Società italiana di economia pubblica

WORKING PAPER
No 597
November 2007

WHAT'S SO SPECIAL ABOUT EUCLIDEAN DISTANCE? A CHARACTERIZATION WITH APPLICATIONS TO MOBILITY AND SPATIAL VOTING

MARCELLO D'AGOSTINO AND VALENTINO DARDANONI

JEL Classification: C0, C53, D63, D72.

Keywords: Euclidean Distance, Axiomatic Method, Mobility Measurement, Spatial Voting.

WHAT'S SO SPECIAL ABOUT EUCLIDEAN DISTANCE?

A CHARACTERIZATION WITH APPLICATIONS TO MOBILITY AND SPATIAL VOTING

MARCELLO D'AGOSTINO AND VALENTINO DARDANONI

Abstract. In this paper we provide an application-oriented characterization of a class of distance measures monotonically related to the Euclidean distance in terms of some general properties of distance functions between real-valued vectors. Our analysis hinges upon two fundamental properties of distance measures that we call "value-sensitivity" and "order-sensitivity". We show how these two general properties, combined with natural monotonicity considerations, lead to characterization results that single out several versions of Euclidean distance from the wide class of separable distance measures. We then discuss and motivate our results in two different and apparently unrelated application areas — mobility measurement and spatial voting theory — and propose our characterization as a test for deciding whether Euclidean distance (or some suitable variant) should be used in your favourite application context.

JEL Classification Numbers: C0, C53, D63, D72.

Keywords: Euclidean Distance, Axiomatic Method, Mobility Measurement, Spatial Voting.

1. Introduction

The problem of measuring the distance between real-valued vectors arises in most areas of scientific research. In particular, variants of the familiar Euclidean distance play a prominent role in many important application contexts not only in economics, statistics, political science and decision theory, but in such diverse fields as DNA sequencing, cryptography, image recognition, and so on. But what's so special about Euclidean distance? How can we judge the appropriateness of adopting this conventional distance measure in some specific application context? In what contexts are we *forced* to use it (or some monotonic transformation of its) as *the* appropriate distance measure?

In this paper we provide an application-oriented characterization of a class of functions monotonically related to the Euclidean distance in terms of five general properies which are intuitively (and perhaps empirically) testable. We show that Euclidean distance is (up to a monotonic transformation) the *only* function that satisfies them all. We also show that, by replacing some of these five properties with suitable variants, one obtains a similar characterization of an averaged version of the Euclidean distance

Date: 1 November 2007.

which we call *Averaged Euclidean Distance*, and unique characterizations of closely related distance measures that we call *Generalized Euclidean Distance* and *Averaged Generalized Euclidean Distance*:

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \qquad \text{(ED)}$$

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - y_i)^2} \qquad \text{(AED)}$$

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^n (g(x_i) - g(y_i))^2} \qquad \text{(GED)}$$

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^n (g(x_i) - g(y_i))^2} \qquad \text{(AGED)}$$

where g is a continuous and increasing function. Our results may then be helpful in testing — intuitively or empirically — whether one or the other of these metrics fits a specific application and should, therefore, be preferred to alternative measures.

Our use of the expression "distance function", throughout the paper, is informal. In several application contexts a variety of functions which do not fully satisfy the standard textbook definition have been taken into consideration and are regarded as intuitively measuring some sort of distance. So, in this paper we are not committing to any specific definition, let alone to the standard one. By "distance measure" we shall refer to any continuous function of two real-valued vectors (of the same finite size) which can be intuitively considered as measuring their distance, even if in some cases, such a function may not satisfy one or the other property of the standard definition. Thus, we shall speak of a "distance measure over $\mathbb{D}^n$" to mean simply a continuous function $d_n : \mathbb{D}^n \times \mathbb{D}^n \mapsto \mathbb{R}_+$ for some suitable interval $\mathbb{D} \subseteq \mathbb{R}$.

Our characterization hinges upon two fundamental properties that a distance measure between two real-valued vectors may satisfy. These properties deal with two basic ways in which the distance between vectors $\mathbf{x}$ and $\mathbf{y}$ may intuitively increase/decrease, namely: (a) by changing the value of one component of one of the two vectors, leaving everything else unchanged; (b) by "swapping" the values of two elements in one of the two vectors, leaving everything else unchanged.

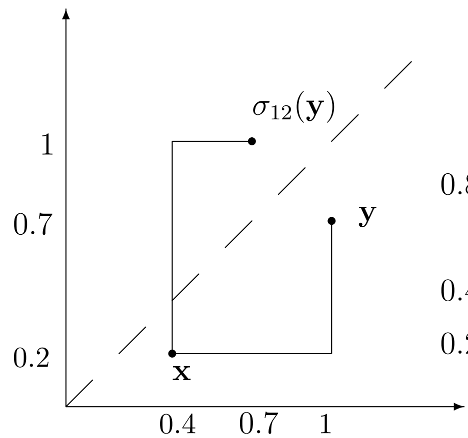
More precisely,

- we say that a distance measure $d_n : \mathbb{D}^n \times \mathbb{D}^n \mapsto \mathbb{R}_+$ (for some interval $\mathbb{D} \subseteq \mathbb{R}$) is *value-sensitive* if, for any $\mathbf{x}, \mathbf{y}, \mathbf{y}' \in \mathbb{D}^n$ such that $|x_j - y_j| < |x_j - y'_j|$ and $y_i = y'_i$ for $i \neq j$, we have

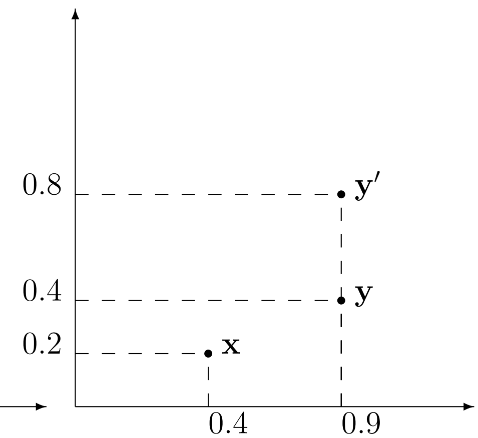
$$d_n(\mathbf{x},\mathbf{y}) < d_n(\mathbf{x},\mathbf{y}');$$

- let $\sigma_{ij}(\mathbf{u})$ denote the vector obtained from $\mathbf{u}$ by "swapping" the values of u_i and u_j; a distance measure d_n is called *order-sensitive* if, for all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$, whenever $(x_i - x_j)(y_i - y_j) > 0$ we have

$$d_n(\mathbf{x},\mathbf{y}) < d_n(\mathbf{x},\sigma_{ij}(\mathbf{y})).$$



(a) Order-sensitive



(b) Value-sensitive

FIGURE 1. Order-sensitive and value-sensitive measures

While value-sensitivity requires that the distance between two vectors should be a monotonic function of the absolute difference between their corresponding coordinates, order-sensitivity requires that it should depend on the order association between the two vectors. Suppose that, for some i and j, there is a positive order association between the corresponding coordinates, that is $(x_i - x_j)(y_i - y_j) > 0$. In this case a swap between y_i and y_j (or between x_i and x_j) is *order-reversing*, since it turns the positive association into a negative one.[1] Order-sensitivity requires that such swaps always increase the distance between the vectors under consideration.

Clearly, not all commonly used distance measures are order-sensitive. In Figure 1(a) on p. 3, it can be immediately verified that the so-called city-block distance $d_n(\mathbf{u}, \mathbf{v}) = \Sigma_{i=1}^n |u_i - v_i|$ declares $\mathbf{x}$ as equally close to $\mathbf{y}$ and $\sigma_{12}(\mathbf{y})$. On the other hand, typical order-sensitive measures, such as the Spearman and Kendall coefficients as well as any other measure based on ranks, are not value-sensitive and, for instance, would declare $\mathbf{x}$ equally close to $\mathbf{y}$ and $\mathbf{y}'$ in Figure 1(b).

Although value-sensitivity and order-sensitivity rely on different intuitions about how the distance between two vectors may change, they are by no means incompatible properties. Each member of the commonly used Minkowsky class of distance measures $d_n(\mathbf{u}, \mathbf{v}) = \left(\Sigma_{i=1}^n (u_i - v_i)^p\right)^{1/p}$ (also known as the *power metric*) is both value and order sensitive whenever $p > 1$.

