Fechnerian Scaling: Dissimilarity Cumulation Theory

Ehtibar N. Dzhafarov and Hans Colonius

Contents

1 Introduction

1.1 What is it about?

In 1860 Gustav Theodor Fechner published the two-volume *Elemente der Psychophysik*. From this event one can date scientific psychology, firmly grounded in mathematics and experimental evidence. One of the main ideas introduced in Fechner's book is that of measuring subjective differences between stimuli **a** and **b** by means of summing (or integrating) just noticeable (or infinitesimal) differences in the interval of stimuli separating **a** and **b**. For Fechner, stimuli of a given kind are always represented by positive reals, so that the interval between them is well-defined.

We use the term "Fechnerian Scaling" to designate any method of computing distances in a stimulus space by means of *cumulating* (summing, integrating) values of a *dissimilarity function* for pairs of "neighboring" stimuli. The term *"dissimilarity cumulation"* can be used as a synonym of "Fechnerian Scaling" or else as designating an abstract mathematical theory of which Fechnerian Scaling is the main application.

A *stimulus space* is a set of stimuli endowed with a structure imposed on this set by an observer's judgments. Thus, a set of all visible aperture colors such that for each pair of colors we have a number indicating how often they appear identical to an observer if presented side by side is an example of a stimulus space. Stimuli in a stimulus space are referred to as its *points*, and generally are denoted by boldface lowercase letters: $\mathbf{x}_k, \mathbf{a}, \mathbf{b}^{(\omega)}$, etc. Dissimilarity function is a generalization of the notion of a *metric*, mapping pairs of stimuli $(\mathbf{x}, \mathbf{y})$ into non-negative numbers $D(\mathbf{x}, \mathbf{y})$. On a very general level, with minimal assumptions about the structure of a stimulus space being considered, Fechnerian Scaling is implemented by summing pairwise dissimilarities $D(\mathbf{x}_1, \mathbf{x}_2)$, $D(\mathbf{x}_2, \mathbf{x}_3)$, etc. along *finite chains* of points $\mathbf{a} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1} = \mathbf{b}$. The distance from $\mathbf{a}$ to $\mathbf{b}$ is then computed as the infimum of these cumulated values over the set of all such chains. Thus obtained distances from $\mathbf{a}$ to $\mathbf{b}$ and from $\mathbf{b}$ to $\mathbf{a}$ need not be the same, and to obtain a conventional, symmetric distance, in Fechnerian Scaling one adds these distances together.

In more specialized stimulus spaces, finite chains can be replaced with continuous or even continuously differentiable *paths*. In the latter case the cumulation is replaced with integration along a path of a certain quantity*, submetric function* $F(\mathbf{x}, \mathbf{u})$, that depends on the location $\mathbf{x}$ of a point and the velocity $\mathbf{u}$ with which it moves along the path. The submetric function is a measure of local discriminability of $\mathbf{x}$ from its "immediate" neighbors $\mathbf{x} + \mathbf{u}\mathrm{d}x$, and it can be empirically estimated by means of one of Fechner's methods for measuring differential thresholds. The original methods are based on one's ability to compare stimuli in terms of "greater than" with respect to some property (brightness, loudness, extent, etc.) In more general situations, stimuli can be compared by a variety of methods based on one's ability to judge whether two stimuli are the same or different.

The structure defining a stimulus space on a set of stimuli is always imposed by an observer's judgements of the stimuli rather than by the way stimuli are measured as physical objects. In this sense, the structure of stimulus space is a psychological rather physical construct. For instance, a drawing of human face has a complex physical description, but if, for example, the faces are compared in terms of greater-less with respect to some property, such as "beauty," then, provided certain assumptions are satisfied, a set of all possible face drawings may form a unidimensional continuum mappable on an interval of reals. However, physical descriptions of the stimuli typically have some properties (e.g., order, closeness) suggestive of the respective properties of the judgements. For instance, if $\mathbf{a}$ and $\mathbf{b}$ have very similar physical descriptions, one can usually expect the results of their comparisons with any stimulus $\mathbf{c}$ to also be very similar — the consideration we use, e.g., in constructing a differential-geometric version of Fechnerian Scaling.

