

$$\|z\| = 1, \qquad |z_i| \leq 1/2, \qquad 1 \leq i \leq n, \qquad (z_i - z_j)^2 \leq 3/4,$$

$$1 \leq i \leq j \leq n.$$

Define the distances by

$$d_{ii} = 0, \qquad 0 \leq i \leq n, \qquad d_{0i} = d_{i0} = \sqrt{1/2 - z_i^2},$$

$$1 \leq i \leq n, \qquad d_{ij} = d_{ji} = \sqrt{1 - (z_i - z_j)^2},$$

$$1 \leq i < j \leq n.$$

Using this definition, one obtains

$$1/2 \leq d_{ij} = d_{ji} \leq 1, \qquad 0 \leq i < j \leq n,$$

and then, for any i, j, k, one has

$$d_{ij} \leq d_{ik} + d_{kj},$$

which is triangle inequality. Since d obviously has the other properties of a distance, it is clear that this is actually a metric for the considered set.

Now, one has also

$$2d_{0i}^2 = 1 - 2z_i^2,$$

$$d_{0i}^2 + d_{0j}^2 - d_{ij}^2 = -z_i^2 - z_j^2 + (z_i - z_j)^2 = -2z_i z_j,$$

that is

$$[d_{0i}^2 + d_{0j}^2 - d_{ij}^2]_{i,j=1,\dots,n} = I - 2zz',$$

where the right member is, by definition, a Householder's matrix. Since one has

$$z'(I - 2zz')z = -1,$$

it is clear that the matrix $[d_{0i}^2 + d_{0j}^2 - d_{ij}^2]_{i,j=1,\dots,n}$ is not positive definite, and then, following Theorem 2, the finite metric set whose distance matrix is $[d_{ij}]_{i,j=0,\dots,n}$ is not isometrically embeddable in an Euclidean space, which proves Lemma 3. □

The above proof concerns metric sets of at least five points. Metric sets of one ore two points are trivially isometrically embeddable in an Euclidean space. Metric sets of three points are also embeddable since the triangle inequality always allows for building a triangle in R^2 with appropriate side lengths. The case of metric sets of four points has been studied by Blumenthal (1936), who found that raising the metric to a positive power, lower or equal to 1/2, compels the embeddability.

4. Monotonic embedding in Euclidean spaces

We have seen in Section 3 that, under quite general conditions, data sets can be monotonically embedded in at least one real metric space in a simple way (Lemma 2), while this does not guarantee the embeddability in an Euclidean space (Lemma 3). In this section, one states general conditions of monotonic embeddability of data sets in Euclidean spaces.

Definition 1. Let S be any set. A 'dissimilarity' function associated to S is a real valued function μ on $S \times S$ such that, for any $a, b \in S$, one has $\mu(a,a) = 0$, and if $a \neq b$ then $\mu(a,b) = \mu(b,a) > 0$. The set of dissimilarity functions associated to S is denoted $\delta(S)$.

Note that metrics are dissimilarities (which, in addition, satisfy triangle inequality), while semi-metrics are not dissimilarities since they allow distinct elements to have a zero distance.

Definition 2. One denotes G the set of functions $g(\mu;p)$ of one real variable $\mu \geq 0$, and one real parameter $p > 0$ such that:

$g(\mu;p)$ is a continuous strictly increasing function of μ,

$g(0;p) = 0$,

for any $\mu > 0$, for any real $\epsilon > 0$, there is a real $a > 0$ such that

$p \leq a \Rightarrow |g^2(\mu;p) - 1| \leq \epsilon$.

Example 1. Power function $g(\mu;p) = \mu^p \in G$, with the restriction that $\sup \mu < \infty$:Without loss of generality, one can assume that $\epsilon < 1$, then

$$a = \begin{cases} \ln(\epsilon+1)/(2\ln\mu), & \text{if } \mu > 1, \\ \ln(1-\epsilon)/(2\ln\mu), & \text{if } \mu < 1, \\ > 0, & \text{if } \mu = 1. \end{cases}$$

Example 2. Weibull's function

$$g(\mu;p) = 1 - \exp(-\mu^r/p) \in G, \qquad \text{with } r > 0:$$

with $\epsilon < 1$ one has

$$a = \frac{-\mu^r}{\ln(1-\sqrt{1-\epsilon})}.$$

Lemma 4. *Let $T = [t_{ij}]_{i,j=0,\ldots,n}$ be a trivial distance matrix associated to $n+1$ distinct elements, that is $t_{ii} = 0$, and if $i \neq j$ then $t_{ij} = 1$. Consider a symmetric matrix of the form $[g^2(\mu_{ij};p)]_{i,j=0,\ldots,n}$, where $\mu \in \delta(S)$, for a given set S of $n+1$ distinct elements* (Definition 1), *and $g \in G$* (Definition 2). *Then for any real $\epsilon > 0$, there is a real $a > 0$ such that*

$$0 < p \leq a \Rightarrow \|[g^2(\mu_{ij};p)]_{i,j=0,\ldots,n} - T\|_\infty \leq \epsilon,$$

where the matricial norm $\|B\|_\infty = \max_i \Sigma_j |b_{ij}|$.