In their well-known [TK70], Tverski and Krantz discuss axiomatic characterizations of various generalizations of the power metric from the point of view of measurement theory. In a different context, motivated by mobility measurement, Fields and Ok ([FO96]) and Mitra and Ok ([MO98]) also provide related characterizations of the Minkowski class. However, since there are infinitely many elements in the Minkowsky

[1] Such order-reversing swaps are discussed in mathematical statistics [Tch80], economics [ET80], and mobility measurement [Atk83, Dar93].

class, such characterizations leave entirely open the choice of the specific metric which is appropriate in each given context, as well as the characterizing properties of the chosen metric. So, their analyses do not, and cannot, assign any special role to the property that we have called "order-sensitivity", for the good reason that the Minkowski class includes, as a special case, the city-block distance which is *not* order-sensitive. One aim of this paper is to provide a sharper analysis, by identifying a set of properties that allow us to single out Euclidean distance, in one or the other version, from the wide class of separable distance measures.[2] A crucial step in this analysis is the recognition that, unlike any other Minkowsky metric, Euclidean distance is order-sensitive in a way which is monotonically related to the distance between the swapped values. This distinctive property plays a crucial role in a variety of applications and can be identified as what makes Euclidean distance so "special".

In Section 2 we present three properties (Properties 3, 4 and 5) that refine, via natural monotonicity considerations, the general definitions of value-sensitive and order-sensitive distance measures informally introduced above. We call them "monotonicity properties". We also introduce two invariance properties, Permutation Invariance and Extension Invariance (Properties 1 and 2). We also consider alternative versions of Properties 2 and 3 which may turn out to be more adequate for some applications. We then show, in Section 3, that Euclidean Distance is the only distance measure which satisfies the monotonicity properties, and the invariance properties (Theorem 1). We also show that Averaged Euclidean distance is the only distance measure that satisfies the monotonicity properties, Permutation Invariance and an alternative to Extension Invariance, called Replication invariance (Theorem 2). Finally, we show that by replacing Property 3 with a suitable alternative version in the first two theorems, one obtains a characterization of Generalized Euclidean Distance and Averaged Generalized Euclidean Distance (Theorem 3).

As a vehicle to appreciate the applicative potential of our results, we shall discuss, in Section 4, two case-studies from very different areas: *social mobility measurement* and *spatial voting theory*.[3] In case-study 1, we make a critical analysis of the characterizing properties of Euclidean Distance in the context of mobility measurement, which leads to the rejection of two of these properties on intuitive grounds. Our analysis suggests that Averaged Generalized Euclidean Distance (characterized in Theorem 3) may be more suitable for this kind of application. In case-study 2, the intuitive rejection of two characterizing properties of Euclidean Distance in the context of spatial voting theory is addressed via a different approach. This consists in showing how the intuitively judged distance between candidates, which is *prima facie* incompatible with a Euclidean metric, can be expressed in terms of a restricted "canonical" model for which all the characterizing properties of Euclidean Distance are satisfied.

[2]A similar sharper analysis, but leading to a characterization of the city-block distance, is provided by Fields and Ok in [FO96].

[3]For surveys on social mobility measurement see Maasoumi [Maa98] and Fields and Ok [FO99a]. For an introduction to the spatial theory of voting, see Hinich and Munger [HM70]; for a general advanced treatment of voting theory, Austeen-Smith and Banks [ASB99].

Potential applications of the characterization results presented in this paper, on the other hand, are by no means restricted to the ones discussed in these two case-studies, but extend to many other interesting problems that can be naturally interpreted in terms of choosing a suitable distance function between real-valued vectors.[4]

2. Properties

As mentioned in the Introduction, we shall speak of a "distance measure over $\mathbb{D}^n$" to mean simply a continuous function $d_n : \mathbb{D}^n \times \mathbb{D}^n \mapsto \mathbb{R}_+$ for some suitable interval $\mathbb{D} \subseteq \mathbb{R}$. In particular, we shall restrict our attention to three typical cases: (*i*) $\mathbb{D} = [0, a]$ for some $a \in \mathbb{R}_+$ such that $1 \leq a$; (*ii*) $\mathbb{D} = \mathbb{R}_+$; (*iii*) $\mathbb{D} = \mathbb{R}$. We notice that in such cases, given any two $x, y \in \mathbb{D}$, their absolute difference $|x - y| \in \mathbb{D}_+$; this restriction will simplify the analysis and the notation used in this paper.

In what follows we shall use the lightface letters a, b, c, d, etc. to denote arbitrary real numbers and the boldface letters $\mathbf{x}, \mathbf{y}, \mathbf{w}$, etc. to denote arbitrary vectors. The vector whose only element is the real number a will be denoted simply by "a". We shall write $[\mathbf{x}, \mathbf{y}]$ for the concatenation of the two vectors $\mathbf{x}$ and $\mathbf{y}$.[5]

2.1. Invariance Properties.

The first property of this group requires d_n to be invariant under uniform permutations of both its arguments:

Property 1 (Permutation Invariance)**.** *For all* $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$

$$d_n(\mathbf{x}, \mathbf{y}) = d_n(\pi(\mathbf{x}), \pi(\mathbf{y}))$$

for every permutation π*.*

The second property relates the distance between two n-dimensional vectors to the distance between higher-dimensional "conservative extensions". It requires that, if both vectors are extended by concatenating each of them with the same vector, their distance is left unaltered.

Property 2 (Extension Invariance)**.** *For all* $m, n \in \mathbb{N}$*, all* $\mathbf{x}, \mathbf{y} \in \mathbb{D}^m$ *and all* $\mathbf{z} \in \mathbb{D}^n$*,*

$$d_m(\mathbf{x}, \mathbf{y}) = d_{m+n}([\mathbf{x}, \mathbf{z}], [\mathbf{y}, \mathbf{z}]).$$

[4]The linear regression model, for instance, is without doubts the workhorse of theoretical and applied econometrics. Given an n-sized vector $\mathbf{y}$ ("regressand") and k n-sized vectors $\mathbf{x_1}, \ldots, \mathbf{x_k}$ ("regressors") which are collected into an $n \times k$ matrix $\mathbf{X}$, the linear regression model deals with how to find the point in the linear space spanned by the columns of $\mathbf{X}$ which is closest to $\mathbf{y}$. Thus, the problem is to find a k-sized vector $\boldsymbol{\beta}$ ("regression coefficient") which minimizes the distance between $\mathbf{y}$ and $\mathbf{X}\boldsymbol{\beta}$. The most common method for solving this problem is of course the OLS, which implies that Euclidean distance is the chosen distance concept.

In other applications, distance measures between real-valued vectors are employed to evaluate the amount of "similarity" between two objects, each of which is decomposed into a fixed number of components. (Dis)similarity is then modelled as a suitable metric in the resulting feature space. For a critical discussion of such metric models in the investigation of human similarity judgements and an alternative proposal in terms of fuzzy set theory, see [SJ99].

[5]Since conventionally vectors in $\mathbb{D}^n$ are columns, by $[\mathbf{x}, \mathbf{y}]$ formally we mean $[\mathbf{x}^T, \mathbf{y}^T]^T$.

This Property may appear intuitively sound in some application contexts but not in others, for instance when we are interested in a notion of averaged distance. For an alternative property which is appropriate in such contexts see Section 2.3 below.

2.2. **Monotonicity Properties.** The first two properties of this group articulate the value-sensitivity property informally discussed in Section 1 and refine it via monotonicity considerations. They are both rather natural assumptions to make, and similar properties arise in several characterization results, including those mentioned in the introduction, in a variety of fields.

Property 3 (One-dimensional value-sensitivity)**.** *For all* $a, b \in \mathbb{D}$, $d_1(a, b) = G(|a-b|)$ *for some continuous and strictly increasing function* $G : \mathbb{D}_+ \mapsto \mathbb{R}_+$ *such that* $G(0) = 0$.

This property makes one-dimensional distance depend monotonically on the absolute difference between the values of their single coordinate ($G(0) = 0$ is a harmless normalization requirement) and leaves entirely open the problem of how to measure higher-dimensional distance.[6] Property 3 is closely related to the one called "intradimensional subtractivity" in [TK70]. In fact, the latter corresponds to Property 3^* presented in Section 2.3.

