1. Introduction.

Ultrametric spaces, known since long to mathematicians (1), have recently found applications in physics in connection with Parisi's solution of the Sherrington-Kirkpatrick model for spin glasses [2]. A recent review on the importance of ultrametricity for physics and other sciences is given in [3].

The purpose of this paper is to show that a similar ultrametric structure can be defined on the set of all histories of a general dynamical system at a given time. Such a set is a partition of the phase space that becomes more and more refined as time goes on. Therefore the entropy of the partition increases with time and we shall show that it is possible to define an ultrametric distance between histories from the entropy distance between partitions. Such a definition is quite natural because the complexity of a set of histories is naturally related to its entropy.

(1) For a recent discussion of ultrametricity in mathematics see for example [1].
The paper is organized as follows. In section 2 we shall define a distance between the elements of a partition, a distance that will turn out to be ultrametric. In section 3 we will introduce a new ultrametric distance between two histories at a given time that depends upon the entropy produced during the evolution from the common ancestor (to be defined) of the two histories. Thus this distance depends upon the choice of the probability measure. In section 4 we shall generalize the distance defined in section 3 by considering instead of the usual entropy the set of $\alpha$-entropies [4] for $\alpha \in [0, 1]$. In section 5 we shall show that in the case of disordered systems the analogue of the time evolution is the resolution process of clusters of ergodic states into their components. The order parameter of these systems is related to the 2-entropy, thus showing its true connection with the notion of complexity. Finally we shall give the relation between the geometric distance on phase space and the probabilistic distance provided by the von Neumann-Shannon entropy.

2. The space of histories and its ultrametric structure.

Let us consider a dynamical system $(X, \mathcal{F}, m, T)$ where $(X, \mathcal{F}, m)$ is a probability space and $T$ is a measurable automorphism of $(X, \mathcal{F})$. Here $X$ denotes the Cartesian product of the phase spaces of a physical system at times $t = 1, 2, \ldots$; $\mathcal{F}$ is a $\sigma$-algebra of measurable subsets (events) of $X$ and $m$ a probability measure (a state) not necessarily preserved by $T$.

An experiment with a finite number of outcomes performed at time $t = 1$ on $(X, \mathcal{F}, m)$ is represented by a measurable partition $\gamma = \{A_0, \ldots, A_k\}$ of $X$. The same experiment performed at $t = n$ ($n$ is an integer) is represented by the partition $T^{- (n - 1)} \gamma = \{T^{- (n - 1)} A_0, \ldots, T^{- (n - 1)} A_{k-1}\}$ and the sequence of experiments repeated $n$ times from $t = 0$ to $t = n$ is represented by the partition

$$
\gamma^{(n)} = \gamma \vee T^{-1} \gamma \vee \cdots \vee T^{- (n - 1)} \gamma
$$

$$
= \gamma^{(\ell)} \vee T^{-\ell} \gamma^{(n-\ell)} \quad \text{for} \quad \ell \leq n.
$$

$\{\gamma^{(\ell)}\}_{\ell=1}^{\infty}$ is a sequence of increasingly refined partitions, i.e. $\gamma^{(\ell)} \supseteq \gamma^{(\ell-1)}$, indexed by time.

The number of elements of $\gamma^{(n)}$ is $k^n$. Each element of $\gamma^{(n)}$

$$
A(i_1, \ldots, i_n) = A^{i_1} \cap \cdots \cap T^{- (n - 1)} A^{i_n}
$$

$$
= A(i_1, \ldots, i_\ell) \cap T^{-\ell} A(i_{\ell+1}, \ldots, i_n)
$$

(i_\ell \in \{0, \ldots, k-1\}$ for any $\ell$) represents a sequence of outcomes of the experiments $\gamma$, $T^{-1} \gamma$, $\ldots$, $T^{-n} \gamma$. We call $A(i_1, \ldots, i_n)$ a $n$-history or history of length $n$ originating at time $t = 1$ from $A(i_1)$. Similarly $T^{-j} A(i_1, \ldots, i_n)$ is a $n$-history originating at time $t = j + 1$ from $A(i_1)$. Equation (2.4) shows that each $n$-history can be regarded as the intersection of an $\ell$-history ($\ell \leq n$) starting at time $t = 1$ with a $(n - \ell)$-history starting at $t = \ell + 1$. We shall call $A(i_1, \ldots, i_\ell)$ the common $\ell$-prehistory of all the $k^{n-\ell}$ histories that branch off from it. The $(n - \ell)$-histories are the elements of $T^{-\ell} \gamma^{(n-\ell)}$ so that:

$$
A(i_1, \ldots, i_\ell) = \bigcup_{i_{\ell+1}, \ldots, i_n} A(i_1, \ldots, i_n).
$$

Thus the collection of all the histories can be arranged in a tree, indexed by time, as shown in figure 1. Each point represents a history; those that belong to the same partition occupy the same horizontal line. Each point of $\gamma^{(n)}$ is connected downwards to a single point of $\gamma^{(n-1)}$ and upwards to $k$ points of $\gamma^{(n+1)}$ that represent the $k$ possible extensions of a $n$-history.
Fig. 1. — A tree of histories corresponding to a partition with $k = 2$. The common closest ancestor of $E$ and $F$, i.e. $P(E, G)$, belongs to the partition $\gamma^{(2)}$ and has a prehistory $P(E, F)$, which belongs to $\gamma$ and is the common closest ancestor of $E$ and $F$.

It is easy to see that the collection of all points on a horizontal line, i.e. the elements of a partition, $\gamma^{(n)}$ say, is an ultrametric set. Indeed, let us consider two events $E$ and $F$ of $\gamma^{(n)}$ and let $\gamma^{(\ell)}$ with $\ell \leq n$ be the most refined partition in the collection $\{\gamma^{(j)}\}_{j=1}^{\infty}$ such that one of its elements, denoted by $P(E, F)$, contains both $E$ and $F$. We call $P(E, F)$ the common closest ancestor of $E$ and $F$. One can easily verify that the non negative number.

$$D(E, F) = \frac{n - \ell}{k^\ell}.$$ \hfill (2.6)

is an acceptable definition of distance between $E$ and $F$.

In addition to the properties of a distance $D$ satisfies the more restrictive ultrametric condition

$$D(E, F) \leq \max \{D(E, G), D(G, F)\}$$ \hfill (2.7)

where $E, F, G$ are generic elements of $\gamma^{(n)}$.

The maximum distance between two elements of $\gamma^{(n)}$ is $n$ which obtains when they have no common prehistory.

The distance $(n - \ell)/k^\ell$ is the same for all the $n$-histories, $E, F \in \gamma^{(n)}$ with common closest ancestor $P(E, F) \in \gamma^{(\ell)}$. Hence

$$\sum_{G \in \gamma^{(\ell)}} D(E, F) = n - \ell$$ \hfill (2.8)

where $E, F$ are any pair in $\gamma^{(n)}$ such that $P(E, F) = G$. $D(E, F)$ is the same for all such pairs. Thus the sum is the time difference between the lengths of the histories and that of their prehistory.

3. Entropy ultrametric.

The distance (2.6) is purely dynamical and does not take into account the probabilistic nature of the system. We now want to define a new distance between histories that depends on their
probabilities and on that of their common closest ancestor. We shall remove from the tree histories with zero measure and, in this way, we include the possibility of obtaining irregular trees.

We begin by noticing that the distance between two partitions $\gamma^{(n)}$ and $\gamma^{(f)}$, with $\ell \leq n$, can be measured either by their time separation $n - \ell$ or by their entropy (information) distance which is given by (2):

$$d_m(\gamma^{(n)}, \gamma^{(f)}) = H(\gamma^{(n)}, m) - H(\gamma^{(f)}, m)$$

(3.1)

where $H(\gamma^{(i)}, m)$ is the entropy of the partition $\gamma^{(i)}$ with respect to the probability measure $m$. If we keep $\ell$ fixed $d_m(\gamma^{(n)}, \gamma^{(f)})$ increases with $n$, not linearly as $n - \ell$ but in a way that depends upon the rate of information produced by the mapping $T$ on the partition $\gamma$.

Let $E$ and $F$ be two histories of $\gamma^{(n)}$ with common closest ancestor $P (E, F) \in \gamma^{(f)}$. We can define their distance by

$$D_m(E, F) = m(P(E, F)) H(\gamma^{(n)}, m_P(E, F))$$

(3.2)

where for any $G \in \mathcal{F}$

$$m_p(G) = \frac{m(P \cap G)}{m(P)}$$

(3.3)

($m_p$ is the measure conditioned by $P$). Obviously only those elements of $\gamma^{(n)}$ that belong to the subtree arising from $P (E, F)$ contribute to $H(\gamma^{(n)}, m_p)$. Since the entropy of the partition $\gamma^{(f)}$ in the measure $m_p$, $P \in \gamma^{(f)}$, vanishes we can write the distance (3.2) in the form

$$D_m(E, F) = m(P) d_m(\gamma^{(n)}, \gamma^{(f)})$$

(3.4)

where $d_m$ is defined in (3.1). The distance $D_m(E, F)$ represents the uncertainty about which element of $\gamma^{(n)}$ the transformation $T$ generates from the element $P$ during the time interval $\Delta t = n - \ell$. In particular, if all the histories arising from $P$ (i.e. all the $n$-histories having a common $\ell$-prehistory) have the same probability, $1/\nu$ say, then

$$D_m(E, F) = m(P) \log \nu .$$

(3.5)

The distance increases with the number of descendents of $P$.

We shall now prove the following:

**Proposition 1.** Let $\gamma^{(n)} = \gamma \lor T^{-1} \gamma \lor \cdots \lor T^{-(n-1)} \gamma$ be a finite partition of the dynamical system $(X, \mathcal{F}, m, T)$. The distance $D_m$ defined by (3.2) between elements of $\gamma^{(n)}$ with non zero probability is ultrametric.

**Proof:** $D_m(E, F)$ is obviously symmetric in $E$ and $F$ and non negative. It vanishes if and only if $E = F = P (E, F)$ because $E$ and $F$ have non zero measure. Let $E$, $F$ and $G$ be three histories in the partition $\gamma^{(n)}$ and suppose that $P (E, G)$, the common closest ancestor of $E$ and $G$, be contained in $P (E, F)$. Then $P (G, F) = P (E, F)$. If we set $L = P (E, F) \setminus P (E, G)$, then the probability measure $m_P(E, F)$ can be decomposed as the convex combination

$$m_P(E, F) = m_P(E, F)(P(E, G)) m_P(E, G) + m_P(E, F)(L) m_L .$$

(3.6)

(2) See for example [5-7] ; and references therein ; for a recent application see [8].
Owing to the concavity of the entropy we have
\[
H(\gamma^{(n)}, m_{P(E,F)}) = \frac{m_{P(E,F)}(P(E,G))}{m(P(E,F))} H(\gamma^{(n)}, m_{P(E,G)}) \\
+ m_{P(E,F)}(L) H(\gamma^{(n)}, m_L) \\
\geq \frac{m(P(E,G))}{m(P(E,F))} H(\gamma^{(n)}, m_{P(E,G)})
\] (3.7)
and therefore
\[
m(P(E,F)) H(\gamma^{(n)}, m_{P(E,F)}) = m(P(G,F)) H(\gamma^{(n)}, m_{P(G,F)}) \\
\geq m(P(E,G)) H(\gamma^{(n)}, m_{P(E,G)}).
\] (3.8)

From this it follows
\[
D_m(E, F) = D_m(G, F) \geq D_m(E, G).
\] (3.9)
If instead \( P(E,G) \geq P(E,F) \) we obtain by a similar argument
\[
D_m(E, F) \leq D_m(G, F) = D_m(E, G).
\] (3.10)
From these two inequalities it follows
\[
D_m(E, F) \leq \max \{D_m(E, G), D_m(G, F)\}
\] (3.11)
which proves the statement.

Q.E.D.

If \( \mu \) is the uniform measure, i.e. \( \mu(G) = k^{-j} \) for any history \( G \in \gamma^{(j)} \), then
\[
D_\mu(E, F) = D(E, F) \log k
\] (3.12)
where \( D(E, F) \) is defined in (2.6).
The sum over \( G \) of the distances \( D_m(E, F) \) such that \( E, F \) is any pair in \( \gamma^{(n)} \) with ancestor \( G = P(E, F) \in \gamma^{(l)} \) is
\[
\sum_{G \in \gamma^{(l)}} D_m(E, F) = \sum_{G \in \gamma^{(l)}} m(G) H(\gamma^{(n)}, m_G) = d_m(\gamma^{(n)}, \gamma^{(l)}).
\] (3.13)

This relation for the sum of the entropy distances \( D_m \) generalises equation (2.8).

Remarks:

(i) the above results are valid also for partitions \( \gamma \) with a countable infinity of elements, provided that for each \( n \) the entropy of the partition \( \gamma^{(n)} \) be finite, because this condition ensures that all the distances are well defined. In this case the number of branches at each event in the tree is no longer limited;

(ii) the maximum distance between two histories of \( \gamma^{(n)} \) is \( H(\gamma^{(n)}, m) \), which is the mean information necessary to reconstruct the tree. The difference
\[
s(n) = H(\gamma^{(n+1)}, m) - H(\gamma^{(n)}, m) = d_m(\gamma^{(n+1)}, \gamma^{(n)})
\] (3.14)
coincides with the silhouette slope \( s(n) \) introduced by Bachas and Huberman [9] for the special case of the uniform measure \( \mu \);
(iii) if \( m \) preserved by \( T \) the limit \( \lim_{n \to \infty} s(n) \) exists (it is called in Ref. [9] the tree’s silhouette) and coincides with the rate of entropy production [5-7] by \( T \) on \( \gamma \) defined by

\[
h(\gamma, T, m) = \lim_{n \to \infty} \frac{1}{n} H(\gamma(n), m) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} d_m(\gamma(j), \gamma(j+1)). \tag{3.15}
\]

If \( m \) is not preserved by \( T \) neither the existence of the rate of entropy production, nor, a fortiori, that of the tree’s silhouette can be proved.

4. \( \alpha \)-distances and \( \alpha \)-overlaps.

In the definition (3.2) of the ultrametric distance \( D_m(E, F) \) we have used the von Neumann-Shannon entropy of the partition \( \gamma^{(n)} \) to which \( E \) and \( F \) belong relative to the measure \( m \), namely:

\[
H(\gamma^{(n)}, m) = - \sum_{E \in \gamma^{(n)}} m(E) \log m(E). \tag{4.1}
\]

As is well known \( H(\gamma^{(n)}, m) \) is the limit for \( \alpha \to 1 \) of the information measures of degree \( \alpha \) defined by

\[
H_\alpha(\gamma^{(n)}, m) = \frac{1}{1 - \alpha} \log \sum_{E \in \gamma^{(n)}} [m(E)]^\alpha \tag{4.2}
\]

(see for example [4]; the case \( \alpha = 2 \) was first considered in [10]). For any positive real \( \alpha \), \( H_\alpha \) is a positive function whose value increases with the refinement of the partition, and which takes values in the interval \([0, n \log k]\). It can be verified that for \( \alpha \in [0, 1] \), \( H_\alpha \) is a concave function on the space of probability measures \( m \). Therefore, by the same argument used in the proof of Proposition 1 we can prove.

**Proposition 2.** Let \( \gamma^{(n)} = \gamma \circ T^{-1} \circ \cdots \circ T^{-(n-1)} \gamma \) be a finite partition of the dynamical system \((X, \mathcal{F}, m, T)\). The distance \( D_m^{(\alpha)} \) between the elements of \( \gamma^{(n)} \) with nonzero probability defined by

\[
D_m^{(\alpha)}(E, F) = m(P(E, F)) H_\alpha(\gamma^{(n)}, m_{P(E, F)}) \tag{4.3}
\]

is ultrametric for every choice of \( \alpha \) in the interval \([0, 1]\).

We shall call \( D_m^{(\alpha)} \) a \( \alpha \)-distance.

To each ultrametric distance we can associate a \( \alpha \)-overlap by the relation

\[
Q_m^{(\alpha)}(E, F) = \exp(- D_m^{(\alpha)}(E, F)) \tag{4.4}
\]

Clearly \( Q_m^{(\alpha)}(E, E) = 1 \). Moreover from the ultrametricity of \( D_m^{(\alpha)} \) it follows

\[
Q_m^{(\alpha)}(E, F) \geq \min \{ Q_m^{(\alpha)}(E, G), Q_m^{(\alpha)}(G, F) \}. \tag{4.5}
\]

Let

\[
Y_n^{(\alpha)} = \exp(- H_\alpha(\gamma^{(n)}, m)). \tag{4.6}
\]
Since the maximum $\alpha$-distance from two histories of $\gamma^{(n)}$ is $H_\alpha(\gamma^{(n)}, m)$ it follows that the minimum value of the $\alpha$-overlap on the partition $\gamma^{(n)}$ is
\[ Y_n^{(\alpha)} = \min_{E, F \in \gamma^{(n)}} Q_m^{(\alpha)}(E, F). \quad (4.7) \]

In terms of $Y_n^{(\alpha)}$, the $\alpha$-silhouette slope at time $t = n$, defined (see (3.14)) as the difference of the $\alpha$-entropies of $\gamma^{(n+1)}$ and $\gamma^{(n)}$, reads
\[ s_\alpha(n) = H_\alpha(\gamma^{(n+1)}, m) - H_\alpha(\gamma^{(n)}, m) = \log \frac{Y_n^{(\alpha)}}{Y_{n+1}^{(\alpha)}}. \quad (4.8) \]

The limit, if it exists, of $s_\alpha(n)$ for $n \to \infty$, is the $\alpha$-silhouette of the tree (the usual proof of the convergence of $s_1(n)$ fails when $\alpha \neq 1$ because $\alpha$-entropies are not in general subadditive). If $m$ is a measure for which $T$ is a Bernoulli shift we have
\[ Y_n^{(\alpha)} = \sum_{E \in \gamma^{(n)}} m(E)^\alpha = (Y_1^{(\alpha)})^n. \quad (4.9) \]

In this case the $\alpha$-silhouette is
\[ s_\alpha = \lim_{n \to \infty} s_\alpha(n) = - \log Y_1^{(\alpha)}. \quad (4.10) \]

The quantity $Y_n^{(\alpha)}$ is defined by (4.6) for any positive value of $\alpha$, also for $\alpha > 1$ when no entropic distance is defined. As $H_\alpha$ increases with refinement (and therefore with time) $Y_n^{(\alpha)}$ decreases with $n$, i.e.
\[ 0 \leq Y_n^{(\alpha)} \leq Y_{n-1}^{(\alpha)} \leq 1. \quad (4.11) \]

In the case $\alpha = 2$
\[ Y_n = Y_n^{(2)} = \sum_{E \in \gamma^{(n)}} [m(E)]^2. \quad (4.12) \]
represents the probability that two histories at time $t > n$ have a common $n$-prehistory.

5. Disordered systems.

The ultrametric structure of the partition $\gamma^{(n)}$ depends upon the existence of a sequence of partitions $\{ \gamma^{(i)} \}$ with $i = 0, \ldots, n, \ldots$ such that $\gamma^{(i)}$ is more refined than $\gamma^{(i-1)}$.

In the theory of disordered systems there also appears an ultrametric structure [11] which however has its origin in the metric properties of the phase space and which can only be defined in a probabilistic sense.

Consider a probability space $(X, \mathcal{F}, m)$ where the phase space $X$ is assumed to be equipped with a metric $g(x, y)$, $x, y \in X$, and where $m$ is an equilibrium measure, that is a convex combination of ergodic measures [12] $\rho_i$:
\[ m = \sum_{i \in \mathbb{I}} a^i \rho_i \quad (5.1) \]

with
\[ 0 \leq a^i \leq 1, \quad \sum_{i \in \mathbb{I}} a^i = 1. \quad (5.2) \]
Let $E^j$ be the non empty subset invariant for the ergodic state $\rho_i$, i.e. such that $\rho_i(E^j) = 1$. Then for any ergodic state $\rho_j$ we have

$$\rho_j(E^j) = \delta^i_j$$

(5.3)

because for $i \neq j$

$$E^j \cap E^i = \emptyset.$$  

(5.4)

Therefore

$$m(E^i) = a^i.$$  

(5.5)

If $Z$ is the subspace of $X$ with zero measure for all the equilibrium states $m$, the set $\eta = \{E^i\}_{i \in I}$ is a partition of the space $X \setminus Z$ of equilibrium configurations.

The partition $\eta$ is non trivial only when there is more than one ergodic state, that is when the system undergoes a phase transition. In the mean field solution of spin glasses there is an infinity of ergodic states not related to each other by a symmetry (this is a consequence of what is called frustration [13]).

The entropy of such a partition

$$H(\eta, m) = - \sum_{i \in I} a^i \log a^i = - \sum_{E^i \in \eta} m(E^i) \log m(E^i)$$  

(5.6)

has been considered previously by Palmer [14].

From the metric $g(x, y)$ in $X$ we can define the Hausdorff distance between the elements of $\eta$:

$$d_{ij} = d(E^i, E^j) = \max \left\{ \sup_{x \in E^i} \inf_{y \in E^j} g(x, y), \sup_{y \in E^j} \inf_{x \in E^i} g(x, y) \right\}.$$  

(5.7)

Let $\{d_0 > d_1 > \cdots > d_n > \cdots > 0\}$ be a sequence of different decreasing distances between the elements of $\eta$ and let us consider a sequence of partitions

$$\{\{X\}, \emptyset\} = \eta_0 < \eta_1 < \cdots < \eta_n < \cdots$$

ordered according to refinement, whose elements are subsets of $\eta$ with decreasing diameters, respectively $d_0$, $d_1$, ..., $d_n$, ..., 0, i.e.

$$\eta_n = \left\{ E^i_r = \bigcup_{i \in I_r} E^i_{n+1} \mid \text{diam } E^i_r \leq d_n, \text{ diam } (E^i_r \cup E^i_{s}) > d_n \text{ when } r \neq s \right\}$$

(5.8)

where

$$\text{diam } E^i_r = \max_{E^i, E^j \in E^i_r} d(E^i, E^j).$$

(5.9)

In reference [2] the elements $E^i_r$ are referred to as clusters.

The sequence $\{\eta_n\}$ is not uniquely determined by the sequence $\{d_n\}$ as several subsets $E^i_n$ with the same diameter can in general exist. Moreover it should be remarked that the distance $d(E^i, E^j)$ between two elements $E^i \in E^i_n$ and $E^j \in E^j_n$ may well be smaller than $d_n$. For example, consider four points at the corners of a square. If $d$ is equal to the side of the square there are two ways of partitioning the set of the four corners into sets of diameter $d$, either by putting together the pair of upper corners and that of the lower ones, or by putting
Fig. 2. — The two ways in which the corners of a square can be united to form subsets with a diameter equal to the side of the square.

together the left and the right pairs. Clearly in both cases there are pairs of points belonging to different subsets which have a distance \( d \) (see Fig. 2).

The situation is much simpler when \( \eta \) is ultrametric for the distances \( d_{ij} \). In this case it is easy to verify that the sequence \( \{ \eta_n \} \) is determined uniquely. Furthermore ultrametricity implies that the distance between two elements \( E^i \in E'_n \) and \( E^j \in E^s_n \) is always greater than \( d_n \) and does not depend upon \( i, j \). In this case we have

\[
d(E'_n, E^s_n) = d(E^i, E^j) \quad \text{for} \quad E^i \in E'_n, \ E^j \in E^s_n. \tag{5.10}
\]

Thus if \( \eta \) is ultrametric each \( \eta_n \) is also ultrametric.

Remark that, if \( E'_{n+1} \subset E'_n \) and \( E^s_{n+1} \subset E^s_n \), we get for the entropic distance that we introduced in section 3

\[
D_m(E'_{n+1}, E^s_{n+1}) - D_m(E'_n, E^s_n) = [H(\eta^{(n+1)}, m_P) - H(\eta^{(n)}, m_P)] m(P) \geq 0 \tag{5.11}
\]

where \( P \) is the common ancestor of \( E'_{n+1} \) and \( E^s_{n+1} \), that is we find the silhouette of the subtree which originates from \( P \).

Instead of ordering the partitions \( \{ \eta_n \} \) according to decreasing diameters it is customary in the theory of disordered systems to order them according to increasing maximum overlaps of the subsets within their elements : \( 0 = q_0 < q_1 < \ldots < q_n < \ldots < q_{\text{Max}} \). The exact relation between distances and overlaps is not important. It is enough to assume that the overlap \( q_{ij} = q(E^i, E^j) \) between two elements \( E^i \) and \( E^j \) of \( \eta \) is a decreasing function of their distance.

The distance and the overlap considered in this section are of a pure geometrical nature in contrast with that discussed in section 3 (and generalized in Sect. 4 for \( \alpha \in [0, 1] \)) which depends upon the probability measure. To establish the relation between these two distances we shall introduce, following Parisi, the probability distribution of the overlaps \( q_{ij} \) in the partition \( \eta \):

\[
P(q) = \sum_{E^i, E^j \in \eta} m(E^i) m(E^j) \delta(q - q_{ij}). \tag{5.12}
\]

\( P(q) \) is a measure of the complexity of the partition \( \eta \) which Parisi has called order parameter for the glassy transition [15].

A measure of the departure from ultrametricity of the partition \( \eta_n \) is given by the quantity

\[
\Delta_n = \int_{q_n}^{q_{\text{Max}}} P(q) \, dq - Y_n \tag{5.13}
\]

where \( Y_n \) is defined as in (4.12).
Indeed, we can prove the following:

**Proposition 3.** $\Delta_n$ is non negative and it vanishes if $\eta$ is ultrametric. Conversely, if $\Delta_n$ vanishes, for all choices of $q_n$, then the elements of $\eta$ with non vanishing probability form an ultrametric set.

**Proof:**

\[
\int_{q_n^{\text{Max}}} P(q) \, dq = \sum_{E^i, E^j \in \eta \atop q_{ij} \geq q_n} m(E^i) \, m(E^j) = Y_n + \sum_{E^i_n, E^j_n \in \eta_n \atop E^i \subset E^j_n} \sum_{E^k} \sum_{q_{ij} \geq q_n} m(E^i) \, m(E^j) \quad (5.14)
\]

where the first term arises from the case $r = s$. Therefore

\[
\Delta_n = \sum_{E^i_n, E^j_n \in \eta_n \atop E^i \subset E^j_n \atop E^k} \sum_{r \neq s} \sum_{q_{ij} \geq q_n} m(E^i) \, m(E^j) \geq 0 \quad (5.15)
\]

If $\eta$ is ultrametric the sum vanishes because elements $E^i, E^j$ with overlaps greater than $q_n$ are contained in the same element of $\eta_n$. This concludes the first part of the proposition.

To prove the second statement suppose that the elements of $\eta$ with non vanishing probability are not an ultrametric set. Then there exists at least three elements $E^i, E^j, E^k$ with non vanishing probability for which $q_{ij} > \min (q_{ik}, q_{kj}) = q_{jk}$. Choose $q_n = q_{ij}$. Then the three elements $E^i, E^j, E^k$ cannot be contained in the same element of $\eta_n$, because this would have a diameter larger than $d_n$. But if $E^i$ is not contained in the same element of $\eta_n$ that contains $E^j$ or $E^k$ there is a non negative contribution to the r.h.s. of (5.15).

Q.E.D.

In the theory of spin glasses ultrametricity occurs only after the average over disorder. Indeed, disorder is together with frustration a necessary ingredient of the theory of spin glasses, which is taken into account by a probability distribution of couplings between spins. In our description disorder isinstead represented by a probability distribution for the coefficients $a_i$'s. Usually the average over the equilibrium measures is denoted by a bar ($\bar{\cdots}$) and one can consider the averaged versions of the previous relations.

The statement of ultrametricity in the mean field theory of spin glasses is expressed by the weaker relation

\[
\bar{\Delta}_n = 0 \quad (5.16)
\]

which means that the probability of finding three elements of $\eta_n$ with non zero measure that do not satisfy the ultrametricity inequality is zero.

Specifically, let $f_n(W)$ be the probability that an event of the partition $\eta^{(n)}$ has probability $W$, that is

\[
f_n(W) = \sum_{E \in \eta^{(n)}} \delta(W - m(E)) \quad (5.17)
\]

(3) For a discussion of the definition of such average, see for example [16].
$Wf_n(W)$ is the probability for the weight $W$ and we get as a consequence of (5.16)

$$Y_n = \int_0^1 W f_n(W) \, dW = \int_{q_n}^{q_{\text{Max}}} \frac{P(q)}{q_n} \, dq .$$

(5.18)

Similarly the average of the von Neumann-Shannon entropy is given by

$$H(\eta^{(n)}, m) = - \int_0^1 W \log W \, f_n(W) \, dW .$$

(5.19)

For example, a Poisson distribution of exponential weights [16], which is typical of the random energy model [17], will give rise to

$$f_n(W) = \frac{W^{-x_n-1}[1 - W]^{x_n-1}}{\Gamma(1 - x_n) \Gamma(x_n)}$$

(5.20)

with $x_n$ a parameter characterizing the partition. Such a distribution also applies to spin glasses [11, 18] and can be interpreted as a distribution of random free energies.

It follows that

$$Y_n = 1 - x_n$$

(5.21)

where as a consequence of equation (4.11) $0 \leq x_{n-1} \leq x_n \leq 1$. The average value of the von Neumann-Shannon entropy is instead.

$$H(\eta^{(n)}, m) = \psi(1) - \psi(1 - x_n)$$

(5.22)

where

$$\psi(z) = \frac{d}{dz} \log \Gamma(z)$$

(5.23)

is increasing in the interval $(0, 1)$ and diverges at minus infinity when $z$ approaches zero. All the entropy distances we have introduced are well defined when $q < q_{\text{Max}}$ or equivalently $x_n < 1$.

Equations (5.18), (5.21) and (5.22) provide the connection between the order parameter and the average value of the complexity of the system at the scale $q_n$.

Acknowledgment.

One of us (L. A. R.) wishes to thank the hospitality at the Institute des Hautes Études Scientifiques, Bures-sur-Yvette, where part of this work has been done.

References