Proof. The diagonals of the two symmetric matrices $[g^2(\mu_{ij};p)]_{i,j=0,\ldots,n}$ and T are zero, which implies that there are at most n non-zero differences per row of the difference matrix. For each non-diagonal cell of the first matrix, there is a real $a_{ij} > 0$ such that $p \leq a_{ij} \Rightarrow |g^2(\mu_{ij};p) - 1| \leq \epsilon/n$, by definition of the function g. This implies that there is an appropriate real number a such that $a \geq \min_{0\leq i<j\leq n} a_{ij} > 0$, which completes the proof. □

Theorem 3. *Let S be any finite set comprised of $n+1$ distinct elements, consider a dissimilarity $\mu \in \delta(S)$, and a function $g \in G$. Then the following two statements are true*:

(i) *There is a real $\alpha(S) > 0$ such that for any stricly positive real $p \leq \alpha(S)$, the space $(S, g(\mu;p))$ is a metric set isometrically embeddable in an Euclidean space whose dimension is at most n.*
(ii) *There is a real $\beta(S) > 0$, $\beta(S) \leq \alpha(S)$, such that for any stricly positive real $p < \beta(S)$, the space $(S, g(\mu;p))$ is a metric set isometrically embeddable in an Euclidean space whose dimension is exactly n.*

Proof. Step 1. Let $T = [t_{ij}]_{i,j=0,\ldots,n}$ be a trivial distance matrix, whose diagonal coefficients equal 0, and other coefficients equal 1. Then for any vector $v \in R^{n+1}$ such that $\sum_{i=0}^n v_i = 0$, one obtains

$$v'Tv = v'\left[\left(\sum_{j=0}^n v_j\right) - v_i\right]_{i=0,\ldots,n} = v'(-v) = -\|v\|^2 \leq 0,$$

which proves that T is almost negative definite.
Step 2. Consider the symmetric matrix $D_p = [g^2(\mu_{ij};p)]_{i,j=0,\ldots,n}$ associated to the space $(S, g(\mu;p))$, for a given $p > 0$. One can write

$$D_p = T + (D_p - T),$$

and for any vector $v \in R^{n+1}$ such that $\sum_{i=0}^n v_i = 0$, one has

$$v'D_pv = v'Tv + v'(D_p - T)v = -\|v\|^2 + v'(D_p - T)v,$$

after Step 1. Then one obtains the equivalence

$$v'D_pv \leq 0 \Leftrightarrow v'(D_p - T)v \leq \|v\|^2,$$

while a sufficient condition for obtaining the last inequality is that the greatest eigenvalue of the symmetric matrix $(D_p - T)$ not be greater than 1 (after usual properties of Rayleigh's ratio for symmetric matrices). Now, after the well-known theorem of Gerschgorin, one knows that the greatest eigenvalue of any square matrix B cannot exceed $\|B\|_\infty$, while after Lemma 4, there is a real $a > 0$ such that

$$0 < p \leq a \Rightarrow \|D_p - T\|_\infty \leq 1.$$

Then, clearly, for any p in the above interval, the matrix D_p is almost negative definite, and (i) of Theorem 3 is proved (in account of Theorem 2 and Corollary 2). Note, however, that the condition used for defining the upper bound (a) of the critical interval of p is sufficient but not necessary, and one has in fact $\alpha(S) \geqq a$.
Step 3. After Lemma 4, for any ϵ such that $0 < \epsilon < 1$, there is a real $b > 0$ such that $0 < p \leq b \Rightarrow \|D_p - T\|_\infty \leq \epsilon$. Then, for any p in this interval, for any non-zero vector $v \in R^{n+1}$ such that $\sum_{i=0}^n v_i = 0$, one has $v'D_pv < 0$, which

implies that the matrix $[g^2(\mu_{0j};p) + g^2(\mu_{0j};p) - g^2(\mu_{ij};p)]_{i,j=1,\ldots,n}$ is strictly positive definite (after Corollary 1), and then the dimension of the embedding space is exactly n (after Corollary 2). Note that one has in fact $\beta(S) > b$. Moreover, comparing (i) with (ii), it is clear that $\beta(S) \leq \alpha(S)$, and Theorem 3 is proved. □

Corollary 3. *Let S be a given set of n distinct elements, let Ω be a generalization set including S, while $\mu \in \delta(\Omega)$, and $g \in G$. Then there is a real $\alpha(\Omega|S) > 0$ such that for any stricly positive real $p \leq \alpha(\Omega|S)$, and for any $y \in \Omega$, the space $(S \cup \{y\}, g(\mu;p))$ is a metric set isometrically embeddable in a n-dimensional Euclidean space.*

Proof. One obviously has $\alpha(\Omega|S) = \inf_{y\in\Omega} \alpha(S \cup \{y\})$, while after Theorem 3:
if $y \in S$ then $\alpha(S \cup \{y\}) = \alpha(S) > 0$, if $y \notin S$ then $\alpha(S \cup \{y\}) > 0$,
which completes the proof. □

When the embedding parameter p tends to 0, the space $(S, g(\mu;p))$ tends to (S,t), where t is the trivial distance. Then the points of the mapping in an Euclidean space tend to the vertices of an equilateral polytope (e.g. an equilateral triangle for $n = 2$, an equilateral tetraedron for $n = 3$, and so on), which account for the equidistance of these points. Then the particular structure of the data set is represented more and more weakly in the embedding space as p tends to 0. However, as long as $p > 0$, the structural information remains available, since the g transform remains inversible, by virtue of its strict monotonicity with respect to μ. Corollary 3 is particularly important for neural network applications since it states that one can embed any generalization item (current input), belonging to a given generalization set Ω, together with a finite learning set (or a set of landmarks) S. Note, however, that this does not mean that one can globally embed an infinite set Ω.