The second property requires that the distance between two vectors is monotonically consistent with the distance between their subvectors:

Property 4 (Subvector Consistency)**.** *For all* k, j *and whenever* $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathbb{D}^k$, $\mathbf{u}, \mathbf{v}, \mathbf{u}', \mathbf{v}' \in \mathbb{D}^j$,

$$d_k(\mathbf{x}, \mathbf{y}) > d_k(\mathbf{x}', \mathbf{y}') \text{ and } d_j(\mathbf{u}, \mathbf{v}) = d_j(\mathbf{u}', \mathbf{v}') \Longrightarrow$$
$$\Longrightarrow d_{k+j}([\mathbf{x}, \mathbf{u}], [\mathbf{y}, \mathbf{v}]) > d_{k+j}([\mathbf{x}', \mathbf{u}'], [\mathbf{y}', \mathbf{v}']).$$

A similar axiom is commonly used in the literature on income inequality [Sho88], poverty [FS91] and mobility measurement [FO99b],[7] and implies (as shown in the proof of Theorem 1 provided in the Appendix, see also [FS91]) the fundamental independence assumption which plays a crucial role in the theory of additive conjoint measurement ([Deb60]). Tverski and Krantz show how to derive the latter from three more primitive axioms (see Theorem 1 in [TK70]).

The third property of this group is best understood in connection with the order-sensitivity property discussed in Section 1. It can be interpreted as requiring that the increase in the distance between two vectors caused by any "order-reversing swap" depends monotonically on the distance, in each vector, between the components that are involved in the swap:[8]

[6]Notice that it implies that one-dimensional distance satisfies the conditions for a "semimetric", namely non-negativity ($d_1(a, b) \geq 0$), symmetry ($d_1(a, b) = d_1(b, a)$), and identity of indiscernibles ($d_1(a, b) = 0$ if and only if $a = b$).

[7]Though such axioms are widely accepted in these contexts, for a critical discussion see Foster and Sen [FS97].

[8]Recall that we denote by $\sigma_{ij}(\mathbf{y})$ the vector obtained from $\mathbf{y}$ by "swapping" y_i and y_j, i.e. the vector $\mathbf{y}'$ such that: (i) $y'_k = y_k$ for all $k \neq i, j$, (ii) $y'_i = y_j$, and (iii) $y'_j = y_i$. Observe that, strictly

Property 5 (Monotonic Order-Sensitivity)**.** *For all $n \geq 4$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$ and all $i, j, k, m \in \{1, \ldots, n\}$, if*

- $(x_i - x_j)(y_i - y_j) > 0$
- $(x_k - x_m)(y_k - y_m) > 0$
- $d_1(x_i, x_j) \leq d_1(x_k, x_m)$ *and* $d_1(y_i, y_j) \leq d_1(y_k, y_m)$,

then

$$d_n(\mathbf{x}, \sigma_{ij}(\mathbf{y})) \leq d_n(\mathbf{x}, \sigma_{mk}(\mathbf{y})).$$

In other words, the effect of an order-reversing swap in the y's depends monotonically both on the distance between the swapped y's and on the distance between the corresponding x's.[9] As will be shown in the next section, this property provides a sharp means to single out Euclidean distance (or some suitable variant of its) from the much wider classes of separable distance measures. It can be verified, for instance, that while all the Minkowski distance measures with $p > 1$ are order-sensitive, none of them is monotonically order-sensitive except for the case of $p = 2$.

2.3. **Variants.** Extension Invariance (Property 2 above) may be considered inappropriate in some application contexts—such as social mobility comparisons—where we are interested in a notion of *average* distance. In this kind of applications one may want to consider, instead of Property 2, a standard "replication" property:

Property 2* (Replication Invariance)**.** $d_k(\mathbf{x}, \mathbf{y}) = d_{nk}([\overbrace{\mathbf{x}, \cdots, \mathbf{x}}^{n}], [\overbrace{\mathbf{y}, \cdots, \mathbf{y}}^{n}])$ *for every $\mathbf{x}, \mathbf{y} \in \mathbb{D}^k$ and every $n \in \mathbb{N}$, where $[\overbrace{\mathbf{u}, \cdots, \mathbf{u}}^{n}]$ denotes the result of concatenating the vector $\mathbf{u}$ with itself n times.*

As will be shown in Section 4, in some application contexts, One-dimensional Value-Sensitivity (Property 3 above) may also be rejected, on intuitive or empirical grounds, in favour of the following variant:

Property 3* (Generalized value-sensitivity)**.** *For all $a, b \in \mathbb{D}$, $d_1(a, b) = |g(a) - g(b)|$, for some continuous and strictly increasing function $g : \mathbb{R}_+ \mapsto \mathbb{D}$,*

the nature of the function g depending on the application. In Section 4.1, for instance, we argue that this variant should be preferred in the context of social mobility measurement, where g is interpreted as an "economic status" function. In the Appendix we show how this property can be derived from more primitive ones which have a natural interpretation in the context of mobility measurement. [TK70] discuss this property in connection with psychological applications and also present (Theorem 1) a different

speaking, this Property does not imply, by itself, that the distance measure be order-sensitive in the sense explained in the Introduction, that is, it does not imply that any order-reversing swap be distance-increasing. However, it does so in conjunction with the other properties, as shown in the Appendix, section 6.2.

[9]The latter must also be taken into consideration, since the pair of vectors $(\mathbf{x}, \sigma_{ij}(\mathbf{y}))$ could be seen as obtained from the pair $(\sigma_{ij}(\mathbf{x}), \sigma_{ij}(\mathbf{y}))$ by an order-reversing swap in $\sigma_{ij}(\mathbf{x})$.

derivation of it from axioms expressed in terms of a "betweennes" relation. Observe also that if Property 3 is replaced by Property 3*, *Property 5 needs to be reinterpreted* by noticing that the distance between the swapped values must be measured taking into account the function g.

3. Characterizations

Theorem 1. *A distance measure d_n satisfies Properties 1, 2, 3, 4, 5, if and only if for all n and all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$:*

$$d_n(\mathbf{x}, \mathbf{y}) = H\big[\sum_{i=1}^{n}(x_i - y_i)^2\big]$$

for some continuous and strictly increasing function $H : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that $H(0) = 0$.

The proof of this theorem is contained in the Appendix, and is based on the fact that Properties 1, 2, 3 and 4 induce a separable structure, while the addition of Property 5 induces the quadratic functional form.

Notice that the monotone function H emerging in Theorem 1 is related to the choice of an appropriate function G in Property 3, since clearly $H(t) = G(\sqrt{(t)})$. For example: (i) the standard Euclidean Distance emerges from our properties simply by letting G be the identity function; (ii) the Squared Euclidean Distance is obtained when G is the quadratic function.

Replacing Property 2 with Property 2* suffices to characterize (still up to a monotonic transformation) *AED*:

Theorem 2. *A distance measure d_n satisfies Properties 1, 2*, 3, 4, 5, if and only if for all $n \in \mathbb{N}$ and all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$,*

$$d_n(\mathbf{x}, \mathbf{y}) = H\big[\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2\big]$$

for some continuous and strictly increasing function $H : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $H(0) = 0$.

When Property 3 is replaced by its variant 3*, one obtains a characterization of *GED* and *AGED*:

Theorem 3.

1. *A distance measure d_n satisfies Properties 1, 2, 3*, 4, 5, if and only if*

$$d_n(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n}(g(x_i) - g(y_i))^2}.$$

2. *If Property 2 is replaced by Property 2*, then:*

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(g(x_i)-g(y_i))^2}.$$

It may be worth noticing that replacing Property 3 with Property 3* forces the monotonic function H in Theorems 1 and 2 to be equal to the identity function.

Theorems 1, 2 and 3 can be directly applied to justify the use of one or the other version of Euclidean distance, in all application contexts in which the relevant properties are satisfied. Typically, an *application context* C is represented by a pair $(\mathcal{S}, d)$ where $\mathcal{S}$ is the *data space* and d is some intended *intuitive distance measure* over $\mathcal{S}$, i.e. some distance measure partially specified by a set of intuitive comparative judgments of the form "$\mathbf{x}$ is closer to $\mathbf{y}$ than $\mathbf{z}$". The problem is determining the functional form that d must have in order to comply with such intuitive judgments. If the characterizing properties are satisfied by the intuitive judgments which can be obtained in the application context C – and may sometimes be revealed experimentally – then our theorems dictate (up to a monotonic transformation) the functional form of the distance measure: Euclidean Distance (Theorem 1), Averaged Euclidean Distance (Theorem 2), Generalized Euclidean Distance (Theorem 3.1) or Averaged Generalized Euclidean Distance (Theorem 3.2).