1.2 Unidimensional Fechnerian Scaling

Various aspects of Fechner's original theory are subject to competing interpretations because they are not presented in his writings with sufficient clarity. The following therefore is not a historical account. Rather it is a modern theory that preserves the spirit of Fechner's idea of cumulation of small differences.

Let us assume that stimuli of a particular kind are represented (labeled, encoded) by values on an interval of positive real numbers $[t, u[$, where t is the absolute threshold value, and u is an appropriately defined upper threshold, or infinity. (Throughout this chapter, half-open or open intervals of reals will always be presented in the form $[t, u[$, $]t, u]$, $]t, u[$, using only square brackets.) The space structure on $[t, u[$ is defined by a *psychometric function* $\gamma(\mathbf{x}, \mathbf{y})$ that gives us the probability with which a stimulus $\mathbf{y}$ (represented by a value $y \in [t, u[$) is judged to be greater than stimulus $\mathbf{x}$ (represented by a value $x \in [t, u[$). In this special case, it is convenient to simply replace stimuli with their representations, and write x, y in place of $\mathbf{x}, \mathbf{y}$:

$$\gamma(x, y) = \Pr[y \text{ is judged to be greater than } x]. \quad (1)$$

We will make the simplifying assumption that

$$\gamma(y, x) = 1 - \gamma(x, y), \quad (2)$$

with the consequence

$$\gamma(x, x) = 1/2. \quad (3)$$

This will allow us to proceed in this special case without introducing the notions of observation areas and canonical transformations that are fundamental for the general theory.

Next, we will make a relatively innocuous assumption that $\gamma(x, y)$ is strictly increasing in y in the vicinity of $y = x$, and that it is continuously differentiable in y at $y = x$. That is, the derivative

$$F(x) = \left.\frac{\partial \gamma(x, y)}{\partial y}\right|_{y=x} \quad (4)$$

exists, is positive, and continuous in x. This is the slope of the psychometric function at its median, and the intuitive meaning of the differential $F(x)\,\mathrm{d}x$ is that it is proportional to the dissimilarity between x and its "immediate" neighbor, $x + \mathrm{d}x$. We can write this as

$$D(x, x + \mathrm{d}x) = cF(x)\,\mathrm{d}x,$$

where c is a positive constant specific to a given stimulus space. The intuition of cumulation of differences in this unidimensional setting is captured by the summation property

$$D(a, b) = D(a, c) + D(c, b),$$

for any $a \leq c \leq b$ in stimulus set $\mathfrak{S}$. It follows that

$$D(a,b) = c\int_a^b F(x)\,\mathrm{d}x. \tag{5}$$

This quantity can be interpreted as the subjective distance between a and b for any $a \leq b$ in $\mathfrak{S}$. We take the relations (4) and (5) for the core of the Fechnerian Scaling in stimulus continua (presented here with simplifying assumptions).

1.3 Historical Digression: Fechner's Law

One can easily check that the *logarithmic law* advocated by Fechner,

$$D(t,x) = K\log\frac{x}{t}, x \geq t, \tag{6}$$

where K is a positive constant, corresponds to

$$F(x) = \frac{K}{x}, \tag{7}$$

which can be viewed as a differential form of the so-called *Weber's law.* Recall that t designates absolute threshold.

This is an example of the so-called *psychophysical law*, the relationship between a physical description of a stimuli x, and the value of $D(x,t)$, referred to as the *magnitude of sensation.* In this chapter we attach little importance to this or other psychophysical laws. In view of the generalization of Fechnerian Scaling to stimulus spaces with more complex descriptions than real numbers, such laws have limited scope of applicability.