5. Relation with the city-block metric

There is a particular relation between city-block metric sets and Euclidean metric sets. This relation allows for a very simple determination of an appropriate g function with an appropriate embedding parameter p, whenever one knows that a data set is a city-block metric set. Moreover, it can be important to know such a relation for the study of psychological spaces, since various observations suggested that objects described on dimensions of the same nature (e.g. width and height) are encoded by humans in an Euclidean space, while objects described on heterogeneous dimensions (e.g. size and colour) are encoded in a city-block space. However, there has been some debate on this question (Ennis, 1988; Nosofsky, 1986; Shepard, 1986).

Lemma 5. *Let S be a finite set of n + 1 distinct points of R^m, and d be a city-block metric (M_1) on $S \times S$, where d is finite on this set. Then the set $(S, d^{1/2})$ is isometrically embeddable in the Euclidean space (R^n, M_2).*

Proof. After the definition of a city-block metric, one has

$$[d_{ij}]_{i,j=0,\ldots,n} = \sum_{k=1}^{m} [d_{ij}^{(k)}]_{i,j=0,\ldots,n},$$

where the upper index (k) indexes the dimensions, and each matrix $[d_{ij}^{(k)}]_{i,j=0,\ldots,n}$ is the matrix of a Minkowskian metric on a real space of dimension 1. In dimension 1, all Minkowskian metrics are equal, and in particular they are equal to the Euclidean one. Since the power function is in G for finite measures, Theorem 3 guarantees that if one raises an Euclidean metric to the power 1/2, it remains an Euclidean metric (for a different set of points), and after Theorem 2, if one squares this new Euclidean metric one obtains that all matrices $[d_{ij}^{(k)}]_{i,j=0,\ldots,n}$, $k = 1,\ldots,m$, are almost negative definite. Since a sum of almost negative definite matrices is an almost negative definite matrix, one concludes that $[d_{ij}]_{i,j=0,\ldots,n}$ is almost negative definite, which implies that $(S, d^{1/2})$ is isometrically embeddable in the Euclidean space (R^n, M_2), and Lemma 5 is proved. □

6. Cholesky factorization of a singular matrix

In this section, one states a result which will be necessary for defining an embedding algorithm in Section 7. One knows that a symmetric strictly positive definite matrix B can be factorized as a product of the form $B = X'X$, where X is a non-singular upper triangular matrix. This is the well-known Cholesky factorization of B. Now, a difficulty arises whenever B is singular, since in this case the Cholesky factorization leads to a division by zero resulting from an indetermination of the form $0 \cdot x = 0$ in Cholesky's equations. The following result allows for removing this indetermination, and then to define a simple variant of the Cholesky factorization for possibly singular matrices.

Theorem 4. *Let B be a symmetric, possibly singular, positive definite matrix of order $n \times n$. Then there is an upper triangular matrix X such that $X'X = B$, $x_{ii} \geq 0$, $1 \leq i \leq n$, and if for an index i one has $x_{ii} = 0$, then $x_{ij} = 0$, $1 \leq j \leq n$. Moreover, the matrix X with these properties is unique.*

Proof. *Existence proof.* Since B is symmetric positive definite, there is a square matrix Y such that $Y'Y = B$ (for example $Y = \Delta^{1/2}Q'$, as in the proof of Corollary 2). One can define a special variant of the usual QR factorization, this variant being of the form $Y = HX$, where H is an orthogonal matrix and X is upper triangular:

1. $H_0 = 0$ (auxiliary vector),
2. for $j = 1$ to n do (3)–(5)
3. $Z_j = Y_j - \sum_{k=0}^{j-1} (Y_j \cdot H_k) H_k$,
4. if $Z_j = 0$ then replace it by any non-zero vector Z_j such that $Z_j \perp \{H_1, \ldots, H_{j-1}, Y_j, \ldots, Y_n\}$,
5. $H_j = Z_j / \|Z_j\|$,
6. $X = H'Y$.

In the above procedure, H_j or Y_j, with $j > 0$, stands for the jth column vector of the corresponding matrix. Steps (1), (2), (3) and (5) correspond to the well-known Gram–Schmidt orthonormalization process, while a variant is introduced at step (4). Then a crucial point is the existence of an appropriate non-zero vector Z_j when step (3) provides a zero vector. All vectors have the dimension n, while if step (3) provides a zero vector, this means that Y_j is a linear combination of the vectors H_k, $k = 1, \ldots, j-1$. Then the rank of the matrix whose collumn vectors are $\{H_1, \ldots, H_{j-1}, Y_j, \ldots, Y_n\}$ is at most $n-1$, which implies that there is a non-zero vector of dimension n which is orthogonal to all these vectors, and step (4) is valid. Now, one verifies that step (6) provides a matrix X which has all the desired properties.

X is upper triangular since $H_j \perp Y_k$, $k = 1, \ldots, j-1$.

Using the equation of step (3), one obtains that

$$Y_j \cdot Z_j = \|Y_j\|^2 - \sum_{k=1}^{j-1} (Y_j \cdot H_k)^2 \geq 0,$$

since the vectors H_k, $k = 1, \ldots, j-1$, form an orthonormalized basis (complete or not for Y_j), and $\sum_{k=1}^{j-1} (Y_j \cdot H_k)^2 \leq \|Y_j\|^2$, by virtue of Bessel–Parseval's inequality. If the basis is not complete for Y_j, then the above inequality is strict, which implies that $x_{jj} = H_j \cdot Y_j > 0$, while if the basis is complete for Y_j, then step (3) provides a zero vector, and step (4) guarantees that H_j is orthogonal to all columns of Y, which implies that the whole jth row of X is zero.