4. Applications: two case-studies

4.1. **Case study 1: social mobility measurement.** When discussing mobility issues, a basic distinction is usually made between intergenerational and intragenerational mobility. The first concept concerns the study of how the distribution of some relevant measure of individual status changes between different generations in a given society. Alternatively, intragenerational mobility studies how the distribution of individual status changes among a group of individuals over a given period of their lifetime. In general, the simplest framework to capture either of these aspects is to consider how, in a society of n individuals, a vector $\mathbf{x}$ is transformed into another vector $\mathbf{y}$, where the i-th element x_i denotes the value of a relevant indicator of the social and economic status of individual i, and y_i denotes its value in the next generation (intergenerational case) or in the next time period (intragenerational case). Typical variables employed in most mobility studies for measuring socioeconomic status are income, wage, consumption, education, and occupational prestige. Focusing on intergenerational income mobility, a *social mobility index* is then a function from $\mathbb{R}^n_+ \times \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ which can be naturally interpreted as a distance function between fathers' and sons' incomes. So, in this application context the data space is $\mathbb{R}^n_+ \times \mathbb{R}^n_+$, where each vector represents social status of individuals in a society, and the intended intuitive distance d is some distance measure that complies with intuitive mobility comparisons.

In the context of mobility measurement, Property 1 seem unexceptionable. Property 4 is also almost invariably assumed in the mobility measurement literature where it is known as "subgroup consistency". On the other hand, Property 2 is quite inappropriate if one believes that social mobility in a society should be measured in *per-capita* terms. So, Property 2 can be dropped and replaced by the Replication Invariance property (Property 2*) which led us to Theorem 2.[10] We will then consider the *per-capita* Euclidean distance as a good candidate for the sound distance measure in the social mobility context, and we will focus then on the remaining properties, namely 3 and 5.

If the domain of the mobility index are dollar-incomes, as often assumed, it may be unreasonable to postulate that a father-son movement, say, between \$100 and \$110 has the same level of mobility than a movement between \$1000 and \$1010, as implied by Property 3. If one interprets a social mobility index as measuring the distance between the economic status of two generations, this kind of objection would lead to rejecting the identification of economic status with dollar-income, and suggests replacing Property 3 with an analogous property which is not subject to this criticism. As a replacement of Property 3 consider then Property 3*, where g is interpreted as an appropriate function measuring "economic status" whose form is application-dependent (but see below for a brief discussion). Notice that this property implies the symmetry of upward and downward mobility. As remarked by Fields and Ok [FO99b], this symmetry is unexceptionable if one does not distinguish between "good" and "bad" movements of income, that is, one is not motivated by welfaristic concerns. Then, Theorem 3, implies that $AGED$ may be a good choice in this application context. Compare this distance measure with the "city-block" and Minkowsky distance measures applied to dollar-incomes, and the "city-block" measure applied to log-dollar incomes, proposed in the seminal characterizations of mobility indices ([FO96], [MO98] and [FO99b]).

As for Property 5, observe that once Property 3 has been replaced by Property 3*, its interpretation inherits the consideration of economic status which is incorporated into the latter. As a consequence, the distance between the swapped values is measured taking into account the effect of the status function g. In particular, observe that, if g is a concave function (such as, for example the log) the distance between two swapped values in the application of Property 5 will be smaller for incomes of high range even when the difference between their absolute values is the same. Having clarified this, Property 5 seems a natural monotonic extension of the order-reversing swap property much discussed in the mobility literature.

Applications of this theorem depend on the choice of an appropriate status function g. Some may argue that a good choice would be, for example, to take g as the log function, so that an income movement from \$1000 to \$2000 would have the same mobility effect as a movement from \$100 to \$200. In general, the most appropriate choice of g depends on the application context and must be justified independently. Property 3* (and, as a consequence, Property 5), can therefore be seen as a general

[10]On the other hand, the Replication Invariance Property too may be criticized in mobility applications. For example, one may legitimately perceive less mobility in a society with two families with incomes, say, $(1, 2)$ and $(2, 1)$ than in a society which replicates these two families a million times. If this view is strongly held, our theorem implies that averaged euclidean distance is not a good candidate for mobility measurement.

constraint on this choice.[11] In the Appendix we discuss how Property 3* can be derived by three more primitive properties which seem rather natural in the context of mobility measurement.

4.2. **Case study 2: multidimensional spatial voting.** The standard univariate spatial theory of voting assumes the existence of a *policy space* $\mathcal{P}$ which is typically an interval $[0, a]$, such that different alternatives can be represented as points in $\mathcal{P}$. A *multidimensional policy space* can then be represented as the Cartesian product $\mathcal{P}^n$, such that each issue has a well-defined unit of measurement which is shared by all voters, and the different alternatives over which voters are assumed to vote are elements of $\mathcal{P}^n$ (real-valued n-dimensional vectors). A voter's preferences are then characterized by an *ideal point* in $\mathcal{P}^n$ and a distance measure on $\mathcal{P}^n$.

Given an ideal point $\mathbf{x} \in \mathcal{P}^n$, the Euclidean distance induces a preference ordering $\succeq_{\mathbf{x}}$ on $\mathcal{P}^n$ such that

$$\mathbf{y} \succeq_{\mathbf{x}} \mathbf{z} \Leftrightarrow -\sum_{i=1}^{n}(x_i - y_i)^2 \geq -\sum_{i=1}^{n}(x_i - z_i)^2$$

that is, the individual will vote for $\mathbf{y}$ whenever $\mathbf{y}$ is "closer" to her ideal point $\mathbf{x}$ than $\mathbf{z}$. For illustration, in the two-dimensional case the ordering can be represented by the utility function $U_{\mathbf{x}}(\mathbf{y}) = -(x_1 - y_1)^2 - (x_2 - y_2)^2$, and the indifference curves are circles centered at (x_1, x_2). Theorem 1 shows that using Euclidean preferences in a given context is equivalent to deeming its characterizing properties to comply with intuitive judgments in that particular context. Here the data space is $[0, a]^n \times [0, a]^n$, and an intuitive distance function measures the "closeness" between two political platforms.

In multidimensional spatial voting, Property 3 is indeed uncontroversial since it can be seen as the Blackian and Downsian starting point for any multidimensional extension. Similarly, Properties 4 and 2 seem natural properties to impose in this context. However, Property 5 and 1 appears to be intuitively sound *only* in situations where different issues are regarded as equally important from the voter's viewpoint. Hence, while Theorem 1 justifies the use of Euclidean distance in all such restricted situations, it can be argued that the restriction is quite unrealistic, since different issues are often given different *saliency*.

When issues are assumed to have different saliency, the usual approach consists in "weighting" different issues by means of a real number expressing the relative importance assigned to them by an individual voter. Given a set of positive weights $w_1, \ldots, w_n$, *weighted Euclidean preferences* can then be represented by a utility function $U_{\mathbf{x},\mathbf{w}}(\mathbf{y}) = -\sum_{i=1}^{n} w_i(x_i - y_i)^2$. It is easy to see that in the bidimensional case indifference curves are *ellipses* centered at the ideal point $\mathbf{x}$, parallel to the horizontal/vertical axis depending on whether w_1 is lower/greater than w_2. It is clear that, given an ideal point $\mathbf{x}$, it is easy to construct examples where, say, $\mathbf{y}$ is preferred to $\mathbf{z}$ by weighted Euclidean preferences but not by simple ones.

[11] Linking the concept of economic status to individual welfare, by interpreting g as an empirically revealed utility function, could help this process.

The intuition underlying the use of a weighted distance measure can be expressed more precisely as follows. Let V be the set of all possible voters. If issues have different saliency for different voters, the distance function changes from voter to voter, and so we can denote by d_n^v the distance function of voter v. Among all possible voters, there are some, that we call *the canonical voters*, for whom all the different issues have the same saliency. Let V_0 be the subset of V consisting of the canonical voters. Using a weighted distance measure (whatever it may be) means that the distance of a candidate $\mathbf{y}$ from the ideal candidate $\mathbf{x}$ for a given voter v is equal to the distance between $\mathbf{a}(\mathbf{y})$ and $\mathbf{a}(\mathbf{x})$ for a *canonical* voter u, where $\mathbf{a}$ is a standardization function that multiplies every coordinate of the vector to which it applies for the "weight" assigned to the corresponding issue.[12] In this way the preferences of each non-canonical voter v can be represented in terms of the preferences of a canonical voter u. So, by connecting the preferences of non-canonical voters to those of a canonical voter, one may obtain a precise mathematical expression which represents the former in terms of the latter. The advantage of this approach is that the distance function of canonical voters may be determined by means of some known characterization. To summarize, the assumption underlying the use of weighted distance functions is the following

(1) *For all $v \in V$ and all $u \in V_0$, there are $\alpha_{v.1}, \ldots, \alpha_{v.n} \in \mathbb{R}_+$ such that $d_n^v(\mathbf{x}, \mathbf{y}) = d_n^u(\mathbf{a_v}(\mathbf{x}), \mathbf{a_v}(\mathbf{y}))$,*

where $\mathbf{a_v}(\mathbf{z}) = \alpha_{v.1}z_1, \ldots, \alpha_{v.n}z_n$.

Now, this assumption leaves open how the distance function of a canonical voter should be appropriately determined. Why should weighted *Euclidean* distance be applied and not, for instance, weighted city-block distance? Weighting, as explained before, is just a way of standardizing the elements of different policy spaces into elements of a reference policy space (that of a canonical voter) and tells us nothing about the appropriate distance function for this reference space. Now, in the restricted domain of canonical voters we have already argued that the distance function satisfies Property 1 and Property 5. Moreover, Properties 3, 2 and 4 are intuitively sound for all d_n^v, no matter whether v is canonical or not. Hence, we can assume that all properties of Theorem 1 are satisfied by d_n^u, for any canonical voter u. Therefore:

C1: The distance function d_n^u of any canonical voter u satisfies 1, 2, 3, 4 and 5,

C2: For all $v \in V$ and all $u \in V_0$, there are $\alpha_{v.1}, \ldots, \alpha_{v.n} \in \mathbb{R}_+$ such that $d_n^v(\mathbf{x}, \mathbf{y}) = d_n^u(\mathbf{a_v}(\mathbf{x}), \mathbf{a_v}(\mathbf{y}))$,

hold true if and only if, for every $v \in V$,

$$d_n^v(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} w_i(x_i - y_i)^2},$$

with $w_i = \alpha_{v.i}^2$.

[12]Such weights can be determined empirically by observing the indifference curves of each voter or group of voters.

For, by Theorem 1, d_n^u must be equal to the Euclidean distance for every canonical voter u, and so, by *C1* and *C2*:

$$d_n^v(\mathbf{x},\mathbf{y}) = d_n^u\big(\mathbf{a_v}(\mathbf{x}),\mathbf{a_v}(\mathbf{y})\big) = \sqrt{\sum_{i=1}^{n}\big(a_{v,i}x_i - a_{v,i}y_i\big)^2}.$$

Thus, d_n^v must be equal to the *weighted Euclidean distance*, with the weight for issue i given by $\alpha_{v.i}^2$.

We conclude this section by stressing that this result may be helpful in guiding empirical research. For example [EMR88] consider weighted Euclidean distance and weighted city-block distance to determine which is empirically closer to voting behavior. By means of our characterization, empirical testing can be performed separately on each of the characterizing properties of the distance function.

5. Conclusions

As should emerge from our discussion, the results presented in Section 3 may be useful in understanding:

- whether some suitable monotonic transformation of Euclidean Distance, in one of the variants investigated in this paper, naturally fits a given application context, by checking whether its characterizing properties are satisfied in it (case study 1);
- when the characterization results cannot be directly applied, how the original application context can be reduced to a "canonical" one in which the properties of some of the characterized distance measures are all satisfied (case study 2).

Our analysis suggests that some variant of the Euclidean distance is likely to be appropriate in many contexts requiring a distance measure which is both monotonically value-sensitive (as made precise by Properties 3/3* and 4) and monotonically order-sensitive (as made precise by Property 5).

6. Appendix

6.1. **Proof of Theorem 1.** To prove the theorem, we first need the following

Lemma 1. *A distance measure d_n satisfies Properties 1, 2, 3, 4 if and only if there exists a continuous and strictly increasing function $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$, with $f(0) = 0$, such that for all n and all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$,*

$$d_n(\mathbf{x},\mathbf{y}) = H\big[\sum_{i=1}^{n} f(|x_i - y_i|)\big],$$

for some continuous and strictly increasing function $H : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $H(0) = 0$.

6.1.1. *Proof of Lemma 1.* The "if" direction is left to the reader. For the "only if" direction, let us suppose that a distance measure d satisfies Properties 1 and 2, 3, 4. First, observe that Property 4, given Permutation Invariance, is equivalent to the following:

$$d_{k+j}([\mathbf{x},\mathbf{u}],[\mathbf{y},\mathbf{v}]) \leq d_{k+j}([\mathbf{x}',\mathbf{u}'],[\mathbf{y}',\mathbf{v}']) \Longleftrightarrow d_k(\mathbf{x},\mathbf{y}) \leq d_k(\mathbf{x}',\mathbf{y}'), \quad \text{whenever } d_j(\mathbf{u},\mathbf{v}) = d_j(\mathbf{u}',\mathbf{v}'), \tag{2}$$

for all k, j, all $\mathbf{x},\mathbf{y},\mathbf{x}',\mathbf{y}' \in \mathbb{D}^k$ and all $\mathbf{u},\mathbf{v},\mathbf{u}',\mathbf{v}' \in \mathbb{D}^j$.

It can be easily verified that Property 4 is implied by (2). To see the converse, notice that under Permutation Invariance (2) is equivalent to:

$$d_{k+j}([\mathbf{x},\mathbf{u}],[\mathbf{y},\mathbf{v}]) \leq d_{k+j}([\mathbf{x}',\mathbf{u}],[\mathbf{y}',\mathbf{v}]) \Longleftrightarrow d_k(\mathbf{x},\mathbf{y}) \leq d_k(\mathbf{x}',\mathbf{y}'), \tag{3}$$

for all k, j, all $\mathbf{x},\mathbf{y},\mathbf{x}',\mathbf{y}' \in \mathbb{D}^k$ and all $\mathbf{u},\mathbf{v} \in \mathbb{D}^j$. We then show that, under Permutation Invariance, Property 4 implies (3), and therefore also (2).

Suppose first that $d_{k+j}([\mathbf{x},\mathbf{u}],[\mathbf{y},\mathbf{v}]) \leq d_{k+j}([\mathbf{x}',\mathbf{u}],[\mathbf{y}',\mathbf{v}])$; then by Property 4 and contrapposition, we have that $d_k(\mathbf{x},\mathbf{y}) \leq d_k(\mathbf{x}',\mathbf{y}')$. On the other hand, suppose that

$$(i)\ d_k(\mathbf{x},\mathbf{y}) \leq d_k(\mathbf{x}',\mathbf{y}') \text{ and } (ii)\ d_{k+j}([\mathbf{x},\mathbf{u}],[\mathbf{y},\mathbf{v}]) > d_{k+j}([\mathbf{x}',\mathbf{u}],[\mathbf{y}',\mathbf{v}]).$$

Now, if $d_k(\mathbf{x},\mathbf{y}) < d_k(\mathbf{x}',\mathbf{y}')$, by Property 4, $d_{k+j}([\mathbf{x},\mathbf{u}],[\mathbf{y},\mathbf{v}]) < d_{k+j}([\mathbf{x}',\mathbf{u}],[\mathbf{y}',\mathbf{v}])$, against the assumption *(ii)*. If $d_k(\mathbf{x},\mathbf{y}) = d_k(\mathbf{x}',\mathbf{y}')$, it follows from *(ii)*, by Property 4 again, that $d_{2k+j}([\mathbf{x},\mathbf{u},\mathbf{x}'],[\mathbf{y},\mathbf{v},\mathbf{y}']) > d_{2k+j}([\mathbf{x}',\mathbf{u},\mathbf{x}],[\mathbf{y}',\mathbf{v},\mathbf{y}])$ against Permutation Invariance. Hence, if $d_k(\mathbf{x},\mathbf{y}) \leq d_k(\mathbf{x}',\mathbf{y}')$ it must hold true that $d_{k+j}([\mathbf{x},\mathbf{u}],[\mathbf{y},\mathbf{v}]) \leq d_{k+j}([\mathbf{x}',\mathbf{u}],[\mathbf{y}',\mathbf{v}])$.