Nevertheless, it is appropriate to take a historical detour and look at how Fechner's law was justified by Fechner himself, in the second volume of his landmark work*, Elemente der Psychophysik.* The relationship (6) is referred by Fechner as the *measurement formula (Massformel).* More generally, Fechner's law can be written as

$$D(a,b) = D(t,b) - D(t,a) = K\log\frac{b}{a}, b \geq a \geq t, \tag{8}$$

for two stimulus magnitudes a, b. Fechner calls this *difference formula* (*Unterschiedsformel*).

In an addendum to his work *Zen Avesta*, Fechner describes how the idea of this law occurred to him in the morning of October 22, 1950 (this date is nowadays celebrated as the *Fechner Day*): he had an insight that an arithmetic progression of sensation magnitude should correspond to a geometric progression of stimulus magnitudes. Fechner's insight on that day is all one needs to derive the law, as logarithm is the only function with non-chaotic behavior that can transform a geometric progression into an arithmetic one. The derivation of the law, however, had to wait for 10 more year before it appeared in vol. 2 of the *Elemente der Psychophysik*, in two different forms (Chapters 16 and 17).

Unfortunately, the second volume has not been translated into English. As we learn from a letter written by E. G. Boring to S. Rosenzweig on February 23, 1968, "Just now I'm spending long hours working over translation into English of the second volume of the Fechner's *Elemente*, because put literally into English it is about as dull and confusing and sometimes uninterpretable as it always was in the German. Holt, Rinehart and Winston published the first volume and someday we will get this second half done, but we do not have much help after NIH stopped supporting translation. We have to get it done by little bits." It seems that Boring has not completed this work.

By a historical happenstance, one of Fechner's derivations of his law was criticized as mathematically incorrect, and the other simply forgotten. In addition, the law itself was criticized as empirically incorrect. However, by careful examination of the premises of Fechner's derivations the mathematical criticisms can be deflected, while empirical falsifications of the law often involve empirical procedures (e.g., direct estimation of sensation magnitudes) that go beyond those Fechner would consider legitimate. In a paper of rejoinders published in 1877, Fechner reacts to the criticisms known to him and makes a bold prediction for the future: "The tower of Babel was never finished because the workers could not reach an understanding on how they should build it; my psychophysical edifice will stand because the workers will never agree on how to tear it down."

The difficulty in understanding Fechner's derivations of his logarithmic law is in that he uses the term "Weber's law" in the meaning that is logically independent of the empirical law established by Ernst Heinrich Weber (which Fechner, to add to the confusion, also calls "Weber's law"). According to Weber's law, if x and $x + \Delta x$ are separated by a *just-noticeable difference*, then

$$\frac{\Delta x}{x} = c^*, \tag{9}$$

where c^* is a constant with respect to x (but generally depends on the stimulus continuum used). In Fechner's mathematical derivations, however, the term "Weber's law" stands for the following statement, essentially a form of his October 1850 insight :

> *the subjective dissimilarity $D(t,b) - D(t,a)$ between stimuli with physical magnitudes a and b (provided $t \leq a \leq b$) is determined by the ratio of these magnitudes, b/a.*

We propose calling this statement "W-principle" to disentangle it from Weber's law. The only relationship between the W-principle and Weber's law can be established through so-called "*Fechner's postulate*," according to which all just-noticeable differences Δx (within a given continuum) are subjectively equal,

$$D(x, x + \Delta x) = c. \tag{10}$$

Any two of the three statements, Fechner's postulate, Weber's law (in its usual meaning), and the W-principle implies the third.

