Submajorization and the Geometry of Unordered Collections

Rachel Wells Hall and Dmitri Tymoczko

Abstract. In this paper, we use submajorization to compare distances between either multisets of real numbers or multisets modulo translation on the real line. We provide a geometrical interpretation in which multisets are represented by points in an orbifold and bijections between multisets are represented by paths between these points. This interpretation shows that submajorization is closely related to the geometrical principle that the shortest path between two points is a straight line. Our results have applications to diverse problems from economics to music theory; moreover, they suggest generalizations of statistical measures of the center and spread of a distribution.

1. INTRODUCTION. Consider the plot of income levels for four individuals shown in Figure 1(a). In which year did incomes change the most? It is easy to answer this question for any one individual, but it is unclear how to aggregate changes in the group as a whole. Should we compare the sum of all individual changes, the largest individual change, or something else entirely? The answer is even less clear if we want to measure "relative" income mobility in a way that is insensitive to overall growth or inflation, as in Figure 1(b).

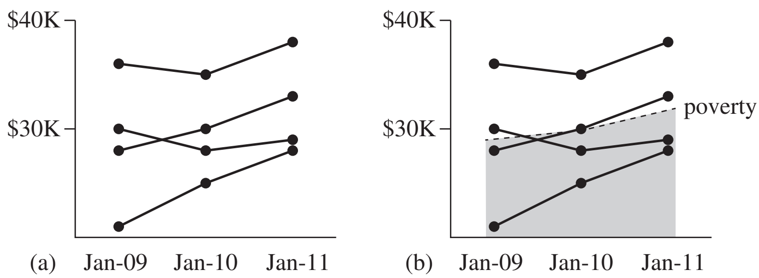

Figure 1. (a) Possible income changes for four individuals. (b) Income changes showing overall economic growth.

These questions have been well-studied in the voluminous economics literature (see Lorenz [**8**], Zheng [**14**], Mitra and Ok [**10**], and D'Agostino and Dardanoni [**3**]). Here, however, we treat them as special cases of a more general problem. We would like to assign an overall "size" to n simultaneous motions in the same one-dimensional metric space, but we have no principled reason for choosing one particular metric in the product space. In the cases to be considered, it is reasonable to assume that the assignment will depend only on the multiset of distances moved by each of the individual actors. Formulated in this general way, we will see that similar issues arise in many different contexts.

http://dx.doi.org/10.4169/amer.math.monthly.119.04.263
MSC: Primary 57R18, Secondary 91B15; 00A65

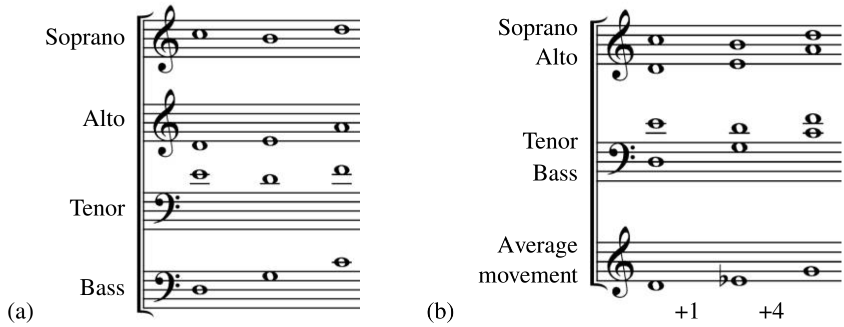

Figure 2. (a) Polyphony. (b) Average vocal movement.

Consider music theory, for example. Western polyphonic music articulates simultaneous melodies that also form meaningful harmonies (see Tymoczko [**12, 13**]). In standard musical notation, these two dimensions are represented spatially, with melodies notated horizontally and harmonies notated vertically. Each dimension is subject to its own constraint. To be perceived as coherent, melodies must move by short distances while harmonies must sound similar to one another. The challenge in composing is to design arrays of notes that are coherent in both ways at the same time.

Here again we face the problem of aggregating multiple motions in several independent metric spaces. The composer's job is essentially to minimize the "distance" between successive verticalities (chords), where that "distance" is itself a composite formed by motions in many distinct instrumental lines or "voices." While it is relatively clear how to measure the distance moved by any one musical line, it is unclear how to trade off motion in one line against motion in another. Is it better for one line to move by two units, or for two lines to each move by one unit? Remarkably, theorists have sometimes asked these same questions in the presence of a musical analogue to inflation, as it can be important to know how much the voices move *relative to one another*, rather than absolutely [**1**], [**13**, Chapter 3]. In these circumstances, we factor out motion shared by all voices, much as we might factor out inflation when measuring changes in income.

In such circumstances, the familiar *submajorization partial order* usefully demarcates a set of basic metrical intuitions about multisets of distances. Given two n-element multisets of nonnegative real numbers, X and Y, we say that X *submajorizes* Y if the sum of the largest k elements of X is at least as large as the sum of the largest k elements of Y, for all k from 1 to n [**9**]. The submajorization partial order encodes the intuition that the size of a multiset is related to evenness. Given two n-element collections summing to x, a more even distribution of values is no larger than a very uneven distribution. Economically, this corresponds to the principle that a large number of small changes in income, distributed among many individuals and summing to x, is no larger than a few changes in a small number of individuals, again summing to x. Moreover, increasing any element of X does not decrease the multiset's overall size. These basic principles are common to many familiar metrics. In fact, if X submajorizes Y then all L^p norms for $p \geq 1$ agree that X is at least as large as Y (the converse does not hold, however).

In this paper, we show that the submajorization partial order has a natural geometrical description in the symmetric product spaces $\mathbb{R}^n/\mathcal{S}_n$, where $\mathcal{S}_n$ is the symmetric group in n letters. These spaces are orbifolds—singular quotient spaces of $\mathbb{R}^n$ by $\mathcal{S}_n$. Points in these orbifolds represent "states" of the system. For example, in the economic case they might represent distributions of wealth in a society, while in the musical case they might represent a particular chord. Use of the symmetric product expresses the thought that the states we wish to consider are fundamentally unordered. The state in which individual A has \$2 and individual B has \$5 is equivalent to the state in which individual A has \$5 and individual B has \$2. Transitions from one state to another—representing, for example, the richer individual losing \$1 or giving \$1 to the poorer individual—can be associated with paths in these spaces. Each of these paths corresponds to a multiset of distances in $\mathbb{R}^n$. In this paper, we will apply submajorization to these multisets of distances rather than (as is more typical) states themselves.

We begin by proving that this use of the submajorization partial order is equivalent to two intuitive principles: (1) (the *Monotonicity Principle*) increasing the change experienced by any one individual does not decrease the size of the overall change and (2) (the *No-Crossings Principle*) there is a minimal change between any two states that preserves the order relationships among individuals (for instance, their relative ranking in terms of wealth).[1] This in turn allows us to understand submajorization geometrically. The No-Crossings Principle asserts that there exists a minimal-length path between any two points in the symmetric product space that does not touch a singularity. Singularities in the orbifold correspond to multisets with multiple copies of the same number, as when multiple individuals have the same wealth. Formulated in this way, the No-Crossings Principle can be seen to be an analogue of the triangle inequality for a restricted class of triangles. It follows that the submajorization partial order endows the higher-dimensional orbifolds with more-than-topological, but less-than-geometrical structure, allowing us to compare paths' lengths without quantifying these lengths numerically. The geometrical interpretation also suggests a way to apply submajorization when we wish to disregard the effects of uniform translation. In musical terms, this allows us to measure the motion of musical voices relative to one another, as in Figure 2(b). We use this construction to generalize statistical measures of the center and spread of a distribution.

2. SUBMAJORIZATION.

A *state* is a multiset of real numbers (an element of $\mathbb{R}^n/\mathcal{S}_n$). A *change* is a bijection[2] between two states, denoted $(a_1, a_2, \ldots, a_n) \to (b_1, b_2, \ldots, b_n)$, indicating that a_1 is paired with b_1, a_2 is paired with b_2, etc. The multiset $\{|b_1 - a_1|, \ldots, |b_n - a_n|\}$ is called the *displacement multiset* corresponding to the change; it is simply the collection of distances between paired elements. These definitions can be extended to the case where the one-dimensional metric space is the circle rather than the real line, though we will not pursue this issue here.

Our goal in this paper is to explore what we can say about the *size* of a change, in situations where we may not have principled reasons for choosing a particular metric in $\mathbb{R}^n/\mathcal{S}_n$. Tymoczko [**12**] proposes three basic assumptions that are valid in a range of applications: (1) the size of a change depends only on the displacement multiset, (2) increasing any one element of the displacement multiset does not decrease the size of the overall change, and (3) the overall size of a change between two states that pairs elements ordered from smallest to largest is no greater than the size of any other change

[1] In the economics literature, D'Agostino and Dardanoni [**3**] propose "value sensitivity" and "order sensitivity" conditions that are versions of the Monotonicity and No-Crossings Principles, respectively.

[2] The requirement that changes be bijective means that we cannot analyze changes between states with different numbers of elements—a difficulty in economic applications, where individuals may leave or join a society from year to year. However, economists frequently use percentiles or quintiles as proxies for individuals, in which case changes will be bijective. Note also that bijections between *multisets* can be reinterpreted as nonbijective relationships between sets. For example, $(0, 4, 4, 7) \to (0, 3, 5, 9)$ is simultaneously a bijective change between multisets $\{0, 4, 4, 7\}$ and $\{0, 3, 5, 9\}$ and a nonbijective pairing of the sets $\{0, 4, 7\}$ and $\{0, 3, 5, 9\}$.

between the same two states. The last two assumptions define a partial order, $\leq_{\text{MNC}}$, on displacement multisets (that is, on the space of n-element multisets of nonnegative real numbers $\mathbb{R}^n_+/\mathcal{S}_n$).