Now, observe that Property 3 implies

$$d_1(a,b) = d_1(|a-b|, 0). \tag{4}$$

It follows from (2) that, for all $n \in \mathbb{N}$ and all $\mathbf{x},\mathbf{y},\mathbf{w},\mathbf{z} \in \mathbb{D}^n$,

$$d_1(x_i,y_i) = d_1(w_i,z_i) \text{ for all } i = 1,\cdots,n, \Longrightarrow d_n(\mathbf{x},\mathbf{y}) = d_n(\mathbf{w},\mathbf{z}). \tag{5}$$

Therefore, given (4) above, we have that for all n and all $\mathbf{x},\mathbf{y} \in \mathbb{D}^n$,

$$d_n(\mathbf{x},\mathbf{y}) = d_n([|x_1-y_1|,\ldots,|x_n-y_n|],[0,\ldots,0]). \tag{6}$$

Let $M_n : \mathbb{D}^n_+ \mapsto \mathbb{R}$ be defined as follows

$$M_n(\mathbf{z}) = d_n([z_1,\ldots,z_n],[0,\ldots,0]).$$

Now, since d_n is assumed to be continuous, M_n must also be continuous. Moreover, given Property 3, M_n must be strictly increasing in each argument. From (2) it also follows that, for all $\mathbf{u},\mathbf{u}' \in \mathbb{D}^h_+$ and $\mathbf{v},\mathbf{v}' \in \mathbb{D}^k_+$, with $h+k=n$,

$$M_n([\mathbf{u},\mathbf{v}]) \geq M_n([\mathbf{u}',\mathbf{v}]) \Rightarrow M_n([\mathbf{u},\mathbf{v}']) \geq M_n([\mathbf{u}',\mathbf{v}']).$$

So, taking into account Property 1, one can apply Gorman's separability Theorem [Gor68], to show that, for all $n \geq 3$, the function M_n must be separable,[13] that is, there

[13] See [SS02] for a recent thorough discussion on separable preferences.

must exist continuous and strictly increasing functions $F_n : \mathbb{R} \mapsto \mathbb{R}$ and $f_n : \mathbb{D}_+ \mapsto \mathbb{R}$ such that:

$$M_n(\mathbf{z}) = F_n\big[\sum_{i=1}^{n} f_n(z_i)\big].$$

Thus, by (6) and taking $z_i = |x_i - y_i|$, we have that for all $i = 1, \ldots, n$:

$$d_n(\mathbf{x}, \mathbf{y}) = F_n\big[\sum_{i=1}^{n} f_n(|x_i - y_i|)\big]. \tag{7}$$

We can assume, without loss of generality, that for all n, $f_n(0) = 0$. For, whenever $f_n(0) = c \neq 0$, let $h_n : \mathbb{D}_+ \mapsto \mathbb{R}$ and $H_n : \mathbb{R} \mapsto \mathbb{R}$ be defined as follows:

$$h_n(t) = f_n(t) - c \qquad H_n(t) = F_n(t - nc).$$

Then, it is immediately verified that

$$F_n\big[\sum_{i=1}^{n} f_n(|x_i - y_i|)\big] = H_n\big[\sum_{i=1}^{n} h_n(|x_i - y_i|)\big]. \tag{8}$$

Hence, from now on, we shall assume that $f_n(0) = 0$ for all n; notice that since f_n is increasing and $\mathrm{dom}(f_n) = \mathbb{D}_+ \subseteq \mathbb{R}_+$, we have $\mathrm{ran}(f_n) = \mathbb{R}_+$.

Properties 2 and 3 imply that $F_n = G \circ f_n^{-1}$. To see this, just notice that for all $\mathbf{x} \in \mathbb{D}^{n-1}$,

$$d_1(a, b) = d_n([a, \mathbf{x}], [b, \mathbf{x}]) = F_n\big[f_n(|a - b|) + \sum_{i=1}^{n-1} f_n(|x_i - x_i|)\big],$$

and, under the assumption that $f_n(0) = 0$, we have that for all $n \geq 3$,

$$d_1(a, b) = d_n([a, \mathbf{x}], [b, \mathbf{x}]) = F_n[f_n(|a - b|)].$$

Since, by Property 3, $d_1(a, b) = G(|a - b|)$, we have that $G(|a - b|) = F_n[f_n(|a - b|)]$. Let $t = f_n(|a - b|)$. Then $G(f_n^{-1}(t)) = F_n(t)$ and $F_n = G \circ f_n^{-1}$.

We have therefore established that for all $n \geq 3$ ad all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$,

$$d_n(\mathbf{x}, \mathbf{y}) = G\left[f_n^{-1}\big(\sum_{i=1}^{n} f_n(|x_i - y_i|)\big)\right], \tag{9}$$

for some continuous and strictly increasing $f_n : \mathbb{D}_+ \mapsto \mathbb{R}_+$ (such that $f_n(0) = 0$) and $G : \mathbb{D}_+ \mapsto \mathbb{R}_+$.

We now show that Property 2 allows us to choose the functions f_n to be independent of n (i.e. such that for all $n \in \mathbb{N}$, $f_n = f$, for some fixed f) and to extend (9) to the cases where $n < 3$.

Property 2 implies immediately that for every $a, b, c, d \in \mathbb{D}$, all $n \geq 3$ and all $\mathbf{x} \in \mathbb{D}^{n-2}$,

$$d_2([a, b], [c, d]) = d_n([a, b, \mathbf{x}], [c, d, \mathbf{x}]). \tag{10}$$

Recalling that for every $n \geq 3$, $f_n(0) = 0$, and that G is strictly increasing, it follows from (10) and (9) that, for every $n, m \geq 3$ and every $x, y \in \mathbb{D}$,

$$f_n^{-1}(f_n(x) + f_n(y)) = f_m^{-1}(f_m(x) + f_m(y)), \tag{11}$$

and therefore:

$$f_n(x) + f_n(y) = f_n\big[f_m^{-1}(f_m(x) + f_m(y))\big]. \tag{12}$$

The above condition implies that for every $n, m \geq 3$, $f_n(x) = \alpha f_m(x)$ for some constant α. For, let $u = f_m(x)$ and $v = f_m(y)$; then:

$$f_n(f_m^{-1}(u)) + f_n(f_m^{-1}(v)) = f_n(f_m^{-1}(u + v)). \tag{13}$$

Let $h(t) = f_n(f_m^{-1}(t))$. Then (13) can be rewritten as:

$$h(u) + h(v) = h(u + v). \tag{14}$$

But (14) implies that $h(t) = \alpha t$ for some constant α.[14] So, taking $s = f_m^{-1}(t)$,

$$f_n(s) = \alpha f_m(s), \tag{15}$$

for some constant α depending on m and n. Now, let f be equal to f_m for some fixed m (say, $f = f_3$) and let c be the appropriate constant satisfying (15) for this fixed m and a given arbitrary n. Since, by (15), $f_n^{-1}(s) = f^{-1}(s/c)$, it follows that for all $n \geq 3$:

$$\begin{aligned} f_n^{-1}\big[\sum_{i=1}^{n} f_n(|x_i - y_i|)\big] &= f_n^{-1}\big[c\sum_{i=1}^{n} f(|x_i - y_i|)\big] \\ &= f^{-1}\big[\sum_{i=1}^{n} f(|x_i - y_i|)\big]. \end{aligned} \tag{16}$$

To conclude the proof observe that, when $\mathbf{x}$ and $\mathbf{y}$ have size $k < 3$, by Property 2 and given that $f(0) = 0$, it follows that for any $\mathbf{z} \in \mathbb{D}^m$:

$$\begin{aligned} d_k(\mathbf{x}, \mathbf{y}) &= d_{k+m}([\mathbf{x}, \mathbf{z}], [\mathbf{y}, \mathbf{z}]) \\ &= G\Bigg[f^{-1}\big(\sum_{i=1}^{k} f(|x_i - y_i|) + \sum_{j=k+1}^{k+m} f(|z_j - z_j|)\big)\Bigg] \\ &= G\Bigg[f^{-1}\big(\sum_{i=1}^{k} f(|x_i - y_i|)\big)\Bigg], \end{aligned} \tag{17}$$

and so the identity

$$d_n(\mathbf{x}, \mathbf{y}) = G\Bigg[f^{-1}\big(\sum_{i=1}^{n} f(|x_i - y_i|)\big)\Bigg] \tag{18}$$

holds for all $n \in \mathbb{N}$. Given that G and f^{-1} are both continuous and strictly increasing, $H = G \circ f^{-1}$ also is. Notice that, since $G(0) = f(0) = 0$, $H(0) = 0$. This concludes the proof of Lemma 1. □

[14] (14) is a Cauchy equation of the first kind whose solution is given for example in Aczel [Acz66].