Finally, one has $X'X = Y'HH'Y = Y'Y = B$, since H is orthogonal (that is $HH' = I$), which completes the existence proof.

Comment. Assume that a vector H_k has been chosen using step (4). Then this choice does not affect the computation of the next vectors by step (3) since $(Y_j \cdot H_k) = 0$, $j = k+1, \ldots, n$. In particular, the set of vectors for which step (3) generates a zero result remains the same, whatever be the particular choice of H_k. On the other hand, all vectors generated using step (4) provide an identical effect on the matrix X, that is a zero row. This implies that the particular choice of certain vectors by step (4) does not affect the resulting matrix X, and then this matrix is unique for a given matrix Y, while the orthogonal matrix H is not unique if Y (and hence B) is singular.

Unicity proof. It remains to prove that X does not depend on a particular choice of the matrix Y. Since an appropriate matrix X exists (see the existence proof earlier), one can write the Cholesky's equations of the system $B = X'X$. Using an auxiliary row vector $[x_{0j}]_{j=1,\ldots,n} = 0$, for writing convenience, one obtains:

$$j = 1, \ldots, n, \qquad i = 1, \ldots, j:$$

if $i = j$ then

$$b_{ii} = \sum_{k=0}^{i} x_{ki}^2 \Rightarrow x_{ii} = \sqrt{b_{ii} - \sum_{k=0}^{i-1} x_{ki}^2},$$

else

$$b_{ij} = x_{ii} x_{ij} + \sum_{k=0}^{i-1} x_{ki} x_{kj} \Rrightarrow x_{ij}$$

$$= \begin{cases} \left(b_{ij} - \sum_{k=0}^{i-1} x_{ki} x_{kj} \right) / x_{ii}, & \text{if } x_{ii} > 0, \\ 0, & \text{if } x_{ii} = 0 \end{cases}$$

by hypothesis on X properties.

One can note that the solution X is unequivocally determined by the constraints without any reference to a particular matrix Y, which completes the proof of Theorem 4. □

Note that we have just defined a Cholesky's factorization of a possibly singular symmetric positive definite matrix B, which was in fact the main goal of Theorem 4. This factorization is very similar to usual Cholesky's factorizations of strictly positive definite symmetric matrices, except that the singular case ($x_{ii} = 0$) is allowed, with a unique guaranteed solution.

7. Monotonic embedding algorithm

We are now ready to define a monotonic embedding algorithm of data sets in Euclidean spaces with the desired properties for neural network applications. First, one can note that if Y is an isometric embedding mapping of a set S in an Euclidean space, then so is QY, for any orthogonal matrix Q, since the Euclidean distance is invariant by any orthogonal transform of coordinates. In particular $X = H'Y$, where X is upper triangular as in Section 6, is a solution. After Theorem 3 (and Corollary 3), for a given data set S, there is $\alpha(S) > 0$ such that $0 < p \leq \alpha(S)$ implies that

$$\tfrac{1}{2}[g^2(\mu_{0i}; p) + g^2(\mu_{0j}; p) - g^2(\mu_{ij}; p)]_{i,j=1,\ldots,n} = X'X = B,$$

where $\mu \in \delta(S)$, $g \in G$, and the solution X of course depends on p.

After Theorem 4, this system can be solved using an appropriate Cholesky factorization, even if B is singular, which can happen if $p \geq \beta(S)$. It remains to define a way of finding an appropriate p, given that in general, one does not know $\alpha(S)$ a priori. For doing this, one can exploite the fact that if $p \le \alpha(S)$, then B is positive definite and the Cholesky factorization has a real solution, while if $p > \alpha(S)$, then B has negative eigenvalues and there are negative arguments for a square root function in the computation of the diagonal coefficients of X. Hence, each time that one detects such an imaginary value case, this means that $p > \alpha(S)$, and that one must lower p. Then a simple bounding procedure allows for approximating $\alpha(S)$ (resp. $\beta(S)$) as closely as one wants.

7.1. Procedure for embedding one element

The following procedure (EMBED) allows for embedding the element number j of a set, while the elements of number 0 to $j-1$ have been previously embedded. This is the fundamental procedure which is called by all particular application algorithms. One can eventually limit the dimension m of the embedding space, while if no limit of dimension is fixed, it suffices to call EMBED with an arbitrary strictly positive parameter $m \ge j$, remembering that $x_{ij} = 0$ if $i > j$. One can arbitrarily limit the dimension m since the Cholesky factorization of the matrix B is not only incremental with respect to the collumns, but also with respect to the rows. One assumes that a dissimilarity function μ is associated to the data set, that a function $g \in G$ has been fixed, and that the embedding parameter p currently has a defined value. The procedure returns an arbitrary negative diagonal value ($x_{jj} = -1$) when the jth element is not embeddable using the current value of p.

EMBED (j,m,p)
  if $j < m$ then for $i := j + 1$ to m do $x_{ij} := 0$
  if $j > 0$ then
    for $i := 1$ to min(j,m) do
      $b_{ij} := (1/2)(g^2(\mu_{0i};p) + g^2(\mu_{0j};p) - g^2(\mu_{ij};p))$
      if $i > 1$ then $s := \sum_{k=1}^{i-1} x_{ki}x_{kj}$ else $s := 0$
      $b_{ij} := b_{ij} - s$
      if $i = j$ then
        if b_{ij} is very close to 0 then $b_{ij} := 0$
        if $b_{ij} < 0$ then $x_{ij} := -1$ else $x_{ij} := \sqrt{b_{ij}}$
      if $i \ne j$ then
        if $x_{ii} = 0$ then $x_{ij} := 0$ else $x_{ij} := b_{ij}/x_{ii}$.