In what follows, $x_{[i]}$ denotes the ith largest element of the multiset $\{x_1, \ldots, x_n\}$.

Definition 1. Let $\leq_{\text{MNC}}$ be the partial order induced on $\mathbb{R}^n_+/\mathcal{S}_n$ by the following.

1. *The Monotonicity Principle.* If X and Y are n-element multisets of nonnegative real numbers whose elements may be ordered so that Y is no less than X in each coordinate, then $X \leq_{\text{MNC}} Y$.
2. *The No-Crossings Principle.* If $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ are multisets of real numbers, then

$$\{|b_{[1]} - a_{[1]}|, \ldots, |b_{[n]} - a_{[n]}|\} \leq_{\text{MNC}} \{|b_1 - a_1|, \ldots, |b_n - a_n|\}.$$

The No-Crossings Principle means that if X is the displacement multiset of the "uncrossed" change pairing the largest element of some state A with the largest element of another state B, the second largest element of A with the second largest element of B, and so on, and Y is the displacement multiset of some "crossed" change from A to B, then $X \leq_{\text{MNC}} Y$. If we locate A and B on copies of the real line, as in Figure 3, and connect elements that are paired by a change from A to B, the change's displacement multiset consists of the absolute vertical distances traversed by the connecting line segments. According to the partial order $\leq_{\text{MNC}}$, the "uncrossed" change in Figure 3(a) should be at least as small as the "crossed" change in Figure 3(b). In economic applications, the No-Crossings Principle states that an efficient redistribution of income does not create an incentive to destroy money by making a poorer individual better-off than a richer individual. In music, the principle states that a pianist can move efficiently from chord to chord without tangling his or her fingers.

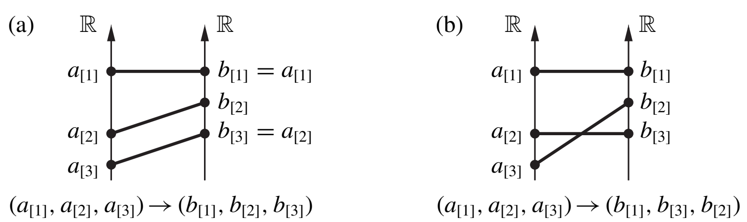

Figure 3. The No-Crossings Principle: the displacement multiset of the change represented in (a) is less than or equal to that in (b).

Now consider the well-known submajorization partial order on $\mathbb{R}^n_+/\mathcal{S}_n$.[3]

Definition 2. Submajorization $\prec_w$ is the partial order induced on $\mathbb{R}^+_n/\mathcal{S}_n$ by the following.

1. *The Monotonicity Principle.* If X and Y are n-element multisets of nonnegative real numbers whose elements may be ordered so that X is less than or equal to Y in each coordinate, then $X \prec_w Y$.
2. *The Dalton Transfer Principle.* Let $\{x_1, \ldots, x_n\}$ be a multiset of nonnegative real numbers. If $x_i \geq x_j$, then

[3]See Marshall and Olkin [**9**] for a thorough introduction to majorization and its associated geometry. Tymoczko refers to this formulation of submajorization as the "distribution constraint" in [**12**].

$$\{x_1, \ldots, x_i - \epsilon, \ldots, x_j + \epsilon, \ldots x_n\} \prec_w \{x_1, \ldots, x_i, \ldots, x_j, \ldots x_n\} \quad (1)$$

for all ϵ such that $0 < \epsilon \leq x_i - x_j$.

In traditional economic applications, the numbers in the multiset would represent *states* (e.g., the number of dollars in one's bank account) and inequality (1) would represent a *Dalton transfer* that redistributes resources from a wealthy individual to a poorer one. A basic principle of economics is that Dalton transfers do not increase income inequality [**4, 5**]. In this paper, however, the numbers will be interpreted as *displacement multisets* representing (for example) the *changes* in individual assets. The Dalton transfer principle, when applied to displacement multisets, asserts that even distributions of motion are smaller than uneven distributions. A few large changes will produce a larger quantity of total motion than will many small changes summing to the same value.

We now show that the submajorization partial order, when applied to displacement multisets, is equivalent to the No-Crossings Principle defined above. An informal proof of this theorem is also given in Tymoczko [**12**].

Theorem 1. *The partial order induced on* $\mathbb{R}^n_+/\mathcal{S}_n$ *by* $\leq_{\text{MNC}}$ *is equivalent to the submajorization partial order* $\prec_w$.

Proof. We claim that the partial order induced on $\mathbb{R}^n_+/\mathcal{S}_n$ by the No-Crossings Principle is equivalent to the partial order $\leq_{\text{NC}}$ induced by the following two conditions:

C1. $\{x_1, \ldots, x_i, \ldots, x_j, \ldots x_n\} \leq_{\text{NC}} \{x_1, \ldots, x_i + \epsilon, \ldots, x_j + \epsilon, \ldots x_n\}$ for all $\epsilon > 0$ and $i \neq j$, and,

C2. if $x_i \geq x_j$, then $\{x_1, \ldots, x_i, \ldots, x_j, \ldots x_n\} \leq_{\text{NC}} \{x_1, \ldots, x_i + \epsilon, \ldots, x_j - \epsilon, \ldots x_n\}$ for all ϵ such that $0 < \epsilon \leq x_j$.

Since condition C2 is equivalent to the Dalton Transfer Principle, while C1 is weaker than the Monotonicity Principle, proof of our claim suffices to prove the theorem.

Suppose $X, Y \in \mathbb{R}^n_+/\mathcal{S}_n$ and $X \leq_{\text{NC}} Y$. We may assume $X \leq_{\text{NC}} Y$ by direct application of either condition C1 or C2. In these cases, X and Y differ in exactly two coordinates, so it suffices to consider $n = 2$.

Case 1. Suppose $X \leq_{\text{NC}} Y$ by C1. Let $X = \{x_1, x_2\}$; then $Y = \{x_1 + \epsilon, x_2 + \epsilon\}$ for some $\epsilon > 0$. Let $A = \{0, \epsilon\}$ and $B = \{-x_1, x_2 + \epsilon\}$.

Case 2. Suppose $X \leq_{\text{NC}} Y$ by C2. Let $X = \{x_1, x_2\}$ where $x_1 \geq x_2$; then $Y = \{x_1 + \epsilon, x_2 - \epsilon\}$ for some $0 < \epsilon \leq x_2$. Let $A = \{0, \epsilon\}$ and $B = \{x_1 + \epsilon, x_2\}$.

In both cases, we conclude that X is less than or equal to Y by the No-Crossings Principle.

Suppose X is less than or equal to Y by the No-Crossings Principle. It suffices to consider instances where the change $A \to B$ contains a single crossing, and hence to restrict our attention to $n = 2$. There exist $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$ where $a_1 \geq a_2$ and $b_1 \geq b_2$ such that $X = \{|b_1 - a_1|, |b_2 - a_2|\}$ and $Y = \{|b_1 - a_2|, |b_2 - a_1|\}$. Without loss of generality, suppose $a_1 \geq b_1$. There are three cases.

Case 1. If $a_1 \geq b_1 \geq b_2 \geq a_2$, then $X = \{a_1 - b_1, b_2 - a_2\} \leq_{\text{NC}} \{a_1 - b_2, b_1 - a_2\} = Y$ by C1, where $\epsilon = b_1 - b_2$.

Case 2. If $a_1 \geq a_2 \geq b_1 \geq b_2$, then $X = \{a_1 - b_1, a_2 - b_2\} \leq_{\text{NC}} \{a_1 - b_2, a_2 - b_1\} = Y$ by C2, where ϵ is the smaller of $\{a_1 - a_2, b_1 - b_2\}$.

Case 3. If $a_1 \geq b_1 \geq a_2 \geq b_2$, then $X = \{a_1 - b_1, a_2 - b_2\} \leq_{\text{NC}} \{0, a_1 + a_2 - b_1 - b_2\}$ by C2, and $\{0, a_1 + a_2 - b_1 - b_2\} \leq_{\text{NC}} \{b_1 - a_2, a_1 - b_2\} = Y$ by C1. ■

The following equivalent definition of submajorization will prove useful (see Lemmas 1 and 2 in [**6**, p. 47] for proof that it is equivalent to definition 2).

Definition 3. Let $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ be multisets of nonnegative real numbers. We say that Y *submajorizes* X and write $X \prec_w Y$ if and only if

$$\sum_{i=1}^{j} x_{[i]} \leq \sum_{i=1}^{j} y_{[i]} \quad \text{for} \quad 1 \leq j \leq n. \tag{2}$$

In other words, the sum of the j largest members of X is less than or equal to the sum of the j largest members of Y, for all j less than or equal to n.

A fourth definition clarifies the sense in which submajorization demarcates a "zone of agreement" among a range of different approaches to quantifying the size of a change.

Definition 4. Let X and Y be n-element multisets of nonnegative real numbers. Then $X \prec_w Y$ if and only if $f(X) \leq f(Y)$ for every real-valued function that is symmetric, convex, and increasing.