Given the above lemma, to prove the "if" direction of Theorem 1 it is sufficient to verify that any strictly increasing function of the Euclidean distance satisfies Property 5, that is, it is sufficient to verify that, whenever $(x_i - x_j)(y_i - y_j) > 0$, $(x_k - x_m)(y_k - y_m) > 0$, $d_1(x_i, x_j) \leq d_1(x_k, x_m)$ and $d_1(y_i, y_j) \leq d_1(y_k, y_m)$, we have

$$(x_i - y_j)^2 + (x_j - y_i)^2 + (x_k - y_k)^2 + (x_m - y_m)^2 \leq \\ \leq (x_i - y_i)^2 + (x_j - y_j)^2 + (x_k - y_m)^2 + (x_m - y_k)^2. \tag{19}$$

After simplification, this equation becomes

$$(x_m - x_k)(y_m - y_k) \geq (x_j - x_i)(y_j - y_i),$$

and the result follows. To prove the "only if" direction, we need to show that the function f in Lemma 1 must be quadratic. Let $a, b, c \in \mathbb{D}$ be arbitrary non negative real numbers with $a \geq c$. Let also $\mathbf{x}, \mathbf{y} \in \mathbb{D}^4$ be such that

- $\mathbf{x} = [a, (a + c), a, (a + c)]$
- $\mathbf{y} = [0, c, a, (a + c)]$

so that

- $\sigma_{12}(\mathbf{y}) = [c, 0, a, (a + c)]$
- $\sigma_{34}(\mathbf{y}) = [0, c, (a + c), a]$.

Let $\sigma_{12}(\mathbf{y}) = \mathbf{w}$ and $\sigma_{34}(\mathbf{y}) = \mathbf{z}$. Since the conditions of Property 5 are satisfied, it follows that $d_n(\mathbf{x}, \mathbf{w}) = d_n(\mathbf{x}, \mathbf{z})$; hence, we have that:

$$H\Big(\sum_{i=1}^{n} f(|x_i - w_i|)\Big) = H\Big(\sum_{i=1}^{n} f(|x_i - z_i|)\Big)$$

and therefore, since H is one-to-one

$$\sum_{i=1}^{n} f(|x_i - w_i|) = \sum_{i=1}^{n} f(|x_i - z_i|).$$

Hence, subtracting $\sum_{i=1}^{n} f(|x_i - y_i|)$ from both sides, we obtain

$$\sum_{i=1}^{n} f(|x_i - w_i|) - \sum_{i=1}^{n} f(|x_i - y_i|) = f(a + c) + f(a - c) - 2f(a)$$

and

$$\sum_{i=1}^{n} f(|x_i - z_i|) - \sum_{i=1}^{n} f(|x_i - y_i|) = 2f(c) - 2f(0).$$

So that, recalling that $f(0) = 0$, we must have for all $a \geq c \geq 0$:

$$f(a + c) - f(a) = f(a) - f(a - c) + 2f(c)$$

This functional equation has a unique solution $f(t) = \alpha t^2$ for some constant $\alpha > 0$.

To see this, consider any sequence $x_m = mc$ with $m \in \mathbb{N}_+$ and $mc \in \mathbb{D}$. Let us first evaluate the difference between two adjacent terms of the sequence $f(x_m)$:

$$\begin{aligned} f(mc) - f((m-1)c) &= f((m-1)c) - f((m-2)c) + 2(f(c)) \\ &= f((m-2)c) - f((m-3)c) + 2(f(c)) + 2(f(c)) \\ &\vdots \\ &= f(c) + 2(m-1)(f(c)) \\ &= (2m-1)(f(c)) \end{aligned}$$

Now, notice that:

$$\begin{aligned} f(mc) &= \sum_{i=1}^{m} \big(f(ic) - f((i-1)c)\big) \\ &= \sum_{i=1}^{m} (2i-1)(f(c)) \\ &= 2\frac{m^2+m}{2}(f(c)) - m(f(c)). \end{aligned}$$

So, we obtain the following functional equation:

$$f(mc) = m^2 f(c). \tag{20}$$

Let $c = \frac{1}{n}$ for some $n \in \mathbb{N}_+$. Then:

$$f\Big(\frac{m}{n}\Big) = m^2 f\Big(\frac{1}{n}\Big). \tag{21}$$

Letting $m = n$, we obtain

$$f(1) = n^2 f\Big(\frac{1}{n}\Big). \tag{22}$$

Hence, by (21),

$$f\Big(\frac{m}{n}\Big) = f(1)\frac{m^2}{n^2}, \tag{23}$$

for all rational $\frac{m}{n} \in \mathbb{D}$. Now $f(1) = \alpha$ is a strictly positive constant, since f is strictly increasing. Thus, by the continuity of f it follows that, for all $t \in \mathbb{D}$:

$$f(t) = \alpha t^2 \tag{24}$$

for some constant $\alpha > 0$, and $f^{-1}(u) = \sqrt{\frac{u}{\alpha}}$. This concludes the proof.

6.2. **Proof that Properties 1–5 imply order sensitivity.** Let $\mathbf{y}'$ be the vector obtained from $\mathbf{y}$ by "swapping" y_i and y_j, i.e. (i) $y'_k = y_k$ for all $k \neq i, j$, (ii) $y'_i = y_j$, and (iii) $y'_j = y_i$, and assume $x_j = x_i + \Delta_x$ and $y_j = y_i + \Delta_y$, with Δ_x and Δ_y having the same sign (i.e. the swap is order-reversing). After the swap, given Theorem 1, distance will increase whenever:

$$(x_i - y_i - \Delta_y)^2 + (x_i + \Delta_x - y_i)^2 > (x_i - y_i)^2 + (x_i + \Delta_x - y_i - \Delta_y)^2;$$

the above inequality is easily seen to be true since $\Delta_x \Delta_y$ is always positive.

6.3. **Proof of Theorem 2.** To prove the theorem, just replace Lemma 1 with the following:

Lemma 2. *A distance measure d_n satisfies Properties 1, 2*, 3, 4, if and only if there exists a continuous and strictly increasing function $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$, with $f(0) = 0$, such that for all n and all $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$,*

$$d_n(\mathbf{x}, \mathbf{y}) = H\big[\frac{1}{n}\sum_{i=1}^{n} f(|x_i - y_i|)\big],$$

for some continuous and strictly increasing function $H : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $H(0) = 0$.

Its proof is equal to that of Lemma 1 up to equation (8), and thereafter continues as follows.

Let $g_n = n f_n$, so that

$$d_n(|x_1 - y_1|, \cdots, |x_n - y_n|) = F_n[\sum_{i=1}^{n} \frac{1}{n} g_n(|x_i - y_i|)]. \tag{25}$$

Then, following the derivation of equation (21) from equation (14) in Foster and Shorrocks [FS91], we have that Property 2* allows us to choose the functions F_n and g_n to be independent of n (i.e. such that for all $m, n \in \mathbb{N}$, $F_m = F_n = H$ and $g_m = g_n = f$, for some fixed H and f) and to extend (25) to the case of $n < 3$. Thus $d_n(\mathbf{x}, \mathbf{y}) = H\big[\frac{1}{n}\sum_{i=1}^{n} f(|x_i - y_i|)\big]$ holds for all $n \in \mathbb{N}$.

Then, a proof of Theorem 2 is easily obtained by using Lemma 2 instead of Lemma 1 and adapting the proof of Theorem 1 accordingly.

6.4. **Proof of Theorem 3.** It is left to the reader to verify that, in both statements, the function on the right-hand side satisfies all the relevant properties. Let us concentrate on showing that it is the only function satisfying them. A proof could be obtained using a similar argument, based on separability properties, as the one used in the proof of Theorems 1 and 2. Here we provide a simpler proof which uses Theorem 1 and 2 as lemmas.

First, (2) implies that for all $n \in \mathbb{N}$ and all $\mathbf{x}, \mathbf{y}, \mathbf{w}, \mathbf{z} \in \mathbb{D}^n$,

$$d_1(x_i, y_i) = d_1(w_i, z_i) \text{ for all } i = 1, \cdots, n, \Longrightarrow d_n(\mathbf{x}, \mathbf{y}) = d_n(\mathbf{w}, \mathbf{z}). \tag{26}$$

This means that

$$d_n(\mathbf{x},\mathbf{y}) = F_n(d_1(x_1,y_1),\ldots,d_1(x_n,y_n)), \tag{27}$$

for some function F_n, and therefore

$$d_n(\mathbf{x},\mathbf{y}) = F_n(|g(x_1)-g(y_1)|,\ldots,|g(x_n)-g(y_n)|). \tag{28}$$

Moreover, it also follows from (2), that F_n is one-to-one in each argument.

Now, let d'_n be defined as follows:

$$d'_n(\mathbf{x},\mathbf{y}) = d_n(\mathbf{g}^{-1}(\mathbf{x}),\mathbf{g}^{-1}(\mathbf{y})) = F_n(|x_1-y_1|,\ldots,|x_n-y_n|), \tag{29}$$

where $\mathbf{g}^{-1}(\mathbf{z})$ stands for the vector $[g^{-1}(z_1),\ldots,g^{-1}(z_n)]$. Clearly, d'_n satisfies Properties 4, 1 and 2 (or 2*) whenever d_n does. Moreover, it is not difficult to verify that, whenever d_n satisfies Properties 3* and 5, d'_n satisfies Properties 3 and 5, with the function G of Property 3 equal to the identity function.

First, consider Property 3*. By (28), $d_1(a,b) = F_1(|g(a)-g(b)|)$, and so $d_1(a,b) = |g(a)-g(b)|$ (Property 3*) whenever F_1 is the identity function. Hence by (29), $d'_1(a,b) = |a-b|$ (Property 3 with G equal to the identity function). As for Property 5, first suppose that d_n satisfies it. It follows that, whenever the conditions of Property 5 are satisfied, it must hold true that:

$$d'_n(\mathbf{x},\mathbf{y}) = d_n\big(\mathbf{g}^{-1}(\mathbf{x}),\mathbf{g}^{-1}(\mathbf{y})\big) \le d_n\big(\mathbf{g}^{-1}(\mathbf{x}),\sigma_{ij}(\mathbf{g}^{-1}(\mathbf{y}))\big) = d'_n\big(\mathbf{x},\sigma_{ij}(\mathbf{y})\big)$$

and

$$d'_n\big(\mathbf{x},\sigma_{ij}(\mathbf{y})\big) = d_n\big(\mathbf{g}^{-1}(\mathbf{x}),\sigma_{ij}(\mathbf{g}^{-1}(\mathbf{y}))\big) \le d_n\big(\mathbf{g}^{-1}(\mathbf{x}),\sigma_{mk}(\mathbf{g}^{-1}(\mathbf{y}))\big) = d'_n\big(\mathbf{x},\sigma_{mk}(\mathbf{y})\big).$$

This implies that d'_n satisfies Property 5.

Therefore, whenever d_n satisfies all the properties of Theorem 3.1, d'_n satisfies all the properties of Theorem 1, and we can apply this theorem to establish that d'_n must be the Euclidean Distance. Thus:

$$F_n(t_1,\ldots,t_n) = \sqrt{\sum_{i=1}^{n} t_i^2}$$

and, by (28),

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(g(x_i)-g(y_i))^2}.$$

Similarly, whenever d_n satisfies all the properties of Theorem 3.2, d'_n satisfies all the properties of Theorem 2, and we can apply this theorem to establish that d'_n must be the Averaged Euclidean Distance and so:

$$d_n(\mathbf{x},\mathbf{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(g(x_i)-g(y_i))^2}$$

6.5. **A derivation of Property 3*.** Let $d_1 : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ be a mobility index for single-family societies. We consider three basic properties that d_1 may satisfy. The first property captures the essence of mobility, since it states that in any society there is mobility if and only if there is a change in income across the two generations:

Property 6. *$d_1(a, b) \geq 0$ for all $a, b \in \mathbb{R}_+$, with the equality holding if and only if $a = b$.*

The second property explicitly states that d_1 is *symmetric*:

Property 7. *For all $a, b \in \mathbb{R}_+$, $d_1(a, b) = d_1(b, a)$*

A third property deals with the decomposition of mobility amongst different generations. Consider the income of three successive generations in a family, a, b, c, and consider the total amount of mobility from the first to the third. In general, it would seem plainly wrong to assume that the total mobility $d_1(a, c)$ should be equal to $d_1(a, b) + d_1(b, c)$, since some families may experience first upward and then downward income mobility which would "wash out" in $d_1(a, c)$. We assume however that this property holds whenever $a \geq b \geq c$ or $c \geq b \geq a$:

Property 8. *For all $a, b, c \in \mathbb{R}_+$ such that either $a \leq b \leq c$ or $a \geq b \geq c$, $d_1(a, c) = d_1(a, b) + d_1(b, c)$.*

Then, we have:

Theorem 4. *Properties 6–8 hold if and only if Property 3* holds.*

Proof. If d_1 satisfies Property 3*, it is easy to verify that it satisfies Properties 6–8. To prove the converse, let $e \geq 0$, be an arbitrary constant and, for any a, let $g_e(a)$ be defined as follows:

$$g_e(a) = \begin{cases} d_1(e, a) & \text{if } a \geq e \\ -d_1(e, a) & \text{if } a < e. \end{cases}$$

Now consider $d_1(a, b)$ and suppose first that $a \leq b$. We distinguish three cases.

(1) If $e \leq a \leq b$, then by Property 8 we have $d_1(e, b) = d_1(e, a) + d_1(a, b)$, that is

$$d_1(a, b) = g_e(b) - g_e(a).$$

(2) If $a < e \leq b$, then $d_1(a, b) = d_1(a, e) + d_1(e, b)$, and it is easy to check that

$$d_1(a, b) = g_e(b) - g_e(a).$$

(3) If $a \leq b < e$, then $d_1(a, e) = d_1(a, b) + d_1(b, e)$, and again

$$d_1(a, b) = g_e(b) - g_e(a).$$

Suppose now that $a > b$. A similar argument, using Property 7, leads to the identity

$$d_1(a, b) = g_e(a) - g_e(b).$$

It then follows that Property 3* holds true. Notice that, if $a > b$, then $d_1(a, b) = g_e(a) - g_e(b)$; since $d_1(a, b) > 0$ (by Property 6), it follows that $g_e(a) > g_e(b)$, that is g is increasing. □

References

[Acz66] J. Aczel. *Lectures on functional equations and their applications*. Accademic Press, New York, 1966.

[ASB99] D. Austen-Smith and J.S Banks. *Positive Political Theory I: Collective Preference*. The University of Michigan Press, 1999.

[Atk83] A.B. Atkinson. The measurement of economic mobility. In A.B. Atkinson, editor, *Social Justice and Public Policy*. Wheatsheaf Books Ltd., London, 1983.

[Dar93] V. Dardanoni. Measuring social mobility. *Journal of Economic Theory*, 61:372–94, 1993.

[Deb60] G. Debreu. Topological methods in cardinal utility. In S. Karlin K. Arrow and P. Suppes, editors, *Mathematical methods in the social sciences*. Stanford University Press, 1960.

[EMR88] J.M. Enelow, N.R. Mendell, and S. Ramesh. A comparison of two distance metrics through regression diagnostics of a model of relative candidate evaluation. *The Journal of Politics*, 50:1057–1071, 1988.

[ET80] L.G. Epstein and S.M. Tanny. Increasing generalized correlation: A definition and some economic consequences. *Canadian Journal of Economics*, 13:16–34, 1980.

[FO96] G.S. Fields and E. Ok. The meaning and measurement of income mobility. *Journal of Economic Theory*, 71:349–77, 1996.

[FO99a] G.S. Fields and E. Ok. The measurement of income mobility. In J. Silber, editor, *Handbook of Income Inequality Measurement*. Kluwer Academic Publishers, Dordrecht, 1999.

[FO99b] G.S. Fields and E. Ok. Measuring movements of incomes. *Economica*, 66:455–71, 1999.

[FS91] J.E. Foster and A.F. Shorrocks. Subgroup consistent poverty indices. *Econometrica*, 59:687–709, 1991.

[FS97] J.E. Foster and A. Sen. On economic inequality after a quarter century. In A. Sen, editor, *On Economic Inequality*. Clarendon Press, Oxford, 1997.

[Gor68] W.M. Gorman. The structure of utility functions. *Review of Economic Studies*, 35:367–390, 1968.

[HM70] M.J. Hinich and M.C. Munger. *Analytical Politics*. Cambridge University Press, Cambridge, UK, 1970.

[Maa98] E. Maasoumi. On mobility. In D. Giles and A. Ullah, editors, *The Handbook of Economic Statistics*. Marcel Dekker, New York, 1998.

[MO98] T. Mitra and E. Ok. The measurement of income mobility: A partial ordering approach. *Economic Theory*, 12:77–102, 1998.

[Sho88] A.F. Shorrocks. Aggregation issues in inequality measurement. In W. Eichorn, editor, *Measurement in Economics*. Springer Verlag, New York, 1988.

[SJ99] S. Santini and R. Jain. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):871–883, 1999.

[SS02] U. Segal and J. Sobel. Min, max, and sum. *Journal of Economic Theory*, 25:126–150, 2002.

[Tch80] A.H.T. Tchen. Inequalities for distributions with given marginals. *Annals of Probability*, 8:814–27, 1980.

[TK70] A. Tverski and D. H. Krantz. The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, 7:572–596, 1970.

Dipartimento di Scienze Umane, Università di Ferrara
E-mail address: dgm@unife.it

Dipartimento di Scienze Economiche, Aziendali e Finanziarie, Università di Palermo
E-mail address: vdardano@unipa.it