l∞-Approximation via Subdominants

Article *in* Journal of Mathematical Psychology · December 2000

DOI: 10.1006/jmps.1999.1270 · Source: PubMed

Victor Chepoi
Laboratoire d'Informatique Fondamentale de …

Bernard Fichet
Aix-Marseille Université

Journal of Mathematical Psychology **44**, 600–616 (2000)
doi:10.1006/jmps.1999.1270, available online at http://www.idealibrary.com on IDEAL®

l_∞-Approximation via Subdominants

Victor Chepoi and Bernard Fichet

Université d'Aix Marseille II

Given a vector $\mathbf{u}$ and a certain subset K of a real vector space $\mathbf{E}$, the problem of l_∞-approximation involves determining an element $\hat{\mathbf{u}}$ in K nearest to $\mathbf{u}$ in the sense of the l_∞-error norm. The subdominant $\mathbf{u}_*$ of $\mathbf{u}$ is the upper bound (if it exists) of the set $\{\mathbf{x} \in K : \mathbf{x} \prec \mathbf{u}\}$ (we let $\mathbf{x} \prec \mathbf{y}$ if all coordinates of $\mathbf{x}$ are smaller than or equal to the corresponding coordinates of $\mathbf{y}$). We present general conditions on K under which a simple relationship between the subdominant of $\mathbf{u}$ and a best l_∞-approximation holds. We specify this result by taking as K the cone of isotonic functions defined on a poset $(X, \prec)$, the cone of convex functions defined on a subset of $\mathbb{R}^N$, the cone of ultrametrics on a set X, and the cone of tree metrics on a set X with fixed distances to a given vertex. This leads to simple optimal algorithms for the problem of best l_∞-fitting of distances by ultrametrics and by tree metrics preserving the distances to a fixed vertex (the latter provides a 3-approximation algorithm for the problem of fitting a distance by a tree metric). This simplifies the recent results of Farach, Kannan, and Warnow (1995) and of Agarwala *et al.* (1996).

1. INTRODUCTION

The basic approximation problem can be formulated as follows: given a vector $\mathbf{u}$ and a certain subset K of a linear space, fit to $\mathbf{u}$ a best vector $\hat{\mathbf{u}} \in K$. The error norms usually chosen are l_1, l_2, or l_∞. Usually K is a certain cone of functions (i.e., if $f \in K$ then $\alpha f \in K$ for every $\alpha \in \mathbb{R}$) of polynomials of given degree, of convex functions, of dissimilarities, etc.

To give an instructive example, consider the isotonic regression problem. Given experimental values $u_1, ..., u_n$ of the dependent variable $\mathbf{u}$, corresponding to values $x_1, ..., x_n$ of the independent variable x, which constitute a set X with a partial order $\prec$, fit to the u_i a best function $\mathbf{u} = \mathbf{f}(x)$ which is nondecreasing (alias isotonic) with respect to $\prec$. Then K is the (convex) cone of all isotonic functions $\mathbf{f}$ defined on the poset $(X, \prec)$ and $\mathbf{u}$ is the vector of numerical data on X; see Barlow, Bartholomew, Bremner, and Brunk (1972) and Robertson, Wright, and Dykstra (1988). Algorithms for this problem under the norm l_2 have received a great deal of attention;

Address correspondence and reprint requests to Bernard Fichet, Laboratoire de Biomathématiques, Université d'Aix Marseille II, 27 Bd Jean Moulin, F-13385 Marseille Cédex 5, France.

0022-2496/00 $35.00

however, polynomial time algorithms are known only when X is a linear order (Barlow *et al.*, 1972) or a rooted tree (Barlow *et al.*, 1972, Chepoi, Cogneau, and Fichet, 1997). The isotonic regression problem with the l_∞-norm has been considered in Chepoi *et al.* (1997) and Ubhaya (1974), and a strikingly simple optimal estimate has been given.

In numerical taxonomy, $\mathbf{u}$ is a distance (more generally, a dissimilarity) on a finite set X and K is the cone of all ultrametrics or tree metrics defined on X; see Barthélemy and Guénoche (1991) and Sneath and Sokal (1973). Fitting distances by trees and ultrametrics is an important problem of data analysis, mathematical psychology, and evolutionary biology. For example, one of the main purposes of analyzing proximity data in psychology is to infer the mental organization of objects. Both continuous spatial models and discrete models have been proposed. An important discrete model is the additive tree model, which represents objects by vertices of a tree and defines dissimilarities by path lengths in this tree (see the papers by Carroll (1976), Carroll and Pruzansky (1980), Cunningham (1978), De Soete (1983), and Sattath and Tversky (1977) for further information and some algorithms for fitting additive trees to proximity data). The ultrametric fitting problem is one of the basic problems of hierarchical clustering of data, while the phylogeny reconstruction problem is closely related to the tree fitting problem (Barthélemy and Guénoche, 1991; Swofford and Olsen, 1990). The ultrametric and the tree fitting problems under l_p-norms first were formulated by Cavalli-Sforza and Edwards (1967). Day (1987) and Křivánek and Morávek (1986) have shown that for norms l_1 and l_2 both problems are NP-hard. Recently, Agarwala *et al.* (1996) established that the tree fitting problem under the l_∞-norm is NP-hard, too. Křivánek (1988) proposed a polynomial algorithm for the ultrametric l_∞-fitting problem. An optimal algorithm for this problem has been developed by Farach, Kannan, and Warnow (1995). Using this result, Agarwala *et al.* (1996) presented a polynomial 3-approximation algorithm for the tree l_∞-fitting problem (i.e., in the worst case the l_∞-error of a tree found by the algorithm is at most three times larger than the l_∞-error of an optimal tree). Most recently, Cohen and Farach (1997) showed that the algorithm from Agarwala *et al.* (1996) and its modification outperform the neighbor joining heuristic of Saitou and Nei (1987).

In this paper we present some general conditions on the set K under which a solution of the l_∞-approximation problem and its relatives is closely related to the subdominant of a vector $\mathbf{u}$ in K. In particular, this significantly simplifies the algorithms presented in Agarwala *et al.* (1996), Farach *et al.* (1995), and Křivánek (1988). To illustrate the algorithms we consider the ultrametric and tree-approximations of the distance matrix between 15 texts obtained by Bartlett (1932) with the method of repeated reproduction.

2. GENERAL RESULTS

Let $\mathbf{E}$ be a finite-dimensional vector space with dimension p. According to a fixed basis of $\mathbf{E}$, a vector $\mathbf{u}$ in $\mathbf{E}$ has coordinates $(u_1, ..., u_p)$. Let $\mathbf{1}$ be the vector with coordinates $(1, ..., 1)$. The set of all $c\mathbf{1}$, $c \in \mathbb{R}$, is the *diagonal* $\mathscr{L}$ of the positive orthant of $\mathbf{E}$. In the following, $\mathbf{E}$ is endowed with the l_∞-norm (i.e., the well-known uniform or Chebychev norm), simply noted $\|\cdot\|$: for a given vector $\mathbf{u}$, $\|\mathbf{u}\| = \max_i |u_i|$. For two vectors $\mathbf{u}$ and $\mathbf{v}$, $\|\mathbf{u}-\mathbf{v}\|$ is the l_∞-distance between $\mathbf{u}$ and $\mathbf{v}$. Given two subsets A and B of $\mathbf{E}$ let

$$\delta(A, B) = \inf_{\mathbf{v} \in B} \sup_{\mathbf{u} \in A} \|\mathbf{u}-\mathbf{v}\|.$$

Note that, in general, $\delta(A, B) \neq \delta(B, A)$. For a vector $\mathbf{x}$, $\delta(\mathbf{x}, B)$ is the l_∞-distance from $\mathbf{x}$ to the set B.

According to the given basis, define a partial order $\prec$ on $\mathbf{E}$ (called the *pointwise order*) by letting $\mathbf{x} \prec \mathbf{y}$ if and only if $x_i \leqslant y_i$ for all $i = 1, ..., p$. Then, every nonempty bounded subset A of $\mathbf{E}$ has a least upper bound (join) and a greatest lower bound (meet), with i-coordinates $\sup\{u_i : \mathbf{u} \in A\}$ and $\inf\{u_i : \mathbf{u} \in A\}$, respectively. In other words, $(\mathbf{E}, \prec)$ is a (conditionally) complete lattice.

Now, we formulate our basic l_∞-approximation problems. Let K be a nonempty subset of $\mathbf{E}$. It will be the approximating reference set. The main problem under consideration in this paper is:

(P_1) given $\mathbf{u} \in \mathbf{E}$, find $\hat{\mathbf{u}} \in K$, if one exists, such that

$$\|\mathbf{u}-\hat{\mathbf{u}}\| = \delta(\mathbf{u}, K).$$

We will also take an interest in approximating $\mathbf{u}$ by an element of K less than or greater than $\mathbf{u}$. To this end, we use the following notations:

$$K_\prec(\mathbf{u}) = \{\mathbf{z} \in K : \mathbf{z} \prec \mathbf{u}\}, \qquad K_\succ(\mathbf{u}) = \{\mathbf{z} \in K : \mathbf{z} \succ \mathbf{u}\}.$$

Then the following problems are introduced, provided the reference sets are nonempty.