A real-valued function on $\mathbb{R}^n_+$ is:

1. *symmetric* if it is well-definted on $\mathbb{R}^n_+/\mathcal{S}_n$ (that is, $f(\mathbf{v}) = f(\sigma(\mathbf{v}))$ for all σ in $\mathcal{S}_n$),
2. *convex* if, for any vectors $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^n_+$, $f\left(\frac{1}{2}(\mathbf{v} + \mathbf{w})\right) \leq \frac{1}{2}(f(\mathbf{v}) + f(\mathbf{w}))$, and
3. *increasing* if it is increasing in each of its elements.

Functions that are symmetric, convex, and increasing are called *strongly isotone*. The equivalence of the convexity condition and submajorization follows from [**9**, Proposition 3.C.2, p. 67]. Thus, submajorization is equivalent to consensus among all strongly isotone measures. Examples of strongly isotone functions include the L^p-norms for $p \geq 1$, where the L^p-norm of a multiset X in $\mathbb{R}^n_+$ is defined to be $||X||_p = \left(\sum_{i=1}^n x_i^p\right)^{1/p}$, and functions Σ_j summing the j largest elements of a multiset (that is, $\Sigma_j(X) := \sum_1^j x_{[j]}$, as in equation (2), above). In fact, any symmetric norm on $\mathbb{R}^n$ is strongly isotone.

2.1. The geometry of submajorization. States can be represented as points in the symmetric product $\mathbb{R}^n/\mathcal{S}_n$; these spaces are singular at points whose coordinates contain two or more instances of the same number. Changes can be identified with the images, in the quotient spaces, of line segments in $\mathbb{R}^n$.[4] When these changes have crossings they will appear in the quotient spaces to "bounce off" singularities. For instance, the shaded region in Figure 4 represents the space $\mathbb{R}^2/\mathcal{S}_2$, a half-plane containing all points $x \geq y$ and singular at the line $x = y$. The change $(0, 3) \to (2, 5)$ is represented by line segment P in the parent space $\mathbb{R}^2$, which appears to be straight in the quotient. The change $(0, 3) \to (5, 2)$ is represented by the "generalized line segment" Q, which appears to "bounce off" the singularity. Note that we distinguish "generalized line segments," which may appear to reflect off a singularity,

[4]Formally, the image in the quotient of the line segment $\mathbf{a} + t(\mathbf{b} - \mathbf{a})$ for $t \in [0, 1]$ represents the change $\mathbf{a} \to \mathbf{b}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.

from "proper line segments" which touch the singularities only at their endpoints, if at all. The advantage of working in a singular orbifold, rather than in the subspace $x_1 \leq x_2 \leq \cdots \leq x_n$, is that the orbifold allows us to define "generalized line segments" which are in one-to-one correspondence with our "changes." In the nonsingular subspace, there is no collection of natural geometrical objects that represents all possible changes.

Submajorization, when applied to displacement multisets, requires that the line segment P be shorter than or equal to the generalized line-segment Q. It is clear from Figure 4 that this constraint is closely related to the triangle inequality, since change Q forms two legs of a triangle, while P forms the third. However, the condition is not equivalent to the triangle inequality—it applies only to triangles created when a line encounters a singularity of the quotient space.

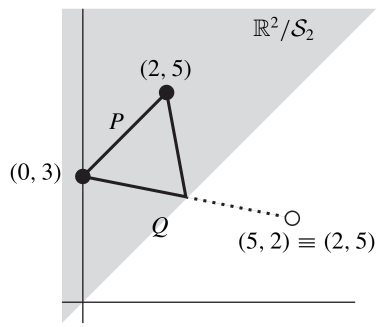

Figure 4. The changes $P : (0, 3) \to (2, 5)$ and $Q : (0, 3) \to (5, 2)$ in $\mathbb{R}^2/\mathcal{S}_2$.

By allowing us to compare distances without quantifying them, submajorization produces a structure intermediate between geometry and topology in the higher-dimensional symmetric product spaces $\mathbb{R}^n/\mathcal{S}_n$. This can be seen by considering the analogue of "spheres" and "balls" in our partial-order geometry.

In what follows, let $\Sigma_j(X)$ denote the sum of the j largest elements of a multiset X of nonnegative numbers (that is, $\Sigma_j(X) = \sum_{i=1}^{j} x_{[i]}$). Note that $\Sigma_1(X) = \max\{x_i\}$ is also known as the L^∞ norm of X (that is, $||X||_\infty$), and $\Sigma_n(X)$ equals $||X||_1$. For a vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, let $\{|v_i|\} = \{|v_1|, \ldots, |v_n|\}$.

Definition 5. The Σ_j*-ball* of a multiset X of nonnegative numbers is the set of vectors $\mathbf{v}$ in $\mathbb{R}^n$ such that $\Sigma_j(\{|v_i|\}) \leq \Sigma_j(X)$. The *inner submajorization ball* is the intersection of the Σ_j-balls, and the *outer submajorization ball* is the union of the Σ_j-balls.

Note that the *inner submajorization ball* of a multiset X of nonnegative real numbers is the set of vectors $\mathbf{v}$ such that $\{|v_i|\} \prec_w X$, while the closure of the complement of the outer submajorization ball is the set of vectors $\mathbf{v}$ such that $\{|v_i|\}_w \succ X$. If a vector $\mathbf{v}$ lies between the inner and outer submajorization balls, then $\{|v_i|\}$ and X are incomparable by submajorization. In general, a Σ_j-ball is the set of simultaneous solutions to the system of linear inequalities of the form $e_1 v_1 + \cdots + e_n v_n \leq \Sigma_j(X)$, where each e_i is chosen from the set $\{0, 1, -1\}$, and exactly j of the e_i are nonzero. Each Σ_j-ball is a convex polytope that is invariant under the group of signed permutations. For example, in $\mathbb{R}^2$, the Σ_1- and Σ_2-balls are filled squares, one rotated by 45°. Figure 5 shows a possible configuration. In $\mathbb{R}^3$, the Σ_1-ball is a filled cube, the Σ_2-ball is a filled rhombic dodecahedron, and the Σ_3-ball is a filled octahedron. The

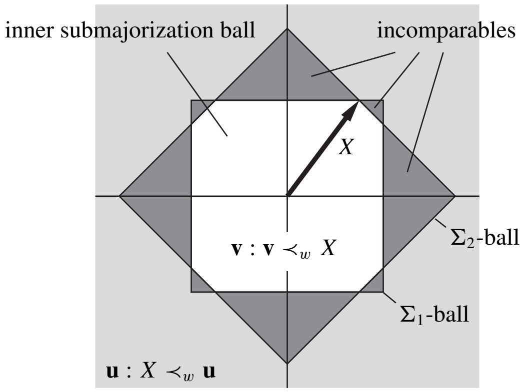

Figure 5. Submajorization balls in $\mathbb{R}^2$. A vector whose coordinates form the multiset X divides space into three regions. The white region contains vectors whose length is less than X by submajorization, while the light gray area contains vectors whose length is more than X by submajorization. The dark area between them represents a "zone of acceptable disagreement" where metrics can disagree about the length of vectors relative to X without violating either the No-Crossings or Monotonicity Principles.

inner submajorization ball—the intersection of the Σ_j-balls for $1 \leq j \leq n$—is the convex polytope whose vertices are the signed permutations of the elements of X [**9**, Prop. C.2, p. 113]. Figure 6 shows how the No-Crossings Principle and the Monotonicity Principle define regions of comparability and incomparability that combine to form the submajorization ball.

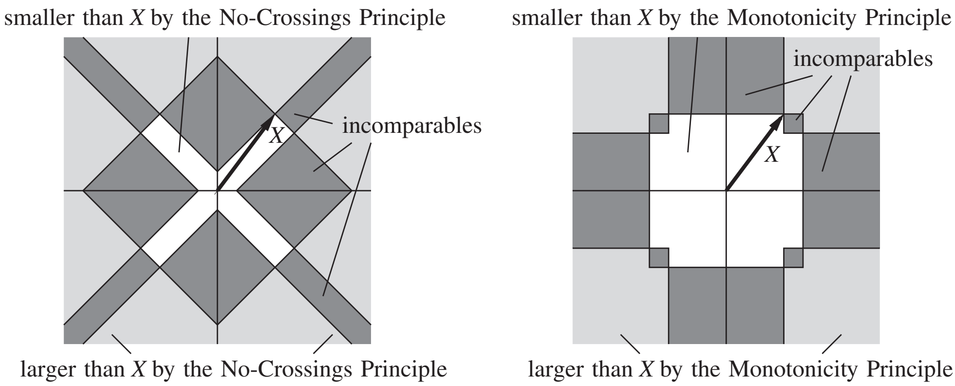

Figure 6. The geometry of Definition 1 in $\mathbb{R}^2$. Both the No-Crossings Principle and the Monotonicity Principle define unbounded regions of incomparability. Note that, for both conditions, the set of points in $\mathbb{R}^2$ whose displacement from the origin is less than X is nonconvex. The two conditions combine to produce the convex submajorization balls in Figure 5(a).

Example. We construct the inner submajorization ball for $\{1, 4, 5\}$. The Σ_1-ball is the filled cube with vertices $(\pm 5, \pm 5, \pm 5)$. The Σ_2-ball is the set of simultaneous solutions to the linear inequalities of the form $e_1x + e_2y + e_3z \leq 9$, where $e_i \in \{-1, 0, 1\}$ and exactly two of the e_i's are nonzero; it is a filled rhombic dodecahedron. The Σ_3-ball is the set of simultaneous solutions to the linear inequalities $e_1x + e_2y + e_3z \leq 10$ where the e_i terms are either 1 or -1; it is a filled octahedron. The inner submajorization ball is the intersection of these; its vertices form the set $\{\sigma(e_1, 4e_2, 5e_3) | \sigma \in \mathcal{S}_n, e_i \in \{-1, 1\}\}$. The set of vectors that submajorize $\{1, 4, 5\}$ is the complement of the union of the interiors of the three Σ_j-balls.

A straightforward argument (see Appendix A.1) shows that the region of incomparability lies between the Euclidean spheres of radius

$$r = \min\left\{ j^{-1/2}\Sigma_j(X) \mid 1 \le j \le n \right\},$$

and

$$R = \max\left\{ n^{1/2}\Sigma_1(X),\ \ \Sigma_n(X) \right\},$$

and these bounds are tight. Therefore, the Euclidean sphere sits in the "middle interval" of the region of incomparability, with either the L^1 or L^∞ ball containing the points that have maximal Euclidean distance from the center. These bounds are useful for error analysis, as they set quantitative limits on the "zone of disagreement" within which acceptable measures can differ.

3. SUBMAJORIZATION MODULO TRANSLATION. Let us now consider how we might use submajorization to measure changes in a way that is insensitive to overall translation. As we will see below, this problem arises when measuring similarity among types of musical chords; it is also related to the problem of measuring income inequality in a way that is insensitive to inflation. The idea is to consider all states $\{x_1 + c, \dots, x_n + c\}$ to be equivalent, since they differ only by addition of the common factor c. Changes between states can now be reinterpreted as changes between the equivalence classes they represent. We would then like to use submajorization to compare distances between these equivalence classes of states, which requires finding minimal changes between them.

Geometrically, our equivalence classes are represented by lines in $\mathbb{R}^n/\mathcal{S}_n$ parallel to the line from the origin to $\mathbf{1} = (1, 1, \dots, 1)$. Each line can be associated with its intersection with the hyperplane containing points whose coordinates sum to zero (the *zero-sum hyperplane* where $\sum x_i = 0$). The space of these equivalence classes, $\mathbb{R}^{n-1}/\mathcal{S}_n$, is therefore endowed with a "barycentric coordinate system" [**2**]. Figure 7 depicts these orbifolds in the case $n = 3$.

Note that although we can use the zero-sum hyperplane to provide a coordinate system for our equivalence classes, we cannot rely on these coordinates when applying

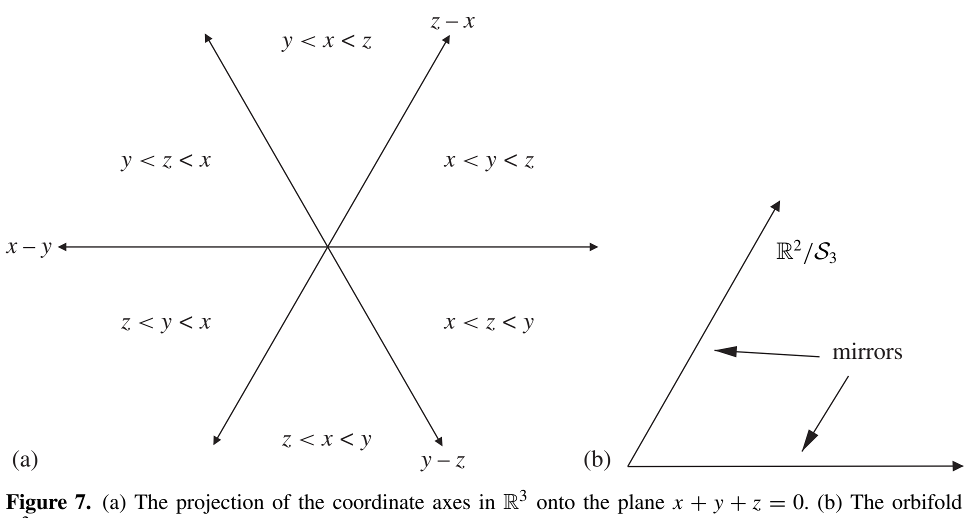

Figure 7. (a) The projection of the coordinate axes in $\mathbb{R}^3$ onto the plane $x + y + z = 0$. (b) The orbifold $\mathbb{R}^2/\mathcal{S}_3$ formed by identifying points whose coordinates are related by permutation.

submajorization to changes between these classes. For example, consider points $A = \{0, 0, 0, 8\}$ and $B = \{0, 0, 6, 6\}$. Let $[A]$ and $[B]$ represent the lines $\{c, c, c, 8 + c\}$ and $\{c, c, 6 + c, 6 + c\}$ respectively. These lines intersect the zero-sum hyperplane at $A' = \{-2, -2, -2, 6\}$ and $B' = \{-3, -3, 3, 3\}$. Submajorization tells us that the *point* B' must be at least as close to the origin as A', since $\{6, 2, 2, 2\}$ submajorizes $\{3, 3, 3, 3\}$. However, it does not tell us that the *equivalence class* $[B]$ is closer to the origin than $[A]$. To see this, note that there are L^p norms that disagree about which line is closer to the line $x_1 = x_2 = x_3 = x_4$ (i.e., the equivalence class containing the origin). According to L^1, $[A]$ is closer to this line than $[B]$ (distance 8 vs. 12), while according to L^∞, the opposite is the case (distance 4 vs. 3).

The issue here is that we do not have metric-independent reasons for using the coordinates in the zero-sum hyperplane. If we are using L^2, then points in the hyperplane minimize the distances among the lines they represent. When we turn to other metrics, this is no longer the case. It is interesting that some applications, including economics, give us independent reasons to choose this particular hyperplane, as we will see shortly. In other cases, however, we need a more sophisticated approach.

Definition 6. Let $V = \{v_1, \ldots, v_n\}$ and $W = \{w_1, \ldots, w_n\}$ be multisets of real numbers. We say that W is **T**-*smaller* than V if, for any real number x, there exists a real number y such that

$$\{|w_i - y|\} \prec_w \{|v_i - x|\}. \tag{3}$$

Intuitively, the definition says that the equivalence class of W-translations is at least as close to the origin as the equivalence class of V-translations, according to the submajorization partial order. More precisely, for every translation of V by x, there exists a translation of W by y which is closer to the origin according to submajorization. Note that here the coordinates of an equivalence class of states are representing *changes* that take the class to the equivalence class containing the origin. Geometrically, W is **T**-smaller than V if and only if the line $\{w_i - y\}$ for $y \in \mathbb{R}$ intersects the inner submajorization ball of $\{|v_i - x|\}$ for all x. Equivalently, W is **T**-smaller than V if and only if the projection of W to the zero-sum plane lies in the intersection, over all real x, of the projections to the zero-sum plane of the inner submajorization balls of $\{|v_i - x|\}$. Figure 8 depicts a possible configuration of submajorization balls.

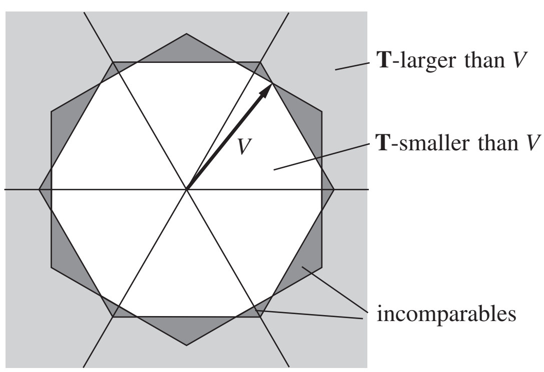

Figure 8. **T**-smallness in the projection of $\mathbb{R}^3$ to the zero-sum plane.

Unfortunately, it is difficult to give an explicit characterization of the submajorization balls in this projective context. The regions are characterized by the simultaneous satisfaction of multiple inequalities, and do not seem to admit of a more intuitive description. Our goal in this section, therefore, will be to provide an efficient algorithm that can be used to determine whether one vector submajorizes another when translations are disregarded. Along the way, we will uncover some links between the current discussion and familiar topics from statistics.

Consider the set $\{\{|v_i - x|\} : x \in \mathbb{R}\}$. An element of a partially ordered set is *minimal* if and only if it is not larger than any other element of the set. Note that there may be multiple minimal elements, all mutually incomparable according to the partial order. A *least element* is less than or equal to any other element in the set. We show in Lemma 2 that an element of $\{\{|v_i - x|\} : x \in \mathbb{R}\}$ is minimal in the submajorization partial order if and only if x lies in a closed interval which we call the "middle interval" of V (see Figure 9).

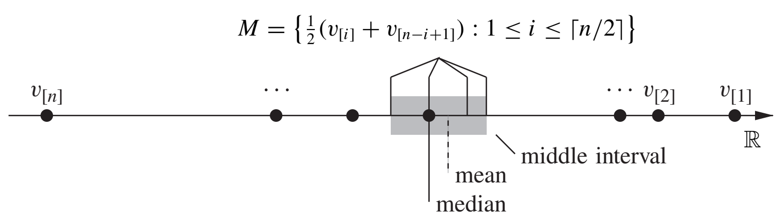

Figure 9. The middle interval of a multiset V. On the real line, the middle interval is the set of "acceptable centers" of V.

Definition 7. Let $V = \{v_1, \ldots, v_n\}$ be a multiset of real numbers and let M be the set

$$M = \left\{\tfrac{1}{2}(v_{[i]} + v_{[n-i+1]}) : 1 \le i \le \lceil n/2 \rceil\right\}.$$

The *middle interval* of V is the interval $[\min M, \max M]$.

For example, if $V = \{9, 8, 0, 0, -1, -3\}$, then $M = \{\frac{1}{2}(9-3), \frac{1}{2}(8-1), \frac{1}{2}(0+0)\} = \{3, 3.5, 0\}$ and the middle interval of V is $[0, 3.5]$.

Lemma 1. *Let $V = \{v_1, \ldots, v_n\}$ be a multiset of real numbers. For any j, $1 \le j \le n$, $\Sigma_j(\{|v_i - x|\})$ is a convex function of x that attains its minimum at $x = m_j$, where $m_j = \frac{1}{2}(v_{[k]} + v_{[n-k+1]})$ for $k = \lceil j/2 \rceil$.*

Proof. Let $S_j(x) \subset V$ be the multiset of the j most distant points in V from x (if there are multiple jth most distant points, choose elements so that the cardinality of $S_j(x)$ is j). Let $r_j(x)$ be number of elements of $S_j(x)$ that are greater than or equal to x, and let $l_j(x)$ be the number of elements that are less than x. The function

$$\Sigma_j(\{|v_i - x|\}) = \sum_{s \in S_j(x)} |s - x|$$

is continuous and piecewise linear. Its slope, when defined, is $l_j(x) - r_j(x)$, which is an increasing function that either equals zero when x equals m_j (for j even), or changes sign at $x = m_j$ (for j odd). Therefore, $\Sigma_j(\{|v_i - x|\})$ is convex and attains its minimal value at $x = m_j$. ∎

Lemma 2. *Let $V = \{v_1, \ldots, v_n\}$ be a multiset of real numbers. The multiset $\{|v_i - m|\}$ is a minimal element of $\{\{|v_i - x|\} : x \in \mathbb{R}\}$ with the submajorization partial ordering if and only if m lies in the middle interval of V.*

Proof. Let M denote the middle interval of V and let

$$\mathcal{D}(V) = \{\{|v_i - x|\} : \min M \leq x \leq \max M\}.$$

There exist odd integers s and t such that: (1) the unique minimum of $\Sigma_s(\{|v_i - x|\})$ occurs at $x = \max M$ and $\Sigma_s(\{|v_i - x|\})$ is strictly decreasing for $x \leq \max M$, and (2) the unique minimum of $\Sigma_t(\{|v_i - x|\})$ occurs at $x = \min M$ and $\Sigma_t(\{|v_i - x|\})$ is strictly increasing for $x \geq \min M$. Suppose there exists an $x < \min M$ such that $\{|v_i - x|\} \prec_w \{|v_i - m|\}$ for some $m \in M$. Therefore, $\Sigma_s(\{|v_i - x|\}) \leq \Sigma_s(\{|v_i - m|\})$, and since Σ_s is strictly decreasing, $x \geq m$, which contradicts our assumption that $x < \min M$. In similar fashion we use Σ_t to prove that no multiset $\{|v_i - x|\}$ for $x > \max M$ is submajorized by an element of $\mathcal{D}(V)$. Moreover, on the interval $[\min M, \max M]$, Σ_s is strictly increasing and Σ_t is strictly decreasing, so no element of $\mathcal{D}(V)$ submajorizes all the other elements of $\mathcal{D}(V)$. ■

Corollary 1. *The poset $\{\{|v_i - x|\} : x \in \mathbb{R}\}$ ordered by $\prec_w$ has a least element if and only if V is symmetric about a value m; this least element is $\{|v_i - m|\}$.*

Definition 6 does not tell us how to determine an ordering based on **T**-smallness in finite time. Theorem 2 shows that it suffices to find y in inequality (3) for finitely many x, thus permitting a polynomial-time algorithm for **T**-smallness.

Theorem 2. *Let $V = \{v_1, \ldots, v_n\}$ and $W = \{w_1, \ldots, w_n\}$ be multisets of real numbers. Let S be the set of points $S = \left\{\frac{1}{2}(v_i + v_j) : 1 \leq i \leq j \leq n\right\}$, and let S' be the set of elements of S that lie in the middle interval of V. Then W is* **T***-smaller than V if and only if for each s in S' there exists a y such that*

$$\{|w_i - y|\} \prec_w \{|v_i - s|\}. \tag{4}$$

Proof. As in the proof of Lemma 1, each Σ_j is continuous and piecewise linear; its slope is undefined only at the discontinuities of r_j, which are precisely the elements of S. Consider any two consecutive elements of S, $s_{[k]}$ and $s_{[k+1]}$. Suppose there exist y_k and y_{k+1} such that

$$\{|w_i - y_k|\} \prec_w \{|v_i - s_{[k]}|\} \quad \text{and} \quad \{|w_i - y_{k+1}|\} \prec_w \{|v_i - s_{[k+1]}|\}.$$

Since each $\Sigma_j(\{|w_i - y|\})$ is convex, and each $\Sigma_j(\{|v_i - x|\})$ is linear for $s_{[k]} \leq x \leq s_{[k+1]}$, we have for every $x, y \in [s_{[k]}, s_{[k+1]}] \times [y_k, y_{k+1}]$ that

$$\Sigma_j(\{|w_i - y|\}) \leq \Sigma_j(\{|v_i - x|\}).$$

This fact, taken together with Lemma 2, implies that it suffices to find a value y in inequality (4) for each $s \in S'$. ■

In Appendix A.2, we sketch an $O(n^4)$ algorithm for determining whether a multiset W is **T**-smaller than a multiset V, where both multisets have n elements. This, in turn, allows us to use submajorization to compare distances among arbitrary equivalence classes of translationally related states.

Our algorithm allows us to assert that the multiset $\{1, 2, 4\}$ is closer to an equal distribution than the distribution $\{2, 2, 5\}$ even when we discount the effects of translation. In musical contexts, it allows us to assert that the chord (C, E, G) divides the octave more evenly than (C, F, G), no matter how we measure musical distance. See section A.3 of the appendix for details of these computations.

3.1. T-smallness and measures of center and dispersion. In this section, we demonstrate that **T**-smallness is a generalization of dispersion measures such as the standard deviation and least absolute deviation, while the middle interval of a multiset is a generalization of various statistical measures of central tendency such as the mean and median. Given a strongly isotone function f, the minimal value of $f(\{|v_i - x|\})$ assesses, in some sense, the dispersion of V. That is, the value represents the size of the smallest change that takes V to the line of agreement $x_1 = x_2 = \cdots = x_n$. This is a kind of "distance" between the equivalence classes containing V and the origin. Different metrics will of course measure this distance in different ways. If $f = \Sigma_n$, then $f(\{|v_i - x|\}) = ||V - x||_1$, and setting x equal to the median of V minimizes $||V - x||_1$. This distance is also known as the least absolute deviation (LAD) of V. Moreover, setting x equal to the mean of V minimizes $||V - x||_2$; in this case, the minimal value of $||V - x||_2$ equals $\sqrt{n}$ times the standard deviation of V. Figure 10 illustrates the minimal distances according to the L^1- and L^2-norms. We can think of the minimal value of $f(\{|v_i - x|\})$ as a "generalized dispersion measure" induced by the strongly isotone function f.[5] Furthermore, since the middle interval of V contains every value x that minimizes $f(\{|v_i - x|\})$ for some f, we can think of the middle interval as the set of "generalized centers" of V. **T**-smallness implies that all strongly isotone functions agree about the relative dispersions of two sets. So, for example, if W is **T**-smaller than V, then the standard deviation, variance, and range of W are all smaller than those of V. If W is *not* **T**-smaller than V, however, Theorem 3 shows us how to construct a strongly isotone function that measures the dispersion of V as smaller than that of W.

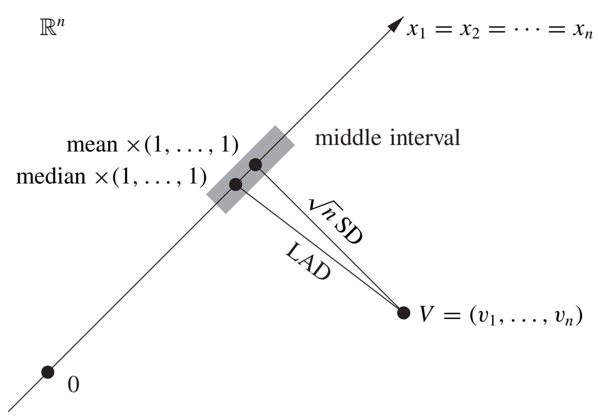

Figure 10. In $\mathbb{R}^n$, the middle interval corresponds to an interval on the line $x_1 = x_2 = \cdots x_n$. Each point in this interval minimizes the distance from the line to point V, as measured by some strongly isotone function. For example, the point whose coordinates equal the mean of V is the closest in the L^2 norm; the Euclidean length of the line segment connecting it to V equals $\sqrt{n}$ times the standard deviation of V. Moreover, the least absolute deviation (LAD) of V is the L^1 distance between V and the point whose coordinates equal the median of V.

[5] Recall that examples of strongly isotone functions include the Σ_j functions and L^p- norms for $p \geq 1$—in fact, any symmetric norm on $\mathbb{R}^n$ is strongly isotone.

Theorem 3. *Let $V = \{v_1, \dots, v_n\}$ be a multiset of real numbers.*

1. *For every strongly isotone function f, there exists at least one value x in the middle interval of V that minimizes $f(\{|v_i - x|\})$.*
2. *A value m lies in the middle interval of V if and only if there exists a strongly isotone function f_m such that*

$$f_m(\{|v_i - m|\}) < f_m(\{|v_i - x|\}) \quad \textit{for all } x \neq m.$$

3. *A multiset of real numbers $W = \{w_1, \dots, w_n\}$ is **T**-smaller than V if and only if*

$$\min_{y \in \mathbb{R}} f(\{|w_i - y|\}) \leq \min_{x \in \mathbb{R}} f(\{|v_i - x|\}) \tag{5}$$

for all strongly isotone functions f.

Proof.

1. This statement follows from Lemma 2.
2. If $f_m(\{|v_i - x|\})$ has a unique minimum at $x = m$, then it follows from statement 1 that m lies in the middle interval of V. Now suppose that m lies in the middle interval of V. If all the elements of V are identical, then $\Sigma_1(\{|v_i - x|\})$ equals zero if and only if $x = m$, and hence $f_m = \Sigma_1$ is minimized at $x = m$. Suppose V has more than one distinct element. We showed in the course of the proof of Lemma 2 that there exist odd integers $s \leq t$ such that $\Sigma_s(\{|v_i - x|\})$ is strictly decreasing on the middle interval of V and $\Sigma_t(\{|v_i - x|\})$ is strictly increasing on the middle interval of V. Let $\alpha_s = \Sigma_s(\{|v_i - m|\})$ and $\alpha_t = \Sigma_t(\{|v_i - m|\})$. Since V has more than one distinct element, α_s and α_t are positive. We claim that the function

$$f_m = \max(\{\alpha_t \Sigma_s, \alpha_s \Sigma_t\})$$

is a strongly isotone function (if g and h are strongly isotone, so is $\max(g, h)$). Moreover, since $\alpha_t \Sigma_s(\{|v_i - x|\})$ is strictly decreasing when $x < m$ and $\alpha_s \Sigma_t(\{|v_i - x|\})$ is strictly increasing when $x > m$, the unique minimum of $f_m(\{|v_i - x|\})$ occurs at $x = m$.
3. Clearly, if W is **T**-smaller than V, then inequality (5) holds for every strongly isotone function. Suppose that W is not **T**-smaller than V. We will construct a strongly isotone function F such that

$$\min F(\{|w_i - y|\}) > \min F(\{|v_i - x|\}). \tag{6}$$

Since W is not **T**-smaller than V, there exists an x_0 such that $\{|w_i - y|\} \not\leq_w \{|v_i - x_0|\}$ for all y. Without loss of generality, we assume x_0 is in the middle interval of V. Let

$$\alpha_j = \Sigma_j(\{|v_i - x_0|\}).$$

If any (and hence all) of the α_j's are zero, V has only one distinct element, and $F = \Sigma_1$ satisfies inequality (6). Alternately, all the α_j's are nonzero, so we define

$$F = \max\left\{\frac{\Sigma_1}{\alpha_1}, \frac{\Sigma_2}{\alpha_2}, \dots, \frac{\Sigma_n}{\alpha_n}\right\}.$$

Then F is a strongly isotone function. Since, as we have noted in the previous section of this proof, at least one of the Σ_j's is strictly increasing and at least one is strictly decreasing on the middle interval of V,

$$\min F(\{|v_i - x|\}) = F(\{|v_i - x_0|\}) = 1;$$

moreover, $F(\{|v_i - x|\})$ is strictly decreasing for $x < x_0$ and strictly increasing for $x > x_0$. Since $\{|w_i - y|\} \not\prec_w \{|v_i - x_0|\}$ for all y, for each y there exists a $j(y)$ such that

$$\Sigma_{j(y)}(\{|v_i - x_0|\}) < \Sigma_{j(y)}(\{|w_i - y|\}). \tag{7}$$

Then, for all y,

$$1 = \frac{\Sigma_{j(y)}(\{|v_i - x_0|\})}{\alpha_{j(y)}} < \frac{\Sigma_{j(y)}(\{|w_i - y|\})}{\alpha_{j(y)}} \leq F(\{|w_i - y|\}). \tag{8}$$

Since $1 = \min F(\{|v_i - x|\})$, F satisfies (6). ■

4. APPLICATIONS. Submajorization is useful when we have greater credence in the No-Crossings Principle than in our ability to chose a particular metric. This can occur when we are trying to quantify intuitions that are inherently nonspecific or vague. It might seem incredible, for example, that there is some particular metric (L^1 as opposed to L^2 or L^3 or L^∞) that exactly captures human intuitions about income inequality; instead, it is reasonable to suppose that intuition supports general comparisons between broadly different societies, without necessarily allowing us to make fine distinctions among very similar ones. At the same time, we have very strong economic reasons to endorse the No-Crossings Principle, which simply expresses the thought that efficient redistribution will not give individuals an incentive to abandon their money. For if the No-Crossings Principle were false, then individual A, richer than B before taxes, could be made poorer than B by the most efficient and egalitarian sequence of taxes and redistributions; under these circumstances, money would become a hot potato, with A having an incentive to trade pre-tax places with B, in order to ensure a better post-tax situation. Orderly economic life would scarcely be possible, as individuals constantly tried to rid themselves of pre-tax money so as to ensure a better post-tax outcome.[6]

In these contexts, submajorization can allow us to escape the tyranny of arbitrary metrics without giving up our ability to make general comparisons among distances. By articulating a class of metrical judgments that remain true regardless of which particular metric we choose—so long as that metric is compatible with the No-Crossings Principle—it demarcates a "zone of agreement" among what we might define as "reasonable" conceptions of distance. For some purposes, these coarse-grained judgments may be entirely sufficient. And even when they are not, the submajorization partial order allows us to distinguish claims that are true regardless of metric from those that depend on a particular choice. In this sense, an approach based on submajorization may help us avoid the false precision that can arise when we take particular metrics too seriously.

[6]Note that this may sometimes occur in the real world; but if so, it is a reflection of inefficiency in the economic system; submajorization merely requires that a perfectly efficient scheme not give people an incentive to abandon money.

In economics, inflation is represented by uniform scaling of wealth, represented as translation when we measure wealth in log dollars. Unfortunately, the formalism of Section 3 cannot be directly adapted to this case, since it is questionable whether the size of a change corresponds to the displacement multiset in log dollar space. However, when income is measured in dollars, there is a metric-independent *economic* justification for comparing income distributions by projecting them into a hyperplane containing a fixed amount of wealth. This simply expresses the basic assumption of the conservation of money. A hyperplane determined by the equation $x_1 + x_2 + \cdots + x_n = c$ represents all the income distributions that are mutually accessible by a sequence of economic exchanges in the absence of inflation (call this the no-inflation hyperplane). It is also the hyperplane that is picked out by the concept of Euclidean orthogonality, as we saw in Section 3, but that fact is irrelevant in this context; regardless of which metric you favor, the no-inflation hyperplane is important for reasons of economic logic alone. Thus an income distribution X determines an equivalence class αX of distributions that are equivalent under inflation (scaling by a factor $\alpha > 0$). Figure 11 shows that these equivalence classes are represented geometrically by a pencil of rays emanating from the origin. Since we have good economic reasons for choosing the no-inflation hyperplane, and since submajorization is invariant under uniform scaling, we can therefore compare inflation-induced equivalence classes. We simply project each class into the same no-inflation hyperplane, and use submajorization on changes between these representatives.

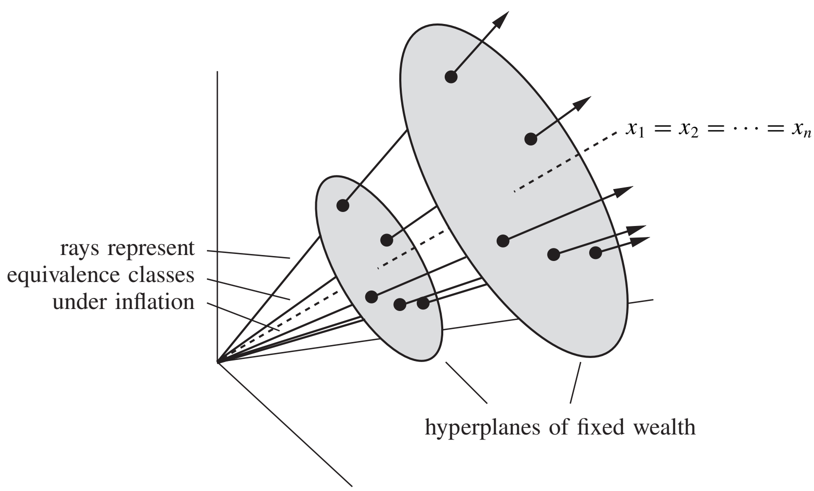

Figure 11. Equivalence classes of income distributions intersect hyperplanes of fixed wealth.

Of course, submajorization has a long and varied history in the economics literature. Economists originally applied submajorization to distributions of income rather than *changes* in income (see Lorenz [**8**] and Zheng [**14**], for example). Much of this work involves majorization, which is identical to submajorization except that it permits comparisons only among multisets with the same sum. Mitra and Ok [**10**] and D'Agostino and Dardanoni [**3**] recognized the usefulness of submajorization in measuring income changes; however, they do not explore the implications of submajorization on its own, nor do they connect submajorization to the No-Crossings Principle.[7] Thus, though the general concept of submajorization is quite familiar to economics, the approach described in this paper is not. By connecting submajoriation to the No-Crossings Principle, we provide a way to characterize the metrical intuitions encoded by this familiar mathematical partial order. By providing a geometrical interpretation in terms of symmetric product spaces, we relate submajorization to the familiar triangle inequality. More generally, by considering these issues in a more general geometrical setting, rather than in the context of particular economic applications, we show that analogous problems appear in a range of different contexts.

Indeed, we confront many of these same issues in music theory [**13**]. Musical pitches can be represented as points on the line, with numerical values corresponding to the logarithm of a soundwave's fundamental frequency. Musicians often treat notes that are an octave apart as being fundamentally similar, referring to them using the same basic letter names. For instance, in scientific pitch notation, middle C is labeled C4, while the note an octave above that is C5. In situations where it is appropriate to ignore octaves, musicians wrap linear pitch space into circular "pitch-class space," where all octave-related notes are represented by the same note. Musical chords (such as "C major" or "F minor") can be represented as multisets of points on this circle. The configuration space of n points on the circle, or the "space of all possible chords," is the orbifold $\mathbb{T}^n/\mathcal{S}_n$, where $\mathbb{T}^n$ is the n-torus. Chord types (such as "major" or "dominant seventh") are equivalence classes of multisets of points on the circle, modulo translation. The configuration space of chord types is $\mathbb{T}^{n-1}/\mathcal{S}_n$ [**2, 12**]. When musicians say "dominant seventh chords are quite close to diminished sevenths," or "the melodic minor scale is very similar to the diatonic," they are referring to distances in this quotient.

If "chords" correspond to states or income distributions, then "voice leadings" represent *changes* between states. In Western music, horizontal melodies define a sequence of bijections between the pitches in adjacent vertical chords, as in Figure 2(a); these "voice leadings" are represented by the images of line segments in the orbifolds $\mathbb{T}^n/\mathcal{S}_n$. The geometry here is closely analogous to the economic case, as can be seen by the resemblance between Figures 1 and 2; however, there are additional complications arising from the fact that the underlying one-dimensional space is a circle. Traditional music pedagogy enjoins students to connect successive chords in a way that minimizes the overall distance traveled by the voices or instruments. But theorists do not specify precisely what quantity is to be minimized. Is it the largest distance moved by any voice, the sum of all distances, or something else entirely?

Here again, it seems implausible that there is some particular metric that exactly captures composers' inherently vague intuitions about musical distance; it is highly implausible that Mozart or Chopin, when improvising at the piano, thought very systematically about "the distance between chords," let alone that their thinking can be modeled by some particular norm such as L^2 or L^3. Nor is it really plausible that composers' improvisatory practice itself, as distinct from their conscious thinking, should determine one such metric. Once again, we have very good musical reasons to endorse the No-Crossings Principle. The avoidance of voice-crossings, besides being central to traditional pedagogy, is well-grounded in listener psychology (since "auditory streaming" often leads listeners to associate notes by register [**7**]) and the ergonomics of instrumental performance (since crossings are often difficult to realize on instruments such as pianos). Furthermore, Tymoczko [**12**] has argued that compositional practice

[7] As already noted, D'Agostino and Dardanoni [**3**] propose "order sensitivity" and "value sensitivity" conditions that are strict versions of the No-Crossings Principle and the Monotonicity Principle, respectively. However, D'Agostino and Dardanoni supplement these conditions with a range of additional constraints.

suggests that composers think of musical distance in a way that is compatible with submajorization, since they use similar voice leadings between chords, regardless of whether the notes are close together or far apart in pitch space.[8]

Unlike the economic case, musical logic gives us no reason to focus on any particular hyperplane when considering distances between chord types, or equivalence classes of chords under translation (musical transposition). Therefore, we must draw on the full apparatus developed in Section 3. Here, the notion of **T**-closeness allows us to compare "distances" between chord types without committing ourselves to any particular metric. For example, theorists have been interested in determining how evenly a particular chord is distributed around the pitch-class circle [**13**]. **T**-closeness allows us to assert that the major triad (C, E, G) is closer to an even division of the octave than is (C, F, G) no matter how we measure musical distance. As shown in Appendix A.3, the multiset $W = \{0, 0, 1\}$, representing the change from (C, E, G) = $\{0, 4, 7\}$ to (C, E, G♯) = $\{0, 4, 8\}$ is **T**-smaller than the multiset $V = \{0, -1, 1\}$, representing the change from (C, F, G) = $\{0, 5, 7\}$ to $\{0, 4, 8\}$. In fact, it is possible to show that **T**-closeness produces a univocal evenness ranking of all the three-note chord types playable on an ordinary piano.

There are a number of other applications in which the goal is to measure simultaneous changes, but there is no obvious reason to choose one metric over another. In transportation and location problems, we think of states as representing locations of goods, and changes as representing transportation of goods from one location to another; the size of a change here represents the "cost" of transporting goods from one point to other. The usefulness of submajorization in this context has been described by Ogryczak [**11**], who notes that analogous problems arise in measuring economic inequality. In curve-fitting and error-analysis one wants to measure the overall distance from a set of data points to a hypothesized curve; here again, there is often no obvious reason, other than practicality, for favoring a particular norm. Our hope is that a unified discussion may help to reveal the similar conceptual issues arising in these various contexts. Submajorization might in some cases provide a more principled alternative to simply selecting a particular metric, allowing us to demarcate relationships that are independent of this arbitrary choice.

A. APPENDIX.

A.1. Euclidean bounds on the submajorization ball.

Lemma 3. *Let* $X = \{x_1, \ldots, x_n\} \in \mathbb{R}^n_+$, *let*

$$r = \min\left\{ j^{-1/2}\Sigma_j(X) \mid 1 \le j \le n \right\},$$

and let

$$R = \max\left\{ n^{1/2}\Sigma_1(X),\ \ \Sigma_n(X) \right\}.$$

Let $\mathbf{v}$ *be any vector in* $\mathbb{R}^n$.

1. *If* $|\mathbf{v}| \le r$, *then* $\{|v_i|\} \prec_w X$.
2. *If* $|\mathbf{v}| \ge R$, *then* $X \prec_w \{|v_i|\}$,

[8] This would not be true if the avoidance of crossings were in conflict with the goal of finding short voice leadings between subsequent chords; when notes were close together, the avoidance of crossings would dominate, leading composers to use inefficient voice leadings, reserving efficient voice leadings for situations in which notes were far apart.

where $|\mathbf{v}|$ *denotes the Euclidean length (*L^2*-norm) of* $\mathbf{v}$*. Moreover, these bounds are optimal:*

3. *For any* $\rho > r$*, there exists a* $\mathbf{z} \in \mathbb{R}^n$ *such that* $|\mathbf{z}| = \rho$ *and* $\{|z_i|\} \not\prec_w X$*.*
4. *For any* $\rho < R$*, there exists a* $\mathbf{z} \in \mathbb{R}^n$ *such that* $|\mathbf{z}| = \rho$ *and* $X \not\prec_w \{|z_i|\}$*.*

Proof.

1. Observe that for any nonnegative numbers $y_1, \ldots, y_j$, we have that

$$y_1^2 + \cdots + y_j^2 \leq (y_1 + \cdots + y_j)^2 \leq j(y_1^2 + \cdots + y_j^2).$$

If $|\mathbf{v}| \leq r$, then for $1 \leq j \leq n$,

$$(\Sigma_j\{|v_i|\})^2 = (|v_{[1]}| + \cdots + |v_{[j]}|)^2 \leq j(v_{[1]}^2 + \cdots + v_{[j]}^2) \leq jr^2 \leq (\Sigma_j(X))^2.$$

We conclude that X submajorizes $\{|v_i|\}$.

2. Suppose $X \not\prec_w \{|v_i|\}$; we claim that $|\mathbf{v}| < R$. Choose j such that $\Sigma_j(X) > \Sigma_j(\{|v_i|\}) = S$. Note that $|v_{[j]}| \leq S/j$. Let $W = (S - (j-1)|v_{[j]}|, |v_{[j]}|, \ldots, |v_{[j]}|)$. Then $\Sigma_j(W) = S$ and $|\mathbf{v}| \leq |W|$ because $\{|v_i|\} \prec_w W$ and the L^2-norm is strongly isotone. To see why $\{|v_i|\} \prec_w W$, note that the largest j elements of $\{|v_i|\}$ are the result of applying a series of Dalton transfers to the largest j elements of W, while the remaining elements of $\{|v_i|\}$ are smaller than the corresponding elements of W by monotonicity. If follows that W belongs to the one-parameter family of multisets

$$\{(S - (j-1)a, a, \ldots, a) \mid 0 \leq a \leq S/j\}.$$

A simple optimization argument shows that no multiset in this collection (in particular, W) has Euclidean norm more than the maximal value of $\{S,\ n^{1/2}S/j\}$. Since $S < \Sigma_j(X) \leq \Sigma_n(X)$ and $S/j < \Sigma_j(X)/j \leq \Sigma_1(X)$, we conclude that $|\mathbf{v}| \leq |W| < R = \max\left\{n^{1/2}\Sigma_1(X),\ \Sigma_n(X)\right\}$.

3. If $\rho > r$, then there exists a j, $1 \leq j \leq n$, such that

$$\rho > j^{-1/2}\Sigma_j(X).$$

Let $\mathbf{z} \in \mathbb{R}^n$ be the vector with j nonzero coordinates of the form

$$\mathbf{z} = \rho j^{-1/2}(1, \ldots, 1, 0, \ldots, 0).$$

Now, $|\mathbf{z}| = \rho$, and

$$\Sigma_j(\{|z_i|\}) = \rho j^{1/2} > \Sigma_j(X),$$

so X does not submajorize $\{|z_i|\}$.

4. If $n^{1/2}\Sigma_1(X) \geq \Sigma_n(X)$, then let $\mathbf{z} = \rho n^{-1/2}\mathbf{1}$. If $n^{1/2}\Sigma_1(X) < \Sigma_n(X)$, then let $\mathbf{z} = \rho(1, 0, \ldots, 0)$. ∎

A.2. Algorithm for determining T-smallness. We sketch an $O(n^4)$ algorithm for determining whether a multiset W is $\mathbf{T}$-smaller than a multiset V, where both multisets have n elements.

1. Find the $n(n-1)$ elements of S as in Theorem 2.
2. For each $s \in S$ and each odd j, find the set of y such that

$$\Sigma_j(\{|w_i - y|\}) \leq \Sigma_j(\{|v_i - s|\}). \tag{9}$$

   We claim that this computation is $O(n^2)$ for each s. Since the Σ_j functions are convex, the set of solutions to inequality (9) is either empty or a closed interval or point. In general, solving an equation of the form $f(y) \leq a$, where f is convex and piecewise linear with k known nonlinear points, requires an algorithm of order k. Evaluate f at each of the k points and determine on which of its linear segments $f(x) - a$ either changes sign or equals the zero function (there are at most two segments). This step requires $k+1$ computations. Since $f(x) - a$ is linear on these segments, we can easily find its zeros. The total number of nonlinear points of the Σ_j functions as j ranges from 1 to n is $n(n-1)$, so this step is $O(n^2)$ for each s.
3. If inequality (9) has no solutions for some j, we conclude that W is not **T**-smaller than V. Otherwise, the set of solutions to the inequality consists of n closed intervals. Comparing the lower endpoints of the intervals to the upper endpoints requires n^2 computations. If every lower endpoint is no larger than all the upper endpoints, W is **T**-smaller than V.

Step 1 involves $O(n^2)$ computations and steps 2 and 3 involve $O(n^2)$ computations for each $s \in S$. Since S has $n(n-1)$ elements, the algorithm is $O(n^4)$.

A.3. Examples. We claim that $W = \{1, 2, 4\}$ is **T**-smaller than $V = \{2, 2, 5\}$. In this case, S, the set of midpoints between pairs of elements of V, is $S = \{2, 3.5\}$. First, $\{|v_i - 2|\} = \{0, 0, 3\}$; we must find y such that $\{|w_i - y|\} \prec_w \{0, 0, 3\}$. Setting $y = 2$ gives us $\{1, 0, 2\}$, which suffices. Second, $\{|v_i - 3.5|\} = \{1.5, 1.5, 1.5\}$; we must find y such that $\{|w_i - y|\} \prec_w \{1.5, 1.5, 1.5\}$. Setting $y = 2.5$ gives us $\{1.5, 0.5, 1.5\}$, which suffices. If we disregard the effects of translation, then we conclude that $\{1, 2, 4\}$ is closer to an even distribution of values than $\{2, 2, 5\}$.

In musical contexts, we are interested in knowing which chord is closest to an even division of the octave. The augmented triad—the equivalence class containing (C, E, G♯) and its transpositions—is the "most even" three-note chord. We claim that the major triad (C, E, G) is closer to an even division of the octave than is (C, F, G), no matter how we measure musical distance. In other words, the multiset $W = \{0, 0, 1\}$, representing the distance between (C, E, G) $= \{0, 4, 7\}$ and (C, E, G♯) $= \{0, 4, 8\}$ is **T**-smaller than the multiset $V = \{0, -1, 1\}$, representing the distance between (C, F, G) $= \{0, 5, 7\}$ and $\{0, 4, 8\}$. In this case, $S = \{-0.5, 0, 0.5\}$. For all $s \in S$, $\{|v_i - s|\}$ submajorizes $\{|w_i|\}$.

ACKNOWLEDGMENTS. We thank two anonymous reviewers, Roland Benabou, Albert Marshall, Barry Arnold, and, especially, Ingram Olkin for their helpful comments on this paper.

REFERENCES

1. C. Callender, Continuous transformations, *Mus. Theory Online* **10** (2004)
2. C. Callender, I. Quinn, D. Tymoczko, Generalized voice-leading spaces, *Science* **320** (2008) 346–348; available at http://dx.doi.org/10.1126/science.1153021.
3. M. D'Agostino, V. Dardanoni, What's so special about Euclidean distance? *Soc. Choice Welf.* **33** (2009) 211–233; available at http://dx.doi.org/10.1007/s00355-008-0353-5.
4. H. Dalton, The measurement of the inequality of incomes, *Econom. Journal* **30** (1920) 348–361.
5. G. Fields, *Distribution and Development: A New Look at the Developing World*, Russell Sage Foundation and MIT Press, New York, Cambridge, and London, 2001.

6. G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.
7. D. Huron, Tone and voice: A derivation of the rules of voice-leading from perceptual principles, *Mus. Perception* **19** (2001) 1–64; available at http://dx.doi.org/10.1525/mp.2001.19.1.1.
8. M. O. Lorenz, Methods of measuring the concentration of wealth, *Publ. Amer. Statistical Assoc.* **9** (1905) 209–219.
9. A. W. Marshall, I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Mathematics in Science and Engineering, Vol. **143**, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979.
10. T. Mitra, E. A. Ok, The measurement of income mobility: a partial ordering approach, *Econom. Theory* **12** (1998) 77–102; available at http://dx.doi.org/10.1007/s001990050211.
11. W. Ogryczak, Inequality measures and equitable locations, *Ann. Oper. Res.* **167** (2009) 61–86; available at http://dx.doi.org/10.1007/s10479-007-0234-9.
12. D. Tymoczko, The geometry of musical chords, *Science* **313** (2006) 72–74; available at http://dx.doi.org/10.1126/science.1126287.
13. ———, *A Geometry of Music. Oxford Studies in Music Theory*, Oxford University Press, Oxford, 2011.
14. B. Zheng, Minimum distribution-sensitivity, poverty aversion, and poverty orderings, *J. Econom. Theory* **95** (2000) 116–137; available at http://dx.doi.org/10.1006/jeth.2000.2687.

RACHEL WELLS HALL is an associate professor of mathematics at Saint Joseph's University, where she researches and teaches both mathematics and music. She is writing a book entitled *The Sound of Numbers: A Tour of Mathematical Music Theory*. She has toured the Mid-Atlantic states since 1995 with the folk trio Simple Gifts and recorded three albums. She plays English concertina, piano, and fiddle.
Department of Mathematics, Saint Joseph's University, Philadelphia, PA 19131
rhall@sju.edu

DMITRI TYMOCZKO is a composer and music theorist who teaches at Princeton University. His book *A Geometry of Music* is available from Oxford University Press, and his CD *Beat Therapy* is available from Bridge Records. His music is influenced by a range of styles, including rock and jazz.
Department of Music, Princeton University, Princeton, NJ 08544
dmitri@princeton.edu

Does Numerology Allow a Group to Have Two Identity Elements?

Benoît Rittaud, Université Paris-13, Institut Galilée, Laboratoire Analyse, Géométrie et Applications.

Laurent Vivier, Université Paris Diderot, Laboratoire de Didactique André Revuz.

In numerology, the pseudoscience that focuses on describing the character of individuals [**1**], a specific binary operation in the set $\mathcal{N} := \{0, \ldots, 9\}$ is often considered. To add "numerologically" two elements of $\mathcal{N}$, first consider their usual sum, then, if bigger than 9, replace it by the sum of its digits. For 8 and 7, for example, we get $8 + 7 = 15$, then $1 + 5 = 6$, so the "numerological sum" of 8 and 7 is: $8 \oplus 7 = 6$.

You might think that all of this is nothing more than usual addition modulo 9. But wait: here, the set $\mathcal{N}$ contains ten elements, not nine as for $\mathbb{Z}/9\mathbb{Z}$. Moreover, since adding 9 to an integer does not change its value modulo 9, $(\mathcal{N}, \oplus)$ seems to contain two identity elements, 9 and 0, a fact that is proved to be impossible from the very beginning of group theory.

Explanation: 0 is indeed an identity element, and 9 is *almost* another one, since $9 \oplus n = n$ for any $n \in \mathcal{N} \ldots$ except for $n = 0$. Hence, $(\mathcal{N}, \oplus)$ provides a limit example of the impossibility of having two identity elements in a set embedded with a binary operation.

Let us emphasize the fact that $(\mathcal{N}, \oplus)$ is *not* itself a group (exercice: prove it!), but that this is not a problem here, since the uniqueness of an identity element is a consequence of the very definition of an identity element, and does not rely on an assumption of a group structure.

REFERENCES

1. Underwood Dudley, *Numerology, Or What Pythagoras Wrought*, The Mathematical Association of America, 1997.