In Chapter 17 of the *Elemente*, Fechner derives his law by using a novel for his time method of *functional equations*. He presents the W-principle as

$$\psi(b) - \psi(a) = F\left(\frac{b}{a}\right)$$

where $\psi(x)$ denotes $D(t, x)$, and observes that this implies

$$F\left(\frac{c}{b}\right) + F\left(\frac{b}{a}\right) = F\left(\frac{c}{a}\right),$$

for any $t \leq a \leq b \leq c$. This in turn means that

$$F(x) + F(y) = F(xy),$$

for any $x, y \geq 1$. Fechner recognizes in this the functional equation introduced only 40 years earlier by Augustin-Louis Cauchy, who showed that its only continuous solution is

$$F(x) = K \log x, x \geq 1.$$

It is known now (Aczél, 1987) that continuity can be replaced with many other regularity assumptions, including monotonicity and nonnegativity, and that it is sufficient to assume that the equation holds only in an arbitrarily small vicinity of 1 (i.e., for very similar stimuli only). It follows that

$$\psi(b) - \psi(a) = K \log \frac{b}{a}, b \geq a \geq t,$$

which is Fechner's *Unterschiedsformel.*

In Chapter 16 of the *Elemente*, Fechner derives the same relationship in a different way. He presents the functional equation as

$$\psi(b) - \psi(a) = G\left(\frac{b-a}{a}\right),$$

and by assuming that G is differentiable at zero gets the differential equation

$$\psi'(x)\,dx = G'(0)\,\frac{dx}{x},$$

whose integration once again leads to Fechner's logarithmic formula.

The novelty of the method of functional equations in the mid-XIX's century is probably responsible for the fact that the Chapter 17 derivation was universally overlooked by Fechner's contemporaries (and then, as it seems, forgotten altogether). The derivation in Chapter 16, through differential equations, was, by contrast, common in Fechner's time, which may be the reason Fechner placed it first. This derivation has been criticized as mathematically or logically flawed by Fechner's contemporaries and modern authors alike. The common interpretation has been that it is based on Fechner's postulate

$$\psi(x + \Delta x) - \psi(x) = c.$$

He is thought to have combined this with Weber's law

$$\frac{\Delta x}{x} = c^*,$$

to arrive at

$$\psi(x + \Delta x) - \psi(x) = \frac{c}{c^*} \frac{\Delta x}{x}.$$

Finally, Fechner is thought to have invoked an "expediency principle" (*Hülfsprinzip*) to illegitimately replace the finite differences with differentials,

$$d\psi = \frac{c}{c^*} \frac{dx}{x}.$$

The integration of this equation with the boundary condition $\psi(x_0) = 0$ yields

$$\psi(x) = \frac{c}{c^*} \log \frac{x}{x_0}.$$

It has been pointed out that this derivation is internally contradictory because it implies

$$\psi(x + \Delta x) - \psi(x) = \frac{c}{c^*} \log \frac{x + \Delta x}{x} = \frac{c}{c^*} \log(1 + c^*),$$

which is not the same as the postulated

$$\psi(x + \Delta x) - \psi(x) = c.$$

Boring's characterization of Fechner's book as "dull and confusing and sometimes uninterpretable" being true, it is not easy to refute this criticism. However, it is clear that Fechner uses neither the Fechner postulate nor Weber's law in deriving his law, although he accepts the truth of both. As explained above, he makes use of the W-principle (which he calls "Weber's law"). It follows from his derivation that if Weber's law holds in addition to the W-principle, then

$$\psi(x + \Delta x) - \psi(x) = K \log(1 + c^*) = c,$$

which is indeed a constant (Fechner's postulate proved as a theorem). As Fechner points out in a book of rejoinders, if the Weber fraction c^* is sufficiently small, the constant K *approximately* equals c/c^*, as in the criticized formula. The "expediency principle" which Fechner's critics especially disparage seems to be nothing more than an inept and verbose explanation of the elementary fact (used in the Chapter 16 derivation) that if a function $f(x)$ is differentiable at zero, then $df(x)$ is proportional to dx.