(P'_1) given $\mathbf{u} \in \mathbf{E}$, find $\mathbf{y} \in K_\prec(\mathbf{u})$, if one exists, such that

$$\|\mathbf{u}-\mathbf{y}\| = \delta(\mathbf{u}, K_\prec(\mathbf{u}))$$

and

(P''_1) given $\mathbf{u} \in \mathbf{E}$, find $\mathbf{y} \in K_\succ(\mathbf{u})$, if one exists, such that

$$\|\mathbf{u}-\mathbf{y}\| = \delta(\mathbf{u}, K_\succ(\mathbf{u})).$$

A solution of (P'_1) can be obtained by solving another problem:

when $K_\prec(\mathbf{u})$ is nonempty, does this set admit a maximum element?

If the answer is yes, such an element $\mathbf{u}_*$ is called the *lower maximum approximation*, or the *subdominant* of the vector $\mathbf{u}$, and $\mathbf{u}_*$ yields a solution, actually the greatest solution, of (P'_1). A similar problem can be formulated for the minimum element of the set $K_\succ(\mathbf{u})$. If it exists, it is denoted by $\mathbf{u}^*$ and is called the *upper minimum approximation* of $\mathbf{u}$.

We formulate the last basic problems, which extend (P_1), (P_1'), and (P_1''): given two vectors $\mathbf{u}$, $\mathbf{v}$, with $\mathbf{u} \prec \mathbf{v}$

(P_2) find $\mathbf{y} \in K$, if one exists, such that

$$\delta(\{\mathbf{u}, \mathbf{v}\}, \mathbf{y}) = \delta(\{\mathbf{u}, \mathbf{v}\}, K);$$

(P_2') find $\mathbf{y} \in K_{\prec}(\mathbf{u})$, if one exists, such that

$$\|\mathbf{v} - \mathbf{y}\| = \delta(\mathbf{v}, K_{\prec}(\mathbf{u}));$$

(P_2'') find $\mathbf{y} \in K_{\succ}(\mathbf{v})$, if one exists, such that

$$\|\mathbf{u} - \mathbf{y}\| = \delta(\mathbf{u}, K_{\succ}(\mathbf{v})).$$

Due to the specificity of the l_∞-norm, some further general problems can be reduced to (P_2). First, assume that A is a bounded subset of $\mathbf{E}$ and we wish to find a vector $\mathbf{y} \in K$, if one exists, such that $\delta(A, \mathbf{y}) = \delta(A, K)$. Define two vectors $\mathbf{u}$ and $\mathbf{v}$ with coordinates $u_i = \inf\{x_i\colon \mathbf{x} \in A\}$ and $v_i = \sup\{x_i\colon \mathbf{x} \in A\}$. One can easily show that a vector $\mathbf{y}$ is a solution of the problem (P_2) if and only if $\delta(A, \mathbf{y}) = \delta(A, K)$. The same reduction acts in the case when the vectors of the set A are defined only partially, i.e., only a few coordinates of each vector are available. (This situation occurs in the isotonic regression problem: the data are given by samples of nonequal size from a set of distributions.) Then u_i and v_i are computed by taking the meet (alias infimum) and the join (alias supremum) of well-defined ith coordinates of $\mathbf{x} \in A$. It remains to replace K by its projection on the coordinate subspace generated by coordinate directions where both $\mathbf{u}$ and $\mathbf{v}$ are defined.

PROPOSITION 1. *Let K be a subset of* $\mathbf{E}$ *invariant under translations along the line $\mathscr{L}$ and let* $\mathbf{u}$, $\mathbf{v}$ *be two vectors with* $\mathbf{u} \prec \mathbf{v}$. *Then*

$$\tfrac{1}{2}\delta(\mathbf{v}, K_{\prec}(\mathbf{u})) = \delta(\{\mathbf{u}, \mathbf{v}\}, K) = \tfrac{1}{2}\delta(\mathbf{u}, K_{\succ}(\mathbf{v})) := \varepsilon.$$

Moreover, in this case the following conditions are equivalent:

(i) $\mathbf{y}$ *is a solution for problem* (P_2);

(ii) $\mathbf{y} - \varepsilon \mathbf{1}$ *is a solution of* (P_2');

(iii) $\mathbf{y} + \varepsilon \mathbf{1}$ *is a solution of* (P_2'').

Proof. Take $\mathbf{y} \in K$. Define $\mathbf{y}' = \mathbf{y} - \delta(\{\mathbf{u}, \mathbf{v}\}, \mathbf{y})\,\mathbf{1}$. By hypothesis, $\mathbf{y}' \in K$. Moreover, $\mathbf{y}' \prec \mathbf{u}$ and $\|\mathbf{v} - \mathbf{y}'\| \leqslant 2\delta(\{\mathbf{u}, \mathbf{v}\}, \mathbf{y})$. Hence $\delta(\mathbf{v}, K_{\prec}(\mathbf{u})) \leqslant 2\delta(\{\mathbf{u}, \mathbf{v}\}, \mathbf{y})$ for every $\mathbf{y}$ in K, so that $\frac{1}{2}\delta(\mathbf{v}, K_{\prec}(\mathbf{u})) \leqslant \delta(\{\mathbf{u}, \mathbf{v}\}, K)$. Conversely, let $\mathbf{x}' \in K_{\prec}(\mathbf{u})$ and define $\mathbf{x} = \mathbf{x}' + \frac{1}{2}\|\mathbf{v} - \mathbf{x}'\|\,\mathbf{1}$. By hypothesis, $\mathbf{x} \in K$ and it is easy to check that $\delta(\{\mathbf{u}, \mathbf{v}\}, \mathbf{x}) \leqslant \frac{1}{2}\|\mathbf{v} - \mathbf{x}'\|$. Therefore, $\delta(\{\mathbf{u}, \mathbf{v}\}, K) \leqslant \frac{1}{2}\|\mathbf{v} - \mathbf{x}'\|$, for every $\mathbf{x}' \in K_{\prec}(\mathbf{u})$, showing that $\delta(\{\mathbf{u}, \mathbf{v}\}, K) \leqslant \frac{1}{2}\delta(\mathbf{v}, K_{\prec}(u))$. Thus the equality $\frac{1}{2}\delta(\mathbf{v}, K_{\prec}(\mathbf{u})) = \delta(\{\mathbf{u}, \mathbf{v}\}, K) := \varepsilon$ holds. Moreover, the previous results show that if $\mathbf{y}$ is a solution of (P_2), then $\|\mathbf{v} - \mathbf{y}'\| \leqslant 2\varepsilon$, whence $\mathbf{y}' = \mathbf{y} - \varepsilon \mathbf{1}$ is a solution of (P_2'). Conversely, if $\mathbf{x}'$ is a solution of (P_2'), then $\delta(\{\mathbf{u}, \mathbf{v}\}, \mathbf{x}) \leqslant \varepsilon$; hence $\mathbf{x} = \mathbf{x}' + \varepsilon \mathbf{1}$ is a solution of (P_2). The other statements follow by duality. ∎

Following the terminology of Rockafellar (1970), the *lineality space* L of a set K is the set of all vectors $\mathbf{y}$ such that $K+\mathbf{y}=K$. Then K can be represented as the direct sum $K=L+(K\cap L^{\perp})$, where $L^{\perp}$ is the orthogonal complement of L (with respect to the usual scalar product in the selected basis of $\mathbf{E}$). Therefore, each set K satisfying the condition of Proposition 1 can be written as the direct sum of the diagonal $\mathscr{L}$ and the intersection of K with the orthogonal complement of $\mathscr{L}$. Geometrically, the sets K occurring in Proposition 1 have a cylindrical shape, and the basis $K\cap\mathscr{L}^{\perp}$ of such a cylinder can be an arbitrary subset of the orthogonal complement of $\mathscr{L}$.

Proposition 1 establishes a relationship between the optimal errors of the problems (P_2), (P_2'), and (P_2''), but does not tell whether optimal solutions exist and how to find them. It does not seem possible to achieve much more in the conditions of Proposition 1. Thus, we descend to a smaller, but quite natural class of sets. Obviously, the optimality sets are nonempty if the reference set K is closed (in the topology of $\mathbf{E}$). Additionally, if y_1 and y_2 are two solutions of (P_2), with $y_1\prec y_2$, then one can easily show that every $y\in K$ obeying $y_1\prec y\prec y_2$ is a solution. If K is convex, then the optimality sets are convex, too. A subset K of $\mathbf{E}$ is called *join-closed* (respectively, *meet-closed*) if for any bounded subset $A\subseteq K$ the join (i.e., the least upper bound) of A (respectively, the meet of A) belongs to the set K; see Birkhoff (1967). In other words, K is *join*-closed if K is a complete join subsemilattice of $\mathbf{E}$ and K is *meet*-closed if K is a complete meet subsemilattice of $\mathbf{E}$. One can easily show that a subset K of $\mathbf{E}$ is *join*-closed if and only if for any vector $\mathbf{u}$ with $K_{\prec}(\mathbf{u})\neq\emptyset$ there exists the lower maximum approximation $\mathbf{u}_*$ in K. Indeed, $\mathbf{u}_*$ will coincide with the join of the nonempty bounded set $K_{\prec}(\mathbf{u})$. Conversely, consider a bounded set $A\subseteq K$ and take its join $\mathbf{u}$. The set $K_{\prec}(\mathbf{u})$ is nonempty, because it contains A. Therefore, $\mathbf{u}$ has the lower maximum approximation $\mathbf{u}_*\in K$. Since $\mathbf{x}\prec\mathbf{u}_*$ for any $\mathbf{x}\in A$ and $\mathbf{u}_*\prec\mathbf{u}$, we conclude that $\mathbf{u}=\mathbf{u}_*$. Dually, one can show that a subset K of $\mathbf{E}$ is *meet*-closed if and only if for any vector $\mathbf{u}$ with $K_{\succ}(\mathbf{u})\neq\emptyset$, there exists the upper minimum approximation $\mathbf{u}^*$ in K. Notice that if K is *join*-closed, then the set S of optimal solutions of (P_2) is a complete join subsemilattice of K.