Note: in the case ($i = j$), if b_{ij} is very close to 0 then it is set to 0 before the test ($b_{ij} < 0$) in order to take into account the rounding errors which occur in computer's floating point arithmetic.

7.2. Multidimensional scaling procedure

The procedure named MDS allows for monotonically embedding a set S of $n + 1$ elements, for given μ and g functions, while one requires that any diagonal value of the matrix X is at least equal to a positive value e. If one calls MDS with $e = 0$ then the final value of p is an approximation of $\alpha(S)$ with the specified 'precision'. The parameter pSup (strictly positive) is choosen as the upper bound of the search interval of p. If the final value of p is equal to pSup, this means that pSup $\le \alpha(S)$. Similarly, one can obtain an approximation of $\beta(S)$ by choosing e just greater than zero. Take care that too large a value of e can be incompatible with the data, which leads p to tend to zero.

MDS(e, n, pSup)
  EMBED(0, n, 1) {the first element is the origin}
  pInf := 0
  p := pSup
  repeat
  $j := 0$
    repeat
    $j := j + 1$
    EMBED(j, n, p)
    until ($j = n$) or ($x_{jj} < e$)
  if ($x_{jj} < e$) then pSup := p else pInf := p
  p := (pInf + pSup)/2
  until (pSup − pInf) < precision.

The procedure MDS1 allows for embedding a new element together with $n + 1$ previously embedded elements. MDS1 must be called with the current value of the embedding parameter p. This value will be lowered in a call to MDS if necessary. For avoiding this modification, replace the call to MDS by any appropriate instruction (for example, display the message 'the new item is not embeddable').

MDS1(e, n, p)
  $m := n + 1$
  EMBED(m, m, p)
  if $x_{mm} < e$ then MDS(e, m, p)

The procedure MDS1 is useful for approximating $\alpha(\Omega|S)$ of Corollary 3, while S is a learning set of $n + 1$ elements, and Ω is a generalization set. Each generalization element must be embeddable together with S, independently of other generalization elements. Then repeatedly sampling Ω and embedding each generalization item together with S by MDS1(0, n, p), one can hope that p converges to $\alpha(\Omega|S)$.

7.3. Fixed dimension spatial maps

The procedure named MAP allows for building a spatial map of fixed dimension m from a set of approximated distances between $n + 1$ objects ($n \ge m$). The first $m + 1$ objects are taken as landmarks, the first one being the origin of coordinates ($X_0 = 0$). The embedding space is completely defined by the landmarks, while the remaining objects are embedded in this space only, and their mutual distances are not taken into account. In this type of application, μ is in general an approximation of an Euclidean distance (however, this is not necessary), and one chooses for g the power function, that is $g(\mu;p) = \mu^p$.

Table 1
Averaged values of $\alpha(S)$ on four runs of MDS(0, n, 10), and dimension (m') of the embedding space as functions of the size of the set ($n+1$), and the type of data space, with $g=\mu^p$, and $m=n$

Data space:	Non-metric		Metric		Euclidean	
n	$\alpha(S)$	m'	$\alpha(S)$	m'	$\alpha(S)$	m'
4	0.202	3	0.894	3	1.102	3
8	0.159	7	0.781	7	1.003	7
16	0.084	15	0.596	15	1.003	15
32	0.054	31	0.540	31	>1	31
64	0.035	63	0.528	63	>1	63
128	0.023	127	0.499	127	>1	127

Hence, if μ is actually an Euclidean distance between the landmarks, then one obtains $p=1$. The bounding method used in the procedure MDS for the search for p is replaced here by a simple decay method, starting from $p=1$, and applying, if necessary, a reduction coefficient ($0<\text{reduc}<1$) close to 1. This strategy is fast whenever μ is close to an Euclidean distance between the landmarks. If this is not the case, then the decay method can of course be replaced by another one, for axample the one used in the MDS procedure. Taking a small $e>0$ ensures that the embedding space is exactly of dimension m.

```
MAP(e, m, n, reduc)
  EMBED(0, m, 1) {the first element is the origin}
  p := 1
  OK := false
  repeat
  j := 0
    repeat
    j := j + 1
    EMBED(j, m, p)
    until (j = n) or ((j ≤ m) and (x_jj < e))
  if (j ≤ m) and (x_jj < e) then p := reduc × p else
                          OK := true
  until OK
```

Note that in real life navigational applications, the landmarks are frequently fixed objects. In this case, the embedding space must be computed only once, possibly in quite good conditions for the approximation of distances (e.g. prior exploration of the environment). Then the on-line computation reduces to embedding the remaining (possibly moving) objects, that is:

```
MAP1(m, n, p)
  for j := m + 1 to n do EMBED(j, m, p).
```

This is very simple, while the main problem is of course the on-line approximation of the distances.

Table 2
Averaged values of $\alpha(S)$ on four runs of MDS(0, 128, 10), and dimension (m') of the embedding space as functions of the type of metric data space, and of its dimension (m), with $g=\mu^p$, and $n=128$

Data space:	Euclidean		Metric		City-block	
m	$\alpha(S)$	m'	$\alpha(S)$	m'	$\alpha(S)$	m'
4	1	4	0.146	127	0.504	127
8	1	8	0.114	127	0.511	127
16	1	16	0.160	127	0.524	127
32	1	32	0.219	127	0.547	127
64	1	64	0.330	127	0.592	127

8. Some numerical results

This section summarises numerical results obtained by the application of the procedure MDS (Section 7.2) to data sets of various types and sizes. For 'non-metric' data sets, dissimilarity symmetric matrices with zero diagonal coefficients were generated by a uniform random sampling of non-diagonal coefficients in the interval [0.01,100]. The distance matrices of metric data sets were generated by computing Minkowskian M_∞ distances between random vectors (uniform random coordinates in $[-1,1]$). After Lemma 1 (Fréchet–Schoenberg), this is representative of the whole set of finite metric sets. The distance matrices of 'Euclidean' data sets were generated in the same way, using the Euclidean metric M_2. while distance matrices of 'City-Block' metric sets were obtained using the metric M_1. The size of the sets ($n+1$ elements) was varied from $n=4$ to 128. The actual dimension (m) of the metric spaces was also varied. Tables present averaged values of the observed embedding parameter bounds ($\alpha(S)$) on four independent problems solved by MDS(0, n, 10), and the obtained dimension of the embedding space ($m'=$ number of non-zero diagonal coefficients of X). Tables 1 and 2 present results obtained with the power function as g function, while Table 3 presents results obtained with two distinct Weibull functions ($r=1$ and 2, respectively) as g functions. The inspection of Table 1 shows that $\alpha(S)$ is lower for non-metric sets than for metric sets, and that it is greater than one for Euclidean sets only. Moreover, $\alpha(S)$ decreases as the

Table 3
Averaged values of $\alpha(S)$ on four runs of MDS(0, n, 10), and dimension (m') of the embedding space as functions of the size of the set ($n+1$), and the type of data space, with $g=1-\exp(-\mu^r/p)$, $m=n$

Data space	Non-metric		Metric		Euclidean	
	$\alpha(S)$	m'	$\alpha(S)$	m'	$\alpha(S)$	m'
$r=1$, $n=16$	1.419	15	1.594	15	>10	(16)
$r=1$, $n=128$	0.092	127	1.273	127	>10	(128)
$r=2$, $n=16$	6.022	15	0.802	15	5.648	15
$r=2$, $n=128$	0.009	127	0.831	127	>10	(128)

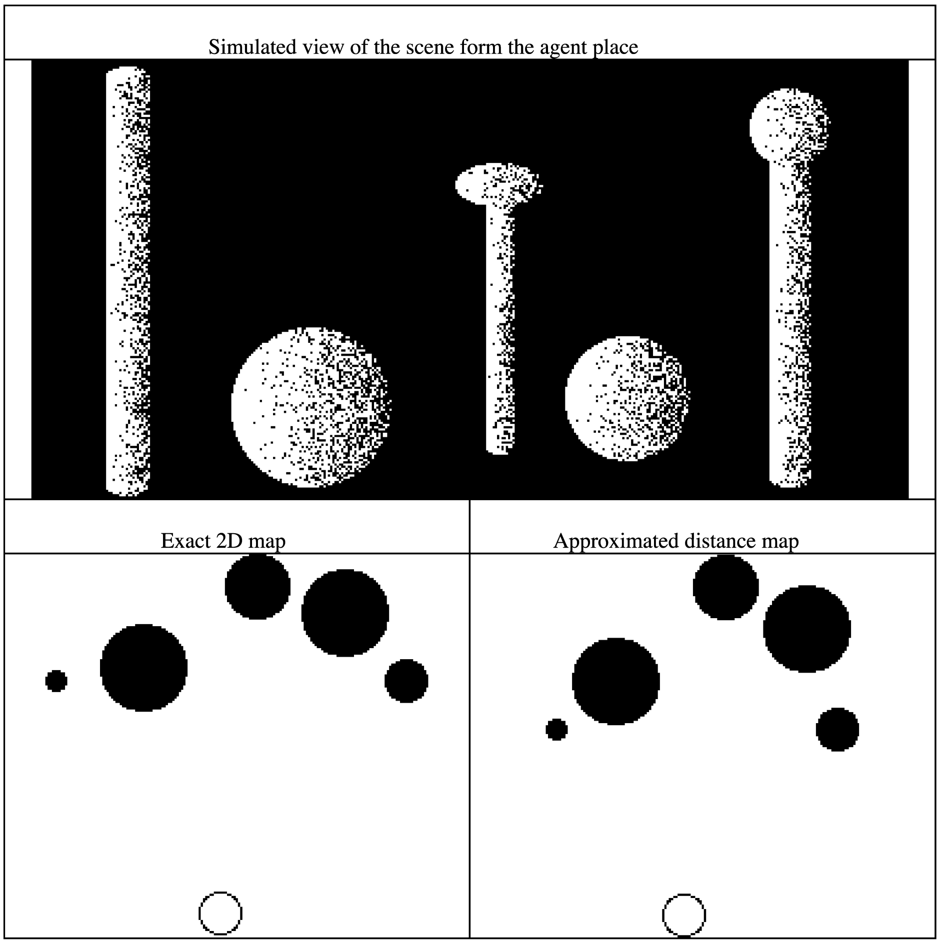

Fig. 1. Simulated view of a scene with three landmarks (pegs) and two possibly moving objects (balls), followed by two maps computed by the embedding algorithm from exact and approximated distance data. In these maps, the white disk stands for the autonomous agent, while the other objects are distinguished by the diameter of their head. The origin of the coordinates is always the thin peg.

size of the set (n) increases, always remaining greater than one for Euclidean sets. The obtained dimension of the embedding space for $p = \alpha(S)$ was always $m' = n - 1$, while the actual dimension of the used metric sets was $m = n$. Additional results were computed for City-Block metric sets. As expected from Lemma 5, $\alpha(S)$ was always greater than 0.5 for these sets and one obtained an averaged $\alpha(S) = 0.673$ with $n = m = 128$.

Table 2 shows the effect of the actual dimension (m) of three types of metric sets (with $n = 128$). The results are clear for Euclidean sets, where the actual dimension was always detected by the algorithm ($m' = m$). For the two other types of metric sets, the actual dimension was not detected ($m' = n - 1$), however, one can observe that $\alpha(S)$ systematically varied as a function of m. Further theoretical investigations are required for understanding this relation.

Table 3 rapidly provides some elements concerning the behavior of the embedding algorithm with Weibull functions. The algorithm works well with these functions which are more appropriate to function approximation contexts than to distance geometry.

Additional results showed, in all cases, that $\beta(S) = \alpha(S)$, that is, taking an embedding parameter p just lower than $\alpha(S)$ always provides an embedding space of dimension n exactly, whatever be m' for $p = \alpha(S)$. Finally, a general observation which can be outlined from the above results is that dissimilarity functions (which were completely random in this study) must be preferably built in a way which provides them with properties close to those of a metric, in order to avoid very low values of $\alpha(S)$ and large transformations of the data space.

9. Example of application to a robot navigation problem

As suggested in Section 1, there are various application fields and various ways of exploiting an embedding algorithm. The example presented in this section has the advantage of providing suggestive visual illustrations. Robotic and compuer vision are very active fields, while robot vision systems can vary considerably in their sophistication. For this illustration, minimal hypothesis concerning the robot technology were retained. We assume

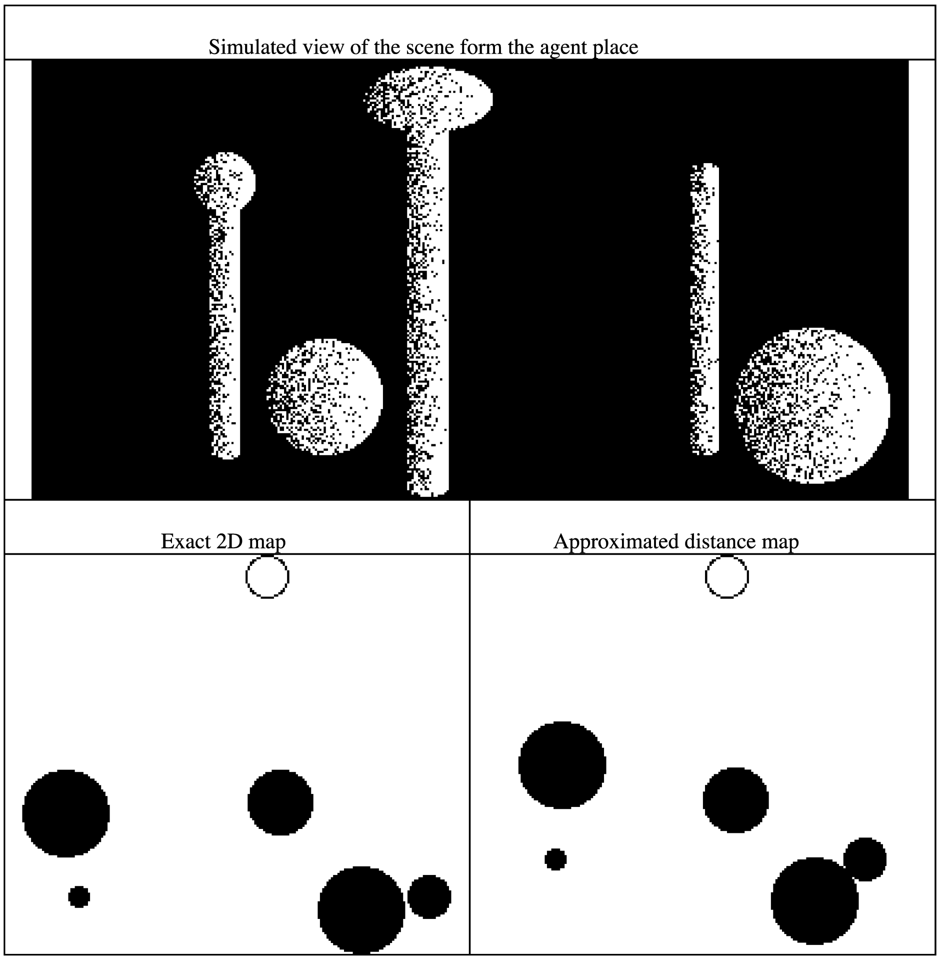

Fig. 2. Similar to Fig. 1, with the same landmarks and a different configuration of the three moving objects.

a robot (autonomous agent) equiped with a simple video camera whose images are numerised in a 200×400 matrix of binary pixels (black/white). We also assume that the robot's computational power is limited, and that the computer vision program is not very sophisticated. However, this program must be able to approximately segment an image into regions corresponding to individual objects, discriminate objects of various shapes, and measure the angular distance between any two points. If the two considered points are not simultaneously present in the visual field (e.g. if the robot lands between two objects), then the angular distance is the absolute rotation angle of the camera required for successively centering the two points. In addition, we assume that the program can use a data base providing the actual size of objects (at least approximately). Such a data base may have been obtained from a prior exploration of the environment. With all this information available, one can approximate the distances between objects (including the agent) by applying usual projection and triangulation formulas (such as those used in astronomy). Then the approximated distance matrix can be used as the input of the monotonic embedding algorithm in order to compute a 2D map of the spatial configuration of the objects (including the agent). Certain fixed objects can be used as landmarks, while at least three objects (fixed or not) are required for computing a 2D map. The main difficulties result from the presence of image numerization noise, shadows, noisy background, partially hidden objects, and poorly discriminable landmarks. This can result in a certain amount of error in the approximation of distances, which leads to distortions in the spatial map. Figs. 1 and 2 show two views of a simulated environment made of three landmarks (pegs) and two moving objects (balls). The choosen order of these objects is: thin peg, round head peg, ellipsoid head peg, and the balls. The two balls have undiscernable shapes, hence their set is a cluster which may eventually be encoded in a special way after a spatial map has been computed (Courrieu, 2001). Below each view, on can see two 2D maps computed by the procedure MAP(e, 2, 5, 0.999), with a small positive e appropriate to the data scale (which is arbitrary). The first map was computed by the embedding algorithm from exact distances (known a priori), while the other map was computed from distances appproximated by a rudimentary (poorly performing) computer vision program. Despite the distance approximation errors generated by this program (up to 20%), one can see in Figs. 1 and 2 that the obtained maps are quite close to the exact ones.

10. Conclusion

Embedding algorithms and multidimensional scaling potentially have a wide set of applications in various fields (data analysis, function approximation on non-Euclidean topological spaces, autonomous agent navigation problems, psychological science). Mathematical fundations of a fast monotonic embedding algorithm of data sets in Euclidean spaces were presented, and then the algorithm was defined, with variants for various types of applications. The particularity of the algorithm, with respect to usual multidimensional scaling methods, is that it is straight, fast, and incremental, which makes it particularly appropriate to Neural Network applications and on-line computation. Some general numerical results were provided, and an illustrative example of application in robot navigation was presented. Further investigations are needed for solving remaining problems such as the optimal reduction of the embedding space dimension, however, this mainly concerns data analysis applications.

Acknowledgement

This work was partially supported by a grant from Fond National pour la Science – ACI "Cognitique" (2000, #90).

References

Blumenthal, L. M. (1936). New theorems and methods in determinant theory. *Duke Mathematical Journal*, *2*, 396–404.

Courrieu, P. (2001). Two methods for encoding clusters. *Neural Networks*, *14*, 175–183.

Cox, T. F., & Cox, M. A. A. (1994). *Multidimensional scaling*. London: Chapman & Hall.

Cressant, A., Muller, R. U., & Poucet, B. (1997). Failure of centrally placed objects to control the firing fields of hippocampal place cells. *The Journal of Neuroscience*, *17*(7), 2531–2542.

Cressant, A., Muller, R. U., & Poucet, B. (1999). Further study of control of place cell firing by intra-apparatus objects. *Hippocampus*, *9*, 423–431.

Demartines, P., & Hérault, J. (1997). Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. *IEEE Transaction on Neural Networks*, *8*, 148–154.

Donoghue, W. F. (1974). *Monotone matrix functions and analytic continuation*. Berlin: Springer.

Ennis, D. M. (1988). Confusable and discriminable stimuli: comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, *117*(4), 408–411.

Fréchet, M. (1910). Les dimensions d'un ensemble abstrait. *Mathematische Annalen*, *68*, 145–168.

Girosi, F., & Poggio, T. (1990). Networks and the best approximation property. *Biological Cybernetics*, *63*, 169–176.

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

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, *29*, 115–129.

Lowrance, R. A., & Wagner, R. A. (1975). An extension of the string to string correction problem. *JACM*, *22*(2), 177–183.

Micchelli, C. A. (1986). Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, *2*, 11–22.

Muller, R. U. (1996). A quarter of century of place cells. *Neuron*, *17*, 813–822.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process model. *Annual Review of Psychology*, *43*, 25–53.

O'Keefe, J., & Nadel, L. (1978). *Hippocampus as a cognitive map*. Oxford: Clarendon Press.

Okochi, M., & Sakai, T. (1982). Trapezoïdal dynamic programming matching with time reversibility. *Proceedings of the IEEE-ASSP Conference, Paris*, 1239–1242.

Poggio, T., & Girosi, F. (1990). Networks for approximation and learning. *Proceedings of the IEEE*, *78*(9), 1481–1497.

Poucet, B., Save, E., & Lenck-Santini, P.-P. (2000). Sensory and memory properties of hippocampal place cells. *Reviews in the Neurosciences*, *11*, 95–111.

Schoenberg, I. J. (1937). On certain metric spaces arising from euclidean spaces by a change of metric and their imbedding in Hilbert space. *Annals of Mathematics*, *38*(2), 787–793.

Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, *44*, 522–536.

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, II. *Psychometrika*, *27*, 219–246.

Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, *115*, 58–61.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Vinstuk, T. K. (1968). Speech discrimination by dynamic programming. *Kibernetika*, *4*(1), 81–88.

Wagner, R. A., & Fischer, M. J. (1974). The string to string correction problem. *JACM*, *21*(1), 168–173.

Young, F. W., & Torgerson, W. S. (1967). Torsca, a Fortran IV program for Shepard–Kruskal multidimensional scaling analysis. *Behavioral Science*, *12*, 468.