1.4 Observation areas and canonical transformation

The elementary but fundamental fact is that if an observer is asked to compare two stimuli, $\mathbf{x}$ and $\mathbf{y}$, they must differ in some respect that allows the observer to identify them as two distinct stimuli. For instance, in the pair written as $(\mathbf{x}, \mathbf{y})$, the first argument, $\mathbf{x}$, may denote the stimulus presented chronologically first, followed by $\mathbf{y}$. Or $\mathbf{x}$ may always be presented above or to the left of $\mathbf{y}$. In perceptual pairwise comparisons, the stimuli must differ in their spatial and/or temporal locations, but the defining properties of $\mathbf{x}$ and $\mathbf{y}$ in the pair $(\mathbf{x}, \mathbf{y})$ may vary. Thus, two line segments to be compared in length may be presented in varying pairs of distinct spatial locations, but one of the line segments may always be vertical (and written first in the pair, $\mathbf{x}$) and the other horizontal (written second, $\mathbf{y}$).

Formally, this means that a stimulus space involves two stimulus sets rather than one. Denoting them $\mathfrak{S}_1^{\star\star}$ (for $\mathbf{x}$-stimuli) and $\mathfrak{S}_2^{\star\star}$ (for $\mathbf{y}$-stimuli), we call them the first and the second *observation areas,* respectively. The space structure is imposed on the Cartesian product of these observation areas by a function

$$\phi^{\star\star} : \mathfrak{S}_1^{\star\star} \times \mathfrak{S}_2^{\star\star} \to R, \tag{11}$$

where R may be a set of possible responses, or possible probabilities of a particular response.

We say that two stimuli $\mathbf{x}, \mathbf{x}' \in \mathfrak{S}_1^{\star\star}$ are *psychologically equal* if

$$\phi^{\star\star}(\mathbf{x}, \mathbf{y}) = \phi^{\star\star}(\mathbf{x}', \mathbf{y})$$

for any $\mathbf{y} \in \mathfrak{S}_2^{\star\star}$. Similarly, $\mathbf{y}, \mathbf{y}' \in \mathfrak{S}_2^{\star\star}$ are psychologically equal if

$$\phi^{\star\star}(\mathbf{x}, \mathbf{y}) = \phi^{\star\star}(\mathbf{x}, \mathbf{y}'),$$

for any $\mathbf{x} \in \mathfrak{S}_1^{\star\star}$. One can always relabel the elements of the observation areas by assigning identical labels to all psychologically equal stimuli. For instance, all metameric colors may be encoded by the same RGB coordinates irrespective of their spectral composition. Objects of different color but of the same weight will normally be labeled identically in a task involving hefting and deciding which of two objects is heavier.

Let us denote by $\mathfrak{S}_1^{\star}$ and $\mathfrak{S}_2^{\star}$ the observation areas in which psychologically equal stimuli are equal. The function $\phi^{\star\star}$ is then redefined into

$$\phi^{\star} : \mathfrak{S}_1^{\star} \times \mathfrak{S}_2^{\star} \to R. \tag{12}$$

We will illustrate this transformation by a toy example. Let the original function be

ϕ^{*}	$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_3$	$\mathbf{y}_4$	$\mathbf{y}_5$	$\mathbf{y}_6$	$\mathbf{y}_7$
$\mathbf{x}_1$	0.7	0.6	0.3	0.4	0.4	0.4	0.4
$\mathbf{x}_2$	0.5	0.3	0.4	0.2	0.2	0.2	0.2
$\mathbf{x}_3$	0.5	0.3	0.4	0.2	0.2	0.2	0.2
$\mathbf{x}_4$	0.2	0.1	0.5	0.3	0.3	0.3	0.3
$\mathbf{x}_5$	0.2	0.1	0.5	0.3	0.3	0.3	0.3
$\mathbf{x}_6$	0.1	0.3	0.8	0.6	0.6	0.6	0.6
$\mathbf{x}_7$	0.1	0.3	0.8	0.6	0.6	0.6	0.6