Since the lower maximum approximation is a solution of problem (P_2'), from Proposition 1 we immediately obtain the following property.

COROLLARY 1. *Let K be a join-closed subset of $\mathbf{E}$ invariant under translations along the diagonal line $\mathscr{L}$. If $\mathbf{u}$, $\mathbf{v}$ are two vectors with $\mathbf{u}\prec\mathbf{v}$ and $\mathbf{u}_*$ is the lower maximum approximation of $\mathbf{u}$ in K, then $\mathbf{u}_*+\frac{1}{2}\,\|\mathbf{v}-\mathbf{u}_*\|\,\mathbf{1}$ is the greatest solution of (P_2). In particular, $\mathbf{u}_*+\frac{1}{2}\,\|\mathbf{u}-\mathbf{u}_*\|\,\mathbf{1}$ is a best l_∞-approximation of $\mathbf{u}$ in K. Additionally, $\mathbf{u}_*+\|\mathbf{v}-\mathbf{u}_*\|\,\mathbf{1}$ is a solution of (P_2'').*

By duality, we can formulate a similar property if K is *meet*-closed. Then $\mathbf{v}^*-\frac{1}{2}\,\|\mathbf{u}-\mathbf{v}^*\|\,\mathbf{1}$ will be a solution of (P_2), while $\mathbf{v}^*-\frac{1}{2}\,\|\mathbf{v}-\mathbf{v}^*\|\,\mathbf{1}$ is a best l_∞-approximation of $\mathbf{v}$ in K.

COROLLARY 2. *Let K be a convex subset of $\mathbf{E}$ invariant under translations along $\mathscr{L}$ and let $\mathbf{u}$, $\mathbf{v}$ be two vectors with $\mathbf{u}\prec\mathbf{v}$. If $\mathbf{x}$ is a solution of (P_2') and $\mathbf{y}$ is a solution of (P_2''), then $\frac{1}{2}(\mathbf{x}+\mathbf{y})$ is a solution of the problem (P_2). If additionally, K is* meet-*closed and* join-*closed, then $\frac{1}{2}(\mathbf{u}_*+\mathbf{v}^*)$ is a solution of (P_2).*

Proof. Let ε be as in Proposition 1. By this proposition the vectors $\mathbf{x}+\varepsilon\mathbf{1}$ and $\mathbf{y}-\varepsilon\mathbf{1}$ are solutions of (P_2). Since K is convex and $\frac{1}{2}(\mathbf{x}+\mathbf{y})=\frac{1}{2}[(\mathbf{x}+\varepsilon\mathbf{1})+(\mathbf{y}-\varepsilon\mathbf{1})]$, we deduce that $\frac{1}{2}(\mathbf{x}+\mathbf{y})$ is a solution of (P_2). ▌

The following *sandwich problem* is a generalization of a similar problem formulated in Farach *et al.* (1995) for ultrametrics. Given $\mathbf{u}$, $\mathbf{v}$ with $\mathbf{u}\prec\mathbf{v}$ find a vector $\mathbf{y}\in K$ (if possible) such that $\mathbf{u}\prec\mathbf{y}\prec\mathbf{v}$.

Corollary 3. *Let K be a join-closed subset of* $\mathbf{E}$. *Then the sandwich problem has a solution if and only if* $\mathbf{u}\prec\mathbf{v}_*$.

Proof. If $\mathbf{u}\prec\mathbf{v}_*$, then $\mathbf{v}_*$ is a solution of the sandwich problem. Conversely, suppose that the sandwich problem has a solution $\mathbf{y}$. Then $\mathbf{u}\prec\mathbf{y}\prec\mathbf{v}_*$ and we are done. ▌

All these rather simple properties can be of use only in the case where meets and joins can be computed efficiently. As we will show below, so it is in many important cases. Before presenting them, we formulate a *consensus problem*, whose solution can be found without such computations. Given a set A of vectors $\mathbf{w}_1, ..., \mathbf{w}_n\in K$ we wish to find a vector $\mathbf{y}\in K$ (if it exists) such that $\delta(A,\mathbf{y})=\delta(A,K)$. As we already noticed, a more general form of this problem can be reduced to (P_2). However, if K is *join*-closed and invariant under translations along $\mathscr{L}$, then the consensus problem can be solved more easily. Namely, find the meet $\mathbf{u}$ and the join $\mathbf{v}$ of the set A. Then $\mathbf{v}\in K$ and by the previous results we deduce that the vector $\mathbf{v}-\frac{1}{2}\|\mathbf{v}-\mathbf{u}\|\,\mathbf{1}$ provides a solution to the consensus problem.

Corollary 4. *Let K be a join-closed subset of* $\mathbf{E}$ *invariant under translations along $\mathscr{L}$. Then the vector $\mathbf{v}-\varepsilon\mathbf{1}$ with $\varepsilon=\frac{1}{2}\|\mathbf{v}-\mathbf{u}\|$ is a solution of the consensus problem.*

We conclude this section with two concrete applications of our remarks.

Example 1 (isotonic l_∞-regression problem). Let $X=\{x_1, ..., x_n\}$ be a set of observation points endowed with a partial order $\prec$, and assume that with each element x_i is associated a set of numbers $y_{i1}, ..., y_{ir_i}$, corresponding, for example, to a sample from the ith distribution. We are looking for an isotonic function $\mathbf{f}$ on X (i.e., $x_i\prec x_j$ implies $\mathbf{f}(x_i)\leqslant\mathbf{f}(x_j)$) that minimizes the l_∞-error

$$D_\infty(\mathbf{f})=\max_{x_i\in X}\ \max_{l=1,...,r_i}|y_{il}-\mathbf{f}(x_i)|.$$

Let $\mathscr{M}(X)$ be the convex cone of isotonic functions on X. It is easy to see that $\mathscr{M}(X)$ is a complete sublattice of the n-dimensional vector space $\mathbb{R}^X$ and is invariant under translations along the diagonal line. Therefore, we can apply the previous results.

For each element x_i define $f_i=\min\{y_{i1}, ..., y_{ir_i}\}$ and $g_i=\max\{y_{i1}, ..., y_{ir_i}\}$. Let $\mathbf{f}=(f_1, ..., f_n)$ and $\mathbf{g}=(g_1, ..., g_n)$. Obviously $\mathbf{f}\prec\mathbf{g}$. The isotonic regression problem reduces to the problem (P_2) with the vectors $\mathbf{f}$, $\mathbf{g}$ and the set $\mathscr{M}(X)$. To solve it we find the lower maximum approximation $\mathbf{f}_*=(f^1_*, ..., f^n_*)$ of $\mathbf{f}$ and the upper minimum approximation $\mathbf{g}^*=(g^*_1, ..., g^*_n)$ of $\mathbf{g}$ in $\mathscr{M}(X)$. One can easily show that for each element x_i we have $f_*^i = \min\{f_j: x_i \prec x_j\}$ and $g_i^* = \max\{g_j: x_j \prec x_i\}$. From Corollaries 1 and 2 we obtain three solutions of (P_2) and of the isotonic regression problem: $\mathbf{f}_* + \frac{1}{2} \|\mathbf{g} - \mathbf{f}_*\| \mathbf{1}$, $\mathbf{g}^* - \frac{1}{2} \|\mathbf{g}^* - \mathbf{f}\| \mathbf{1}$, and $\frac{1}{2}(\mathbf{f}_* + \mathbf{g}^*)$ (see also Chepoi *et al.* (1997) and Ubhaya (1974)). If $(X, \prec)$ is given by the covering graph, then the complexity of this procedure is linear in the number of edges of this graph and the size of the set of observations.

Example 2 (l_∞-approximation by convex functions). The cone $\mathscr{C}$ of convex functions on $\mathbb{R}^p$ is another example of a complete lattice. The meet of a family F of convex functions is $\operatorname{conv}\{\mathbf{f} \in F\}$, while the join is $\sup\{\mathbf{f} \in F\}$; see Rockafellar (1970). In numerical applications, we have a finite subset $X = \{x_1, ..., x_n\}$ of $\mathbb{R}^p$ and the corresponding values $y_1, ..., y_n$ of the dependent variable $\mathbf{y}$. We are searching a convex function $\mathbf{f}$ (or, only the restriction of $\mathbf{f}$ to the set X) that minimizes the l_∞-error

$$D_\infty(\mathbf{f}) = \max_{x_i \in X} |y_i - f(x_i)|.$$

If we will take the restriction of each convex function to the set X, then again we will get a complete lattice $\mathscr{C}(X) \subset \mathbb{R}^X$ which is invariant under translations along the diagonal line. Therefore, Proposition 1 and its corollaries apply in this case as well. Let $F = \{(x_i, y_i): x_i \in X\}$. We compute the convex hull $\operatorname{conv}(F)$ of the set F (for algorithms and other details see Edelsbrunner (1987)) and take its lower part $\operatorname{conv}_*(F)$. This is a part of the graph of a piecewise linear convex function $\mathbf{f}_*$, whose restriction on the set X is nothing but the lower maximum approximation of the vector $\mathbf{y}$ in the cone $\mathscr{C}(X)$. Let $\varepsilon = \frac{1}{2} \max_{x_i \in X}\{y_i - \mathbf{f}_*(x_i)\}$. Then the convex function $\mathbf{f}_* + \varepsilon \mathbf{1}$ minimizes the l_∞-error $D_\infty(\mathbf{f})$ with $\mathbf{f} \in \mathscr{C}$.

3. TREES AND ULTRAMETRICS

In this section we present the principal applications of the properties from the previous section. Throughout the section, X will be a finite set with n elements. Let $\mathscr{D} := \mathscr{D}(X)$ be the vector space of functions $\mathbf{d}: X^2 \to \mathbb{R}$ satisfying the properties $\mathbf{d}(x, x) = 0$ for all $x \in X$ and $\mathbf{d}(x, y) = \mathbf{d}(y, x)$ for all $x, y \in X$. The dimension of $\mathscr{D}$ is $n(n-1)/2$. A *dissimilarity* is a function $\mathbf{d} \in \mathscr{D}$ taking nonnegative values. By a slight abuse of terminology, a *metric* is a dissimilarity $\mathbf{d}$ satisfying the triangle condition:

$$\mathbf{d}(x, y) \leqslant \mathbf{d}(x, z) + \mathbf{d}(z, y) \qquad \text{for all} \quad x, y, z \in X.$$

A dissimilarity $\mathbf{d}$ is said to be an *ultrametric* if it satisfies the *ultrametric inequality*:

$$\mathbf{d}(x, y) \leqslant \max\{\mathbf{d}(x, z), \mathbf{d}(y, z)\} \qquad \text{for all} \quad x, y, z \in X.$$

A metric $\mathbf{d}$ is a *tree metric* if it satisfies the following 4-*point condition*:

$$\mathbf{d}(x, y) + \mathbf{d}(z, w) \leqslant \max\{\mathbf{d}(x, z) + \mathbf{d}(y, w), \mathbf{d}(x, w) + \mathbf{d}(y, z)\} \qquad \text{for all} \quad x, y, z, w \in X.$$

The 4-point condition is equivalent with the tree realizability of a given distance matrix (see, for example, Buneman (1974) and Zaretskii (1965)). This constitutes a well-known result used widely in numerical taxonomy. Every ultrametric is a tree metric; see, for example, Critchley and Fichet (1994).

Denote by $\mathscr{U} := \mathscr{U}(X)$ all $\mathbf{d} \in \mathscr{D}$ obeying the ultrametric inequality for *distinct* elements. It is easy to show (and well known) that any $\mathbf{d} \in \mathscr{D}$ admits the subdominant (alias the lower maximum approximation) $\mathbf{d}_*$ in $\mathscr{U}$. If $\mathbf{d}$ is a dissimilarity, then $\mathbf{d}_*$ is an ultrametric. The approximation $\mathbf{d}_*$ can be computed via minimum spanning trees of $\mathbf{d}$; see Gower and Ross (1969) (another way to compute $\mathbf{d}_*$ is to use the well-known single linkage algorithm). Let T be a minimum spanning tree of $\mathbf{d}$. For any $x, y \in X$ set $\mathbf{d}_*(x, y)$ to be the length of a longest edge on the unique path connecting x and y in T. The algorithm can be easily implemented in optimal $O(n^2)$ time. Equivalently, $\mathbf{d}_*(x, y)$ coincides with the so-called bottleneck distance between x and y; see Hu (1960). Namely, in any path connecting x and y in the complete graph, take the longest edge and among such edges take the shortest one. Its length is $\mathbf{d}_*(x, y)$.

Farach *et al.* (1995) presented an algorithm with complexity $O(n^2)$ for the sandwich problem and the l_∞-fitting problem for ultrametrics. The implementation details and the correctness analysis of this algorithm are rather complicated. The results of the previous section suggest an alternative approach to these problems. Indeed, $\mathscr{U}$ is invariant under translations along the diagonal $\mathscr{L}$ of $\mathscr{D}$ (notice that $\mathscr{U}$ is a join subsemilattice of $\mathscr{D}$). By Proposition 1 and Corollary 1, $\mathbf{d}_* + \frac{1}{2}\,\|\mathbf{d} - \mathbf{d}_*\|\,\mathbf{1}$ is a best l_∞-approximation of $\mathbf{d}$ in $\mathscr{U}$. Again, if $\mathbf{d}$ is a dissimilarity, then $\mathbf{d}_* + \frac{1}{2}\,\|\mathbf{d} - \mathbf{d}_*\|\,\mathbf{1}$ is an ultrametric. More generally, if $\mathbf{d}', \mathbf{d}'' \in \mathscr{D}$ and $\mathbf{d}' \prec \mathbf{d}''$, then $\mathbf{d}'_* + \frac{1}{2}\,\|\mathbf{d}'' - \mathbf{d}'_*\|\,\mathbf{1}$ is a solution of the problem (P_2) related to $\mathscr{U}$. Since $\mathscr{U}$ is *join-closed*, the sandwich problem has a solution if and only if $\mathbf{d}' \prec \mathbf{d}''_*$. The complexity of these procedures is the complexity to find the lower maximum approximation, i.e., $O(n^2)$. To solve the consensus problem for a subset A of $\mathscr{U}$ we find $\mathbf{d}' = \inf\{\mathbf{u}_i \in A\}$ and $\mathbf{d}'' = \sup\,\{\mathbf{u}_i \in A\}$. Then $\mathbf{d}'' \in \mathscr{U}$ and thus $\mathbf{d}'' - \frac{1}{2}\,\|\mathbf{d}'' - \mathbf{d}'\|\,\mathbf{1}$ provides a solution to the consensus problem. Its complexity is $O(n^2\,|A|)$.

Finally, suppose that $\mathbf{d}$ is a partial dissimilarity, that is $\mathbf{d}$ is defined on a certain subset E of pairs. Denote by $\mathbf{d}'$ the extension of $\mathbf{d}$ to the whole set $X \times X$ by setting $\mathbf{d}'(u, v) = M$ on all missing pairs (u, v), where $M > \max\{\mathbf{d}(x, y) \colon (x, y) \in E\}$. Then there is an ultrametric extension of $\mathbf{d}$ if and only if the subdominant of $\mathbf{d}'$ coincides with $\mathbf{d}$ on E. Otherwise, to find a best l_∞-ultrametric approximation of $\mathbf{d}$ one can proceed as follows. Let $\mathscr{D}_E$ denote the set of all partial dissimilarities on E and let $\mathscr{U}_E$ be those vectors of $\mathscr{D}_E$ which have at least one ultrametric extension. Then $\mathscr{U}_E$ is join-closed and contains the diagonal line of $\mathscr{D}_E$. Therefore, for every $\mathbf{d} \in \mathscr{U}_E$ we have a subdominant and an l_∞-approximation in $\mathscr{U}_E$. We acknowledge A. Guénoche for drawing our attention to this question and suggesting that the subdominant in this case is the ultrametric obtained from the minimum spanning forest of $\mathbf{d}$ on E.

Let $\mathscr{T} := \mathscr{T}(X)$ be the set of $\mathbf{d} \in \mathscr{D}$ satisfying the 4-point condition for *distinct* elements. From a result of Bandelt and Steel (1995), all $\mathbf{d} \in \mathscr{T}$ can be realized by weighted trees with real edge weights. Choose a *base-point* a of X, and denote

$X_a := X \backslash \{a\}$, $\mathscr{U}_a := \mathscr{U}(X_a)$. There is a close relationship between the cones $\mathscr{T}$ and $\mathscr{U}_a$. Define $\mathbf{d}^a \in \mathscr{D}$ with zero diagonal by setting

$$\mathbf{d}^a(x, y) = \mathbf{d}(x, y) - \mathbf{d}(x, a) - \mathbf{d}(y, a) \tag{1}$$

for $x \neq y$ in X. The quantity $-\frac{1}{2}\mathbf{d}^a(x, y)$ is known as the *Gromov product* of x, $y \in X_a$ (see Ghys and de la Harpe (1990)), while $\mathbf{d}^a$ plus an additive constant is called the *Farris transform* of $\mathbf{d}$ (see Bandelt (1990) and Leclerc (1995)). The following result in the case of metrics has been suggested by Farris, Kluge, and Eckardt (1970) and was rediscovered by Klotz and Blanken (1981) and Brossier (1985). In the full generality, the result is due to Leclerc (1995).

LEMMA 1. *$\mathbf{d} \in \mathscr{T}$ if and only if $\mathbf{d}^a|_{X_a} \in \mathscr{U}_a$.*

We continue with a characterization of tree metrics.

LEMMA 2. *For a function $\mathbf{d} \in \mathscr{D}$ the following conditions are equivalent:*

(i) *$\mathbf{d}$ is a tree metric;*

(ii) *$\mathbf{d} \in \mathscr{T}$ and for any $x, y \in X$, $|\mathbf{d}(a, x) - \mathbf{d}(a, y)| \leqslant \mathbf{d}(x, y) \leqslant \mathbf{d}(a, x) + \mathbf{d}(a, y)$;*

(iii) *$\mathbf{d}^a \in \mathscr{U}$ and for any $x, y \in X$, $|\mathbf{d}(a, x) - \mathbf{d}(a, y)| \leqslant \mathbf{d}(x, y)$.*

Proof. (i) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (iii). First, $\mathbf{d}^a|_{X_a} \in \mathscr{U}_a$. For any pair of distinct points $x, y \in X_a$, $\mathbf{d}^a(x, y) \leqslant 0$ by (ii), so that $\mathbf{d}^a(x, y) \leqslant \mathbf{d}^a(a, x) = \mathbf{d}^a(a, y)$.

(iii) $\Rightarrow$ (i). Clearly, $\mathbf{d}^a \in \mathscr{U} \subseteq \mathscr{T}$ infers that $\mathbf{d} \in \mathscr{T}$. Pick three distinct elements x, y, z of X. In order to prove the metric inequality observe that $\mathbf{d}(x, y) \leqslant \mathbf{d}(x, z) + \mathbf{d}(y, z)$ if and only if $\mathbf{d}^a(x, y) - 2\mathbf{d}(a, z) \leqslant \mathbf{d}^a(x, z) + \mathbf{d}^a(y, z)$. But, for every $v \in X$, $v \neq z$, we have $\mathbf{d}^a(z, v) + 2\mathbf{d}(a, z) \geqslant 0$. Indeed, $\mathbf{d}^a(z, v) + 2\mathbf{d}(a, z) = \mathbf{d}(z, v) + \mathbf{d}(a, z) - \mathbf{d}(a, v) \geqslant 0$. Thus, $-2\mathbf{d}(a, z) \leqslant \min\{\mathbf{d}^a(x, z), \mathbf{d}^a(y, z)\}$ and $\mathbf{d}^a(x, y) \leqslant \max\{\mathbf{d}^a(x, z), \mathbf{d}^a(y, z)\}$, completing the proof. ∎

Remark. From Lemma 1 one can deduce the following strengthening of the 4-point condition for tree metrics: *Let $\mathbf{d} \in \mathscr{D}$ and let a be fixed point of X. Then $\mathbf{d} \in \mathscr{T}$ if and only if $\mathbf{d}$ satisfies the 4-point condition for a and any distinct points x, y, $z \in X_a$.*

Given a base-point $a \in X$ and a function r: $X \to \mathbb{R}$ with $r(a) = 0$, we denote by $\mathscr{D}_{a, r}$ the collection of all the dissimilarities $\mathbf{d}$ such that $\mathbf{d}(a, x) = r(x)$ for all $x \in X$. $\mathscr{D}_{a, r}$ is an affine vector space of dimension $(n-1)(n-2)/2$. Let $\mathscr{T}_{a, r} = \mathscr{T} \cap \mathscr{D}_{a, r}$. By $\mathscr{T}^m_{a, r}$ we denote all tree metrics from $\mathscr{T}_{a, r}$. One can easily show that $\mathscr{T}_{a, r}$ is invariant under translations along the diagonal line of $\mathscr{D}_{a, r}$ (this is no longer true for the diagonal line of the space $\mathscr{D}$).

PROPOSITION 2. *Any $\mathbf{d} \in \mathscr{D}_{a, r}$ admits the lower maximum approximation in $\mathscr{T}_{a, r}$. A function $\mathbf{d} \in \mathscr{D}_{a, r}$ admits the lower maximum approximation in $\mathscr{T}^m_{a, r}$ if and only if*

$$|\mathbf{d}(a, x) - \mathbf{d}(a, y)| \leqslant \mathbf{d}(x, y) \tag{2}$$

for any x, $y \in X$. If $\mathbf{d}$ is a metric, then the two approximations coincide.

Proof. Let $(\mathbf{d}^a|_{X_a})_*$ be the subdominant of $\mathbf{d}^a|_{X_a}$ in $\mathscr{U}_a$. Extend $(\mathbf{d}^a|_{X_a})_*$ to X by setting $(\mathbf{d}^a|_{X_a})_*(a, x) = 0$ for any $x \in X$. Define $\mathbf{d}_* \in \mathscr{D}_{a,r}$ obtained from $(\mathbf{d}^a|_{X_a})_*$ via the inverse of (1). By Lemma 1, we deduce that $\mathbf{d}_*$ is the lower maximum approximation of $\mathbf{d}$ in $\mathscr{T}_{a,r}$.

Clearly, the triangle condition (2) is a necessary condition for the existence of a metric less than $\mathbf{d}$ and preserving the values $\mathbf{d}(a, x)$, in particular of a tree metric in $\mathscr{T}^m_{a,r}$ less than $\mathbf{d}$. Now, suppose that $\mathbf{d}$ obeys (2). Let $\mathbf{d}^a_*$ be the subdominant of $\mathbf{d}^a$ in $\mathscr{U}$. Since $\mathbf{d}^a_*(a, x) \leqslant \mathbf{d}^a(a, x) = 0$, for every $x \in X$, we obtain that $\mathbf{d}^a_*(x, y) \leqslant 0$ for arbitrary $x, y \in X$. Consequently, we deduce $\mathbf{d}^a_*(a, x) = 0$ for every $x \in X$. Thus, the function $\mathbf{d}_*$ obtained from $\mathbf{d}^a_*$ via the inverse of (1) belongs to the set $\mathscr{T}_{a,r}$. If $\mathbf{d}$ is a metric, then $\mathbf{d}^a(x, y) \leqslant 0$ for any $x, y \in X_a$. From this and previous arguments we deduce that the restriction of $\mathbf{d}^a_*$ on X_a coincides with $(\mathbf{d}^a|_{X_a})_*$, thus proving the third assertion. It remains to prove that $\mathbf{d}_*$ is a metric. By Lemma 2 it suffices to show that $\mathbf{d}_*(a, x) - \mathbf{d}_*(a, y) \leqslant \mathbf{d}_*(x, y)$ for any distinct x and y. But this condition is equivalent to $\mathbf{d}^a_*(a, x) - 2r(y) \leqslant \mathbf{d}^a_*(a, y) + \mathbf{d}^a_*(x, y)$, i.e.,

$$\mathbf{d}^a_*(x, y) + 2r(y) \geqslant 0. \tag{3}$$

To establish (3) for fixed x and y, set $\alpha := \min_{z \neq y} \mathbf{d}^a(y, z)$. The function $\mathbf{d}'$ defined by $\mathbf{d}'(y, z) = \alpha$ for every $z \neq y$ and $\mathbf{d}'(u, v) = \beta < \alpha$ elsewhere is in $\mathscr{U}$ and less than $\mathbf{d}^a$, hence less than $\mathbf{d}^a_*$, for β sufficiently small. Therefore $\mathbf{d}^a_*(x, y) \geqslant \mathbf{d}^a(y, z)$ for some $z \neq y$. Hence $\mathbf{d}^a_*(x, y) + 2r(y) \geqslant \mathbf{d}(y, z) + r(y) - r(z) \geqslant 0$ by (2). Thus $\mathbf{d}_* \in \mathscr{T}^m_{a,r}$ and by Lemma 2, $\mathbf{d}_*$ is the lower maximum approximation of $\mathbf{d}$ in $\mathscr{T}^m_{a,r}$. ∎

From Proposition 2 we obtain a simple optimal algorithm for constructing the subdominant $\mathbf{d}_*$ of $\mathbf{d}$ in $\mathscr{T}_{a,r}$ and even in $\mathscr{T}^m_{a,r}$ when it exists. From $\mathbf{d}$ we find $\mathbf{d}^a$ applying (1); then we compute the subdominants $(\mathbf{d}^a|_{X_a})_*$ and $\mathbf{d}^a_*$. This can be done in time $O(n^2)$ by using a minimum spanning tree T of $\mathbf{d}^a$. Then find $\mathbf{d}_*$ using the inverse transform of (1). The whole computation requires $O(n^2)$ number of operations. This construction implicitly has been used by M. Gromov for finding approximating trees in hyperbolic spaces; see Ghys and de la Harpe (1990).

Let $\hat{\mathbf{d}}$ and $\tilde{\mathbf{d}}$ be best tree metric l_∞-approximations of the metric $\mathbf{d}$ in the sets $\mathscr{T}_{a,r}$ and $\mathscr{T}$, respectively. (We assume $r(x) = \mathbf{d}(a, x)$ for any $x \in X$.) Agarwala *et al.* (1996) established that

$$\|\mathbf{d} - \hat{\mathbf{d}}\| \leqslant 3 \cdot \|\mathbf{d} - \tilde{\mathbf{d}}\|.$$

They showed that the problem of computing $\tilde{\mathbf{d}}$ is NP-hard, while the problem of finding $\hat{\mathbf{d}}$ can be solved by Farris transform and a modification of the algorithm from Farach *et al.* (1995). Namely, expressed here with the equivalent transform (1), their approach is as follows. Starting from $\mathbf{d}$, we get $\mathbf{d}^a$ by (1). Then we may compute an l_∞-ultrametric approximation $\hat{\mathbf{d}}^a$ of $\mathbf{d}^a$ that yields by the converse transform of (1) an element $\hat{\mathbf{d}}$. In order to ensure $\hat{\mathbf{d}} \in \mathscr{D}_{a,r}$, the function $\hat{\mathbf{d}}^a$ must satisfy the constraint

$$\hat{\mathbf{d}}^a(a, x) = 0 \qquad \text{for any} \quad x \in X. \tag{4}$$

Then, $\hat{\mathbf{d}}$ will be an l_∞-approximation of $\mathbf{d}$ in $\mathcal{T}^m_{a,r}$ provided $\hat{\mathbf{d}}$ is a metric. To guarantee that, we have to impose an additive constraint on $\hat{\mathbf{d}}^a$, analogous to (3):

$$\text{for any} \quad x, y \in X, \quad x \neq y, \qquad \hat{\mathbf{d}}^a(x, y) + 2r(y) \geqslant 0. \tag{5}$$

Agarwala *et al.* (1996) modify the algorithm from Farach *et al.* (1995) for computing $\hat{\mathbf{d}}^a$ satisfying the constraints (4) and (5).

We can get the same performances by starting from the subdominant $\mathbf{d}_*$ of $\mathbf{d}$ in $\mathcal{T}_{a,r}$ and $\mathcal{T}^m_{a,r}$. As before, let $\varepsilon = \frac{1}{2}\|\mathbf{d} - \mathbf{d}_*\|$ and $\mathbf{1} \in \mathcal{D}_{a,r}$. From Proposition 1 we obtain the following result.

COROLLARY 5. *Let* $\mathbf{d} \in \mathcal{D}_{a,r}$ *and* $\mathbf{d}_*$ *be the subdominant of* $\mathbf{d}$ *in* $\mathcal{T}_{a,r}$. *Then* $\hat{\mathbf{d}} = \mathbf{d}_* + \varepsilon\mathbf{1}$ *is a best* l_∞*-approximation of* $\mathbf{d}$ *in* $\mathcal{T}_{a,r}$ *and*

$$\|\mathbf{d} - \hat{\mathbf{d}}\| \leqslant 3 \cdot \delta(\mathbf{d}, \mathcal{T}).$$

The proof of the last inequality is identical to the proof of Lemma 3.3 of Agarwala *et al.* (1996).

Now, suppose that $\mathbf{d} \in \mathcal{D}_{a,r}$ fulfills the triangle condition (2). Let $\mathbf{d}_*$ be its subdominant in $\mathcal{T}^m_{a,r}$. Define $\hat{\mathbf{d}}_0 = \mathbf{d}_* + \varepsilon\mathbf{1}$. If $\mathbf{d}$ is a metric, then $\hat{\mathbf{d}}_0$ is a best l_∞-approximation in $\mathcal{T}_{a,r}$ (in fact, it coincides with the l_∞-approximation $\hat{\mathbf{d}}$ from Corollary 5). However, as the following example shows, $\hat{\mathbf{d}}_0$ is not necessarily a metric. Let $X = \{a, b, x, y, z\}$ and

$$\mathbf{d}(a, x) = \mathbf{d}(a, y) = \mathbf{d}(a, z) = \mathbf{d}(b, x) = \mathbf{d}(b, y) = \mathbf{d}(b, z) = 16,$$
$$\mathbf{d}(a, b) = 2, \qquad \mathbf{d}(x, y) = \mathbf{d}(x, z) = 10, \qquad \mathbf{d}(y, z) = 20.$$

Then

$$\mathbf{d}^a(x, y) = \mathbf{d}^a(x, z) = -22, \quad \mathbf{d}^a(b, x) = \mathbf{d}^a(b, y) = \mathbf{d}^a(b, z) = -2, \quad \mathbf{d}^a(y, z) = -12.$$

The subdominant $\mathbf{d}^a_*$ of $\mathbf{d}^a$ in $\mathcal{U}$ is given by $\mathbf{d}^a_*(x, y) = \mathbf{d}^a_*(x, z) = \mathbf{d}^a_*(z, y) = -22$, $\mathbf{d}^a_*(b, x) = \mathbf{d}^a_*(b, y) = \mathbf{d}^a_*(b, z) = -2$. The tree metric $\mathbf{d}_*$ is represented by a tree T; see Fig. 1a. Note that $\varepsilon = 5$ and that $\hat{\mathbf{d}}_0 = \mathbf{d}_* + \varepsilon\mathbf{1}$ is not a metric. Still $\hat{\mathbf{d}}_0$ can be represented by a weighted tree obtained from T by increasing the lengths of its stems by $\varepsilon/2$ and decreasing by the same value the length of the stem incident to a; see Fig. 1b.

To repair this, it suffices instead of $\hat{\mathbf{d}}_0$ to consider $\hat{\mathbf{d}}_1 := \mathbf{d}_* + \varepsilon\mathbf{1}$, where ε is defined as before and where $\mathbf{1}$ is the diagonal line of $\mathcal{D}$. In other words, for any distinct $x, y \in X$ we set $\hat{\mathbf{d}}_1(x, y) = \mathbf{d}_*(x, y) + \varepsilon$. If $\mathbf{d} \in \mathcal{D}_{a,r}$ obeys (2), by Proposition 2, $\mathbf{d}_*$ is a tree metric; thus $\hat{\mathbf{d}}_1$ is a tree metric, too. The tree of $\hat{\mathbf{d}}_1$ has the same topology as the tree representing $\mathbf{d}_*$. It is obtained from T by increasing the lengths of all stems by $\varepsilon/2$.

COROLLARY 6. *If* $\mathbf{d} \in \mathcal{D}_{a,r}$ *fulfills the triangle condition* (2), *then* $\hat{\mathbf{d}}_1$ *is a tree metric and*

$$\|\mathbf{d} - \hat{\mathbf{d}}_1\| = \delta(\mathbf{d}, \mathcal{T}^m_{a,r}).$$

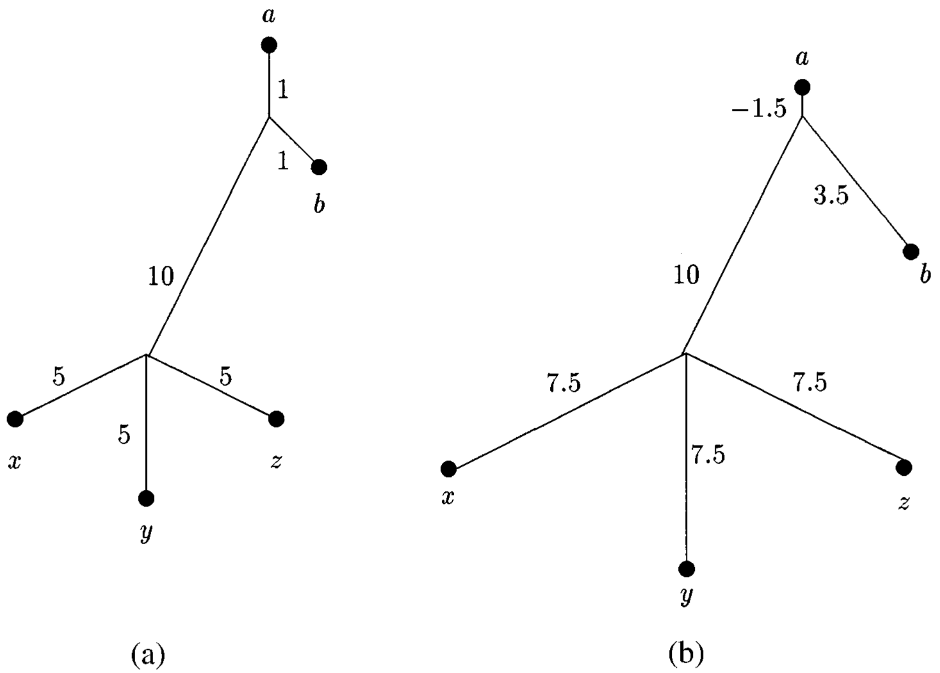

FIG. 1. (a) $\mathbf{d}_*$ and (b) $\hat{\mathbf{d}}_0$.

We may also obtain a best l_∞-approximation in $\mathscr{T}^m_{a,r}$, thus recovering the result from Agarwala *et al.* (1996). Our result is more general and we avoid the solution of a restricted l_∞-approximation problem as in the approach of Agarwala *et al.* (1996). Here $\mathbf{1}$ stands for the diagonal line of $\mathscr{D}$.

COROLLARY 7. *Let* $\mathbf{d} \in \mathscr{D}_{a,r}$ *fulfill the triangle condition* (2) *and suppose that* $\mathbf{d}^a$ *is obtained from* $\mathbf{d}$ *using* (1). *Let* $\mathbf{d}^a_*$ *be the subdominant of* $\mathbf{d}^a$ *in* $\mathscr{U}$. *Define* $\hat{\mathbf{d}}$ *to be obtained from the meet of* $\mathbf{0}$ *and* $\mathbf{d}^a_* + \frac{1}{2}\,\|\mathbf{d}^a - \mathbf{d}^a_*\|\,\mathbf{1}$ *by the transformation inverse to* (1). *Then* $\hat{\mathbf{d}}$ *is a best* l_∞*-approximation of* $\mathbf{d}$ *in* $\mathscr{T}^m_{a,r}$.

Indeed, by Proposition 1, $\mathbf{d}^a_* + \frac{1}{2}\,\|\mathbf{d}^a - \mathbf{d}^a_*\|\,\mathbf{1}$ is an l_∞-approximation of $\mathbf{d}^a$ in $\mathscr{U}$, and clearly the same remains true for the meet of $\mathbf{0}$ and $\mathbf{d}^a_* + \frac{1}{2}\,\|\mathbf{d}^a - \mathbf{d}^a_*\|\,\mathbf{1}$. Moreover, the latter function fulfills (5), since $\mathbf{d}^a_*$ does (see (3)).

Note that if the initial $\mathbf{d}$ is a tree metric, then $\hat{\mathbf{d}}$ and $\mathbf{d}_*$ coincide with $\mathbf{d}$, giving an optimal $O(n^2)$ algorithm for recognizing tree metrics; see Bandelt (1990), Barthélemy and Guénoche (1991), Batagelj, Pisanski, and Simoes-Pereira (1990), and Leclerc (1995) for other recognition procedures.

Let $\mathscr{U}_{a,r} = \mathscr{U} \cap \mathscr{D}_{a,r}$. We conclude this section with a simple procedure for finding a best l_∞-approximation $\hat{\mathbf{d}}$ in $\mathscr{U}_{a,r}$ of a function $\mathbf{d} \in \mathscr{D}_{a,r}$. Let $r_1 < \cdots < r_k$ be the distinct values of $r(x)$, $x \in X_a$ and put $B_i = \{x \in X_a\colon \mathbf{d}(a,x) = r_i\}$. Then $\hat{\mathbf{d}}(a,x) = r_i$ for any $x \in B_i$. By the ultrametric inequality, $\hat{\mathbf{d}}(x,y) = r_j$ for any $x \in B_i$ and $y \in B_j$ with $i < j$. To solve the initial problem, it suffices to find for each $\mathbf{d}|_{B_i}$ a best l_∞-approximation in $\mathscr{U}(B_i)$ all of whose values do not exceed r_i. Let $\hat{\mathbf{d}}_i$ be a best l_∞-approximation in $\mathscr{U}(B_i)$ of $\mathbf{d}|_{B_i}$ (for example, that computed by our algorithm). Define a new $\hat{\mathbf{d}}'_i \in \mathscr{U}(B_i)$ by letting $\hat{\mathbf{d}}'_i(u,v) = \min\{r_i, \hat{\mathbf{d}}_i(x,y)\}$. We assert that $\hat{\mathbf{d}}'_i$ a required l_∞-approximation of $\mathbf{d}|_{B_i}$. Indeed, if $\|\hat{\mathbf{d}}'_i - \mathbf{d}|_{B_i}\| = \|\hat{\mathbf{d}}_i - \mathbf{d}|_{B_i}\|$, there is nothing

TABLE 1

Distance Matrix d of Bartlett

	ORIG	N1	N15	HP1	HP8	L1	L120	P1	P45	P105	P1145	R1	R15	R45	X180
ORIG	0														
N1	81	0													
N15	81	32	0												
HP1	119	102	106	0											
HP8	121	102	108	42	0										
L1	78	85	89	107	113	0									
L120	110	91	99	97	99	88	0								
P1	118	83	101	91	97	102	92	0							
P45	109	88	94	100	102	107	95	59	0						
P105	113	92	104	94	100	105	97	49	52	0					
P1145	115	80	84	76	80	107	87	71	80	76	0				
R1	80	59	65	75	83	92	82	80	77	87	53	0			
R15	91	62	68	76	84	93	79	75	78	86	46	23	0		
R45	101	68	68	78	92	103	77	79	86	90	42	37	20	0	
X180	113	78	90	84	94	111	87	83	92	90	58	65	58	56	0

to prove. So, assume that $\|\hat{\mathbf{d}}'_i - \mathbf{d}|_{B_i}\| > \|\hat{\mathbf{d}}_i - \mathbf{d}|_{B_i}\|$ and that this error is realized by a pair $u, v \in B_i$. Then obviously $\hat{\mathbf{d}}'_i(u, v) = r_i$ and $\mathbf{d}(u, v) \geqslant r_i$. For any other $\hat{\mathbf{d}}''_i \in \mathscr{U}(B_i)$ whose values do not exceed r_i, we have $\mathbf{d}(u, v) - \hat{\mathbf{d}}''_i(u, v) \geqslant \mathbf{d}(u, v) - \hat{\mathbf{d}}'_i(u, v)$, thus establishing the optimality of $\hat{\mathbf{d}}'_i$. If the initial $\mathbf{d} \in \mathscr{D}_{a,r}$ is a dissimilarity, then all $\hat{\mathbf{d}}'_i$, $i = 1, ..., k$, are ultrametrics, and the global solution is an ultrametric, too.

To illustrate the results of this section, we consider the distance matrix between 15 texts obtained by Bartlett (1932) (Table 1) with the method of repeated reproduction of the legend "The war of the ghosts" (we acknowledge J.-P. Barthélemy for suggesting this example). This distance matrix **d** has been

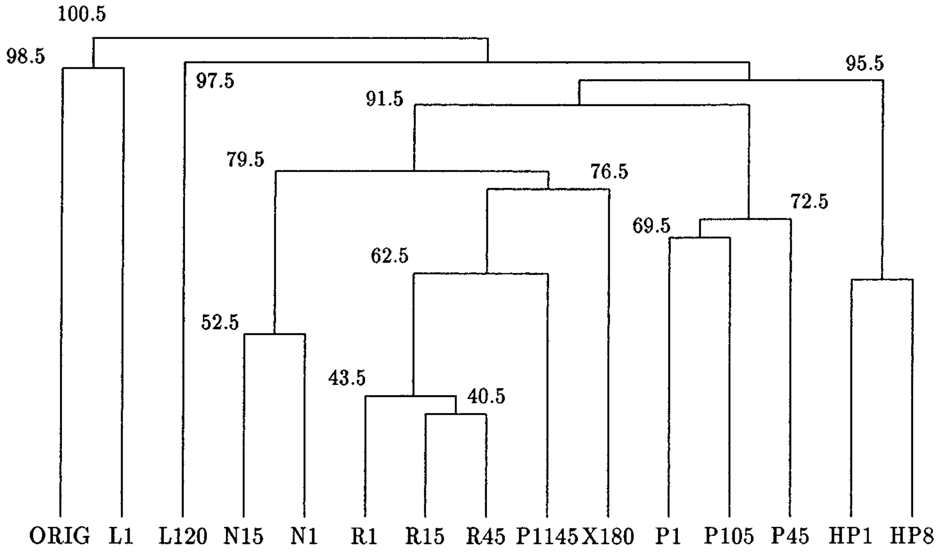

FIG. 2. A best l_∞-ultrametric approximation of **d**.

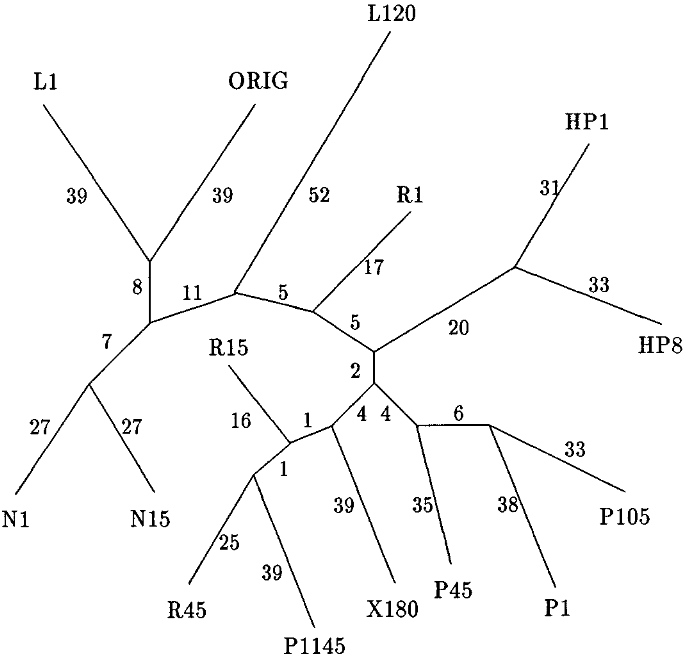

FIG. 3. A best l_∞-approximation of $\mathbf{d}$ in $\mathscr{T}^m_{a,r}$ preserving the distances of ORIG.

approximated by a tree metric by Abdi, Barthélemy, and Luong (1984) using a different approach. Six different subjects are denoted by the letter(s) beginning the labels. The number following the letter(s) indicates the number of days between the first presentation and the recall. ORIG denotes the original text. Although slightly different from the tree of Abdi *et al.* (1984), both the ultrametric and the tree l_∞-approximations of $\mathbf{d}$ indicate a strong fidelity of most subjects to themselves (see Figs. 2 and 3).

4. UNIVERSAL SOLUTIONS

In Section 2 we presented some simple properties of the set S of optimal solutions of the problem (P_2) under the l_∞-norm. In general S is sufficiently large, so that it might be desirable to distinguish a unique element of S, sharing certain additional properties. A major example is provided by the consensus problem. It is in the nature of this problem to return a unique solution or to distinguish one among the set of solutions. In this section we present a method to select a universal solution.

Let K be *join*-closed and invariant under translations along the diagonal line $\mathscr{L}$. By ε denote the optimal error of (P_2) for vectors $\mathbf{u}, \mathbf{v} \in \mathbf{E}$. Additionally we suppose that $K_{\succ}(\mathbf{v})$ has a minimum element $\mathbf{v}'$. Then S has a minimum element $\mathbf{u}_1 := \mathbf{v}' - \varepsilon\mathbf{1}$ and a maximum element $\mathbf{v}_1 := \mathbf{u}_* + \varepsilon\mathbf{1}$. This implies that S coincides with the order interval $[\mathbf{u}_1, \mathbf{v}_1] := \{\mathbf{y} \in K\colon \mathbf{u}_1 \prec \mathbf{y} \prec \mathbf{v}_1\}$.

Define a decreasing chain of nonempty subsets $\{S_i\}$ of K. For this, set $S_1 := S$; now, given S_n, let $S_{n+1} = S_n \cap S'_n$, where S'_n is the set of solutions for the l_∞-consensus of S_n. The next result shows that this procedure stabilizes provided we obtain a one-point set.

PROPOSITION 3. *After at most p steps we will obtain a set with a single solution* $\mathbf{w}$, *called the universal solution. If, in addition, K is meet-closed then* $\mathbf{w}=\frac{1}{2}(\mathbf{u}_* + \mathbf{v}^*)$.

Proof. First, by induction on n we will prove that every S_n is an order interval. As we noticed before, this is true for $n=1$. Suppose, by the induction hypothesis, that $S_n=[\mathbf{u}_n, \mathbf{v}_n]$. Let ε_n be the optimal error for the l_∞-consensus of S_n. The set S'_n has a minimum element $\mathbf{u}'_n := \mathbf{v}_n - \varepsilon_n \mathbf{1}$ and a maximum element $\mathbf{v}'_n := \mathbf{u}_n + \varepsilon_n \mathbf{1}$. Let $\mathbf{u}_{n+1}$ be the join of $\mathbf{u}_n$ and $\mathbf{u}'_n$. Since K is *join*-closed, the vector $\mathbf{u}_{n+1}$ belongs to K and lies between $\mathbf{u}_n$ and $\mathbf{v}_n$, whence $\mathbf{u}_{n+1} \in S_{n+1}$. Moreover, $\mathbf{u}_{n+1}$ is the minimum element of S_{n+1}. Notice that a vector belongs to S_{n+1} if and only if it lies between $\mathbf{u}_{n+1}$ and the meet of $\mathbf{v}_n$ and $\mathbf{v}'_n$. Therefore, if $\mathbf{v}_{n+1}$ is the subdominant of the latter vector, then $S_{n+1}=[\mathbf{u}_{n+1}, \mathbf{v}_{n+1}]$. Moreover, if $\varepsilon_n \neq 0$, then for every $1 \leqslant i \leqslant p$ such that $(\mathbf{v}_n)_i - (\mathbf{u}_n)_i = 2\varepsilon_n$, we infer that $(\mathbf{v}_{n+1})_i = (\mathbf{u}_{n+1})_i$. Hence, after at most p steps, we will arrive at a single solution $\mathbf{w}$.

Now, suppose that K is *meet*-closed. Then $\mathbf{v}'=\mathbf{v}^*$, and, consequently, $\mathbf{w} := \frac{1}{2}(\mathbf{u}_* + \mathbf{v}^*)$ coincides with the vector $\frac{1}{2}(\mathbf{u}_1 + \mathbf{v}_1)$. By the induction on n and the previous proof, one can easily check that $\mathbf{w} = \frac{1}{2}(\mathbf{u}_n + \mathbf{v}_n)$, concluding the proof. ∎

We continue with an efficient procedure for finding the universal solution $\mathbf{w}$. As in some previous methods, we suppose that there is an algorithm for computing the subdominants and, additionally, a method for finding the minimum element $\mathbf{v}'$ of $K_\succ(\mathbf{v})$ (in the consensus problem, $\mathbf{v}'=\mathbf{v}$). The procedure consists of at most p iterations, the complexity of each iteration being the complexity of computing the subdominant of some vector. In the preprocessing step, we find $\mathbf{v}'$, the subdominant $\mathbf{u}_*$ of $\mathbf{u}$, and the error $\varepsilon = \frac{1}{2}\|\mathbf{v}' - \mathbf{u}\|$. Set $\mathbf{u}_1 = \mathbf{v}' - \varepsilon \mathbf{1}$ and $\mathbf{v}_1 = \mathbf{u}_* + \varepsilon \mathbf{1}$.

(1) If $\varepsilon_n = \frac{1}{2}\|\mathbf{v}_n - \mathbf{u}_n\| = 0$ then return $\mathbf{w} := \mathbf{u}_n$.

(2) Compute the join $\mathbf{u}_{n+1}$ of $\mathbf{u}_n$ and $\mathbf{v}_n - \varepsilon_n \mathbf{1}$.

(3) Find the subdominant $\mathbf{v}_{n+1}$ of the meet of $\mathbf{v}_n$ and $\mathbf{v}'_n := \mathbf{u}_n + \varepsilon_n \mathbf{1}$.

(4) Set $n := n+1$ and go to step (1).

This algorithm yields a universal l_∞-consensus. It can be applied to convex regressions, ultrametrics, or trees from $\mathcal{T}_{a,r}$. For isotonic regressions, Proposition 3 provides a new characterization of the solution $\frac{1}{2}(\mathbf{f}_* + \mathbf{g}^*)$.

REFERENCES

Abdi, H., Barthélemy, J.-P., & Luong, X. (1984). Tree representation of associative structures in semantic and episodic memory research. In E. Degreef, & J. Van Buggenhaut (Eds.), *Trends in mathematical psychology*. Amsterdam: North-Holland.

Agarwala, R., Bafna, V., Farach, M., Narayanan, B., Paterson, M., & Thorup, M. (1996). On the approximability of numerical taxonomy: Fitting distances by tree metrics. In *Proceedings of the 7th Annual ACM-SIAM Symposium on Discrete Algorithms.*

Bandelt, H.-J. (1990). Recognition of tree metrics. *SIAM Journal of Discrete Mathematics*, **3**, 1–6.

Bandelt, H.-J., & Steel, M. A. (1995). Symmetric matrices representable by weighted trees over a cancellative abelian monoid. *SIAM Journal of Discrete Mathematics*, **8**, 517–525.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions*. New York: Wiley.

Barthélemy, J.-P., & Guénoche, A. (1991). *Trees and proximity representations*. New York: Wiley.

Bartlett, F. (1932). *Remembering*: *A study in experimental and social psychology*. Cambridge, MA: Cambridge Univ. Press.

Batagelj, V., Pisanski, T., & Simoes-Pereira, J. (1990). An algorithm for tree-realizability of distance matrices. *International Journal of Computer Mathematics*, **34**, 171–176.

Birkhoff, G. (1967). *Lattice theory* (3rd ed.). Providence, RI: Amer. Math. Soc..

Brossier, G. (1985). Approximation des dissimilarités par des arbres additifs. *Mathématiques et Sciences humaines*, **91**, 5–21.

Buneman, P. (1974). A note on the metric properties of trees. *Journal of Combinatorial Theory Ser. B*, **17**, 48–50.

Carroll, J. D. (1976). Spatial, nonspatial and hybrid models for scaling. *Psychometrika*, **41**, 439–463.

Carroll, J. D., & Pruzansky, S. (1980). Discrete and hybrid models for scaling. In E. D. Lanterman & H. Feger (Eds.), *Similarity and choice* (pp. 108–139). Bern: Hans Huber.

Cavalli-Sforza, L., & Edwards, A. (1967). Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chepoi, V., Cogneau, D., & Fichet, B. (1997). Polynomial algorithms for isotonic regression problem. In Y. Dodge (Ed.), *L_1-Statistical procedures and related topics*. Institute of Mathematical Statistics, Lecture Notes Monograph Series, **31**, 147–160.

Cohen, J., & Farach, M. (1997). Numerical taxonomy on data: experimental results. In *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms*.

Critchley, F., & Fichet, B. (1994). The partial order by inclusion of the principal classes of dissimilarity on a finite set, and some of their basic properties. In B. van Cutsem (Ed.), *Classification and dissimilarity analysis*. Lecture Notes in Statistics (pp. 5–65). New York: Springer-Verlag.

Cunningham, J. P. (1978). Free trees and bidirectional trees as representations of psychological distance. *Journal of Mathematical Psychology*, **17**, 165–188.

Day, W. H. E. (1987). Computational complexity of inferring phylogenies from dissimilarity matrices. *Bulletin of Mathematical Biology*, **49**, 461–467.

De Soete, G. (1983). A least squares algorithm for fitting additive trees to proximity data. *Psychometrika*, **48**, 621–626.

Edelsbrunner, H. (1987). *Algorithms in combinatorial geometry*. New York: Springer-Verlag.

Farach, M., Kannan, S., & Warnow, T. (1995). A robust model for finding optimal evolutionary trees. *Algorithmica*, **13**, 155–179.

Farris, J. S., Kluge, A. G., & Eckardt, M. J. (1970). A numerical approach to phylogenetic systematics. *Systematic Zoology*, **19**, 172–189.

Ghys, E., & de la Harpe, P. (1990). *Les Groupes Hyperboliques d'après M. Gromov*. Progress in Mathematics (Vol. 83). Basel: Birkhauser.

Gower, J. C., & Ross, G. K. S. (1969). Minimum spanning tree and single linkage cluster analysis. *Applied Statistics*, **18**, 54–64.

Hu, T. C. (1960). The maximum capacity route problem. *Operation Research*, **8**, 733–736.

Klotz, L. C., & Blanken, R. L. (1981). A practical method for calculating evolutionary trees from sequence data. *Journal of Theoretical Biology*, **91**, 261–272.

Křivánek, M. (1988). The complexity of ultrametric partitions on graphs. *Information Processing Letters*, **27**(5), 265–270.

Křivánek, M., & Morávek, J. (1986). *NP*-Hard problems in hierarchical-tree clustering. *Acta Informatica*, **23**, 311–323.

Leclerc, B. (1995). Minimum spanning trees for tree metrics: abridgement and adjustments. *Journal of Classification*, **12**, 207–241.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton Univ. Press.

Saitou, N., & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–424.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, **42**, 319–347.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy*. San Francisco: Freeman.

Swofford, D. L., & Olsen, G. J. (1990). Phylogeny reconstruction. In D. M. Hillis & C. Moritz (Eds.), *Molecular systematics* (pp. 411–505). Sunderland, MA: Sinauer Associates.

Ubhaya, V. A. (1974). Isotone optimization, I, II. *Journal of Approximation Theory*, **12**, 146–159, 315–331.

Zaretskii, K. (1965). Constructing trees from the set of distances between pendant vertices. *Uspehi Matematiceskih Nauk*, **20**, 90–92.

Received: September 17, 1997