The first observation area, $\mathfrak{S}_1^{\star\star}$, comprises stimuli $\{\mathbf{x}_1, \ldots, \mathbf{x}_7\}$ (e.g., weights placed on one's left palm), the second observation area, $\mathfrak{S}_2^{\star\star}$, comprises stimuli $\{\mathbf{y}_1, \ldots, \mathbf{y}_7\}$ (weights placed on one's left palm), and the entries in the matrix above are values of $\phi^* (\mathbf{x}, \mathbf{y})$, an arbitrary function mapping $(\mathbf{x}, \mathbf{y})$-pairs into real numbers (say, the probabilities of deciding that the two weights differ in heaviness). If two rows (or columns) of the matrix are identical, then the two corresponding $\mathbf{x}$-stimuli (respectively, $\mathbf{y}$-stimuli) are psychologically equal, and can be labeled identically. Thus, the stimuli $\mathbf{x_2}, \mathbf{x_3}$ and $\mathbf{x_4}, \mathbf{x_5}$ and $\mathbf{x_6}, \mathbf{x_7}$ and $\mathbf{y}_4, \mathbf{y}_5, \mathbf{y}_6, \mathbf{y}_7$ are all psychologically equal and they can be replaced by a single symbol, respectively. The redefined spaces $\mathfrak{S}_1^\star$ and $\mathfrak{S}_1^\star$ are then as follows,

$$\begin{array}{cccccccc} \mathfrak{S}_1^{\star\star}: & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 & \mathbf{x}_6 & \mathbf{x}_7 \\ & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ \mathfrak{S}_1^{\star}: & \mathbf{x}_a & \mathbf{x}_b & \mathbf{x}_b & \mathbf{x}_c & \mathbf{x}_c & \mathbf{x}_d & \mathbf{x}_d \end{array}, \quad \begin{array}{cccccccc} \mathfrak{S}_2^{\star\star}: & \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \mathbf{y}_4 & \mathbf{y}_5 & \mathbf{y}_6 & \mathbf{y}_7 \\ & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ \mathfrak{S}_2^{\star}: & \mathbf{y}_a & \mathbf{y}_b & \mathbf{y}_c & \mathbf{y}_d & \mathbf{y}_d & \mathbf{y}_d & \mathbf{y}_d \end{array},$$

and the function ϕ^* transforms into $\phi^\star$ accordingly,

$\phi^\star$	$\mathbf{y}_a$	$\mathbf{y}_b$	$\mathbf{y}_c$	$\mathbf{y}_d$
$\mathbf{x}_a$	0.7	0.6	0.3	0.4
$\mathbf{x}_b$	0.5	0.3	0.4	0.2
$\mathbf{x}_c$	0.2	0.1	0.5	0.3
$\mathbf{x}_d$	0.1	0.3	0.8	0.6

As another example, consider the function $\gamma(x, y)$ of the previous section, and assume that

$$\mathfrak{S}_1^{\star\star} = [t_1, u_1[, \mathfrak{S}_2^{\star\star} = [t_2, u_2[.$$

Assume that $\gamma(x, y)$ is strictly increasing in y and strictly decreasing in x. Then $\gamma(x, y) = \gamma(x, y')$ implies $y = y'$ and $\gamma(x, y) = \gamma(x', y)$ implies $x = x'$, so that in this case

$$\mathfrak{S}_1^{\star\star} = \mathfrak{S}_1^\star, \mathfrak{S}_2^{\star\star} = \mathfrak{S}_2^\star.$$

Staying with this example, $\gamma(x, y) = 1/2$ defines here the binary relation "is matched by": $x \in \mathfrak{S}_1^\star$ is matched by $y \in \mathfrak{S}_2^\star$ if and only if $\gamma(x, y) = 1/2$. The relation "$y \in \mathfrak{S}_2^\star$ is matched by $x \in \mathfrak{S}_1^\star$" is defined by the same condition, $\gamma(x, y) = 1/2$. The traditional psychophysical designation of this relation is that y is the *point of subjective equality* (PSE) for x (and then x is the PSE for y). The assumptions (2)-(3) made in the previous section do not hold generally. In particular, the psychometric function γ, as a rule, has a nonzero *constant error*, i.e., $\gamma(x, y) = 1/2$ does not imply $x = y$ (see Figure 1).

With the monotonicity assumptions about γ made above, if we also assume that the range of the function $y \mapsto \gamma(x, y)$ for every x includes the value $1/2$, and that the same is true for the range of the function $x \mapsto \gamma(x, y)$ for every y, then we have the following properties of the PSE relation (see Figure 2):

1. the PSE for every $x \in \mathfrak{S}_1^\star$ exists and is unique;

2. the PSE for every $y \in \mathfrak{S}_2^\star$ exists and is unique;

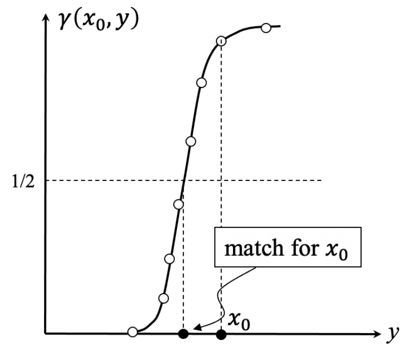

Figure 1: A "greater-less" psychometric function $y \mapsto \gamma(x, y)$ defined on an interval of real numbers.The function shows, for a fixed value of $x = x_0$, the probability with which y is judged to be greater than x_0 with respect to some designated property. The median value of y, one at which $\gamma(x_0, y) = \frac{1}{2}$, is taken to be a match, or point of subjective equality (PSE) for x_0, and the difference between x_0 and its PSE defines constant error. (Note that showing $\gamma(x, y)$ at a fixed value of x does not mean that the value of x was fixed procedurally in an experiment. The graph is simply a cross-section of $\gamma(x, y)$ at $x = x_0$.)

3. $y \in \mathfrak{S}_2^\star$ is a PSE for $x \in \mathfrak{S}_1^\star$ if and only if $x \in \mathfrak{S}_1^\star$ is a PSE for $y \in \mathfrak{S}_2^\star$.

We will assume that these properties generalize to any function $\phi^\star$ in (12). In other words, we assume that $\phi^\star$ is associated with a *bijective function* $\mathbf{h} : \mathfrak{S}_1^\star \longrightarrow \mathfrak{S}_2^\star$ such that for all $\mathbf{x} \in \mathfrak{S}_1^\star$ and $\mathbf{y} \in \mathfrak{S}_2^\star$,

(P1) $\mathbf{y}$ is a PSE for $\mathbf{x}$ if and only if $\mathbf{y} = \mathbf{h}(\mathbf{x})$;

(P2) $\mathbf{x}$ is a PSE for $\mathbf{y}$ if and only if $\mathbf{x} = \mathbf{h}^{-1}(\mathbf{y})$.

This makes the relation of "being a PSE of" or "being matched by" symmetric. As a result, one can always apply to the observation areas a *canonical transformation*

$$\mathbf{f} : \mathfrak{S}_1^\star \longrightarrow \mathfrak{S}, \mathbf{g} : \mathfrak{S}_2^\star \longrightarrow \mathfrak{S},$$

with $\mathbf{f}$ and $\mathbf{g}$ arbitrary except for

$$\mathbf{h} = \mathbf{g}^{-1} \circ \mathbf{f}.$$

A canonical transformation redefines the function $\phi^\star$ into

$$\phi : \mathfrak{S} \times \mathfrak{S} \longrightarrow R,$$

such that, for any ordered pair $(\mathbf{x}, \mathbf{y})$, one of the elements is a PSE for the other element if and only if $\mathbf{x} = \mathbf{y}$. We say that the stimulus space and the space-forming function ϕ here are in a *canonical form.*

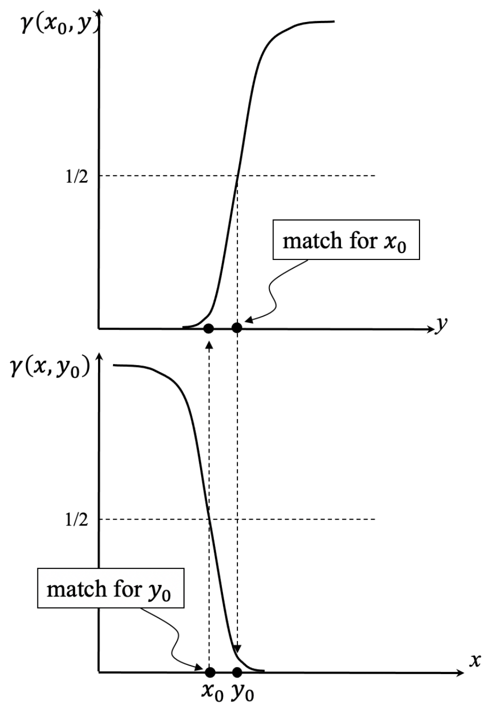

Figure 2: An illustration, for the psychometric function $\gamma\,(x, y)$, of the symmetry of the relation "to be a PSE for." The upper panel shows the function $y \mapsto \gamma\,(x, y)$ at $x = x_0$, and y_0 denotes the PSE for x_0. The lower panel shows the function $x \mapsto \gamma\,(x, y)$ at $y = y_0$, and x_0 then has to be the PSE for y_0. Conversely, if x_0 denotes the PSE for y_0 in the lower panel, then y_0 has to be the PSE for x_0 in the upper panel. This follows from the fact that in both cases the PSE is defined by $\gamma\,(x, y) = \frac{1}{2}$, and the assumption that both $x \mapsto \gamma\,(x, y)$ and $y \mapsto \gamma\,(x, y)$ are monotone functions whose range includes the value $\gamma = \frac{1}{2}$.

Let us use the toy example above for an illustration. We assume that the PSE for any $\mathbf{x}$ is defined here as $\mathbf{y}$ at which $\mathbf{y} \mapsto \phi^{\star}(\mathbf{x}, \mathbf{y})$ reaches its minimum; and the PSE for any $\mathbf{y}$ is defined as $\mathbf{x}$ at which $\mathbf{x} \mapsto \phi^{\star}(\mathbf{x}, \mathbf{y})$ reaches its minimum. The inspection of the matrix for $\phi^{\star}$ shows that the PSEs are well defined for both $\mathbf{x}$-stimuli and $\mathbf{y}$-stimuli:

$\phi^{\star}$	$\mathbf{y}_a$	$\mathbf{y}_b$	$\mathbf{y}_c$	$\mathbf{y}_d$
$\mathbf{x}_a$	0.7	0.6	$\boxed{0.3}$	0.4
$\mathbf{x}_b$	0.5	0.3	0.4	$\boxed{0.2}$
$\mathbf{x}_c$	0.2	$\boxed{0.1}$	0.5	0.3
$\mathbf{x}_d$	$\boxed{0.1}$	0.3	0.8	0.6

.

We also see that in each row the minimal value (shown boxed) is also minimal in its column. That is, $\mathbf{y}$ is a PSE for $\mathbf{x}$ if and only if $\mathbf{x}$ the PSE for $\mathbf{y}$. The graph of the bijective $\mathbf{h}$-function in the formulations of the properties P1 and P2 is given by the pairs

$$\{(\mathbf{x}_a, \mathbf{y}_c), (\mathbf{x}_b, \mathbf{y}_d), (\mathbf{x}_c, \mathbf{y}_b), (\mathbf{x}_d, \mathbf{y}_a)\}.$$

Simple relabeling then allows us to have all PSE-pairs on the main diagonal. Both $\mathfrak{S}_1^{\star}$ and $\mathfrak{S}_1^{\star}$ can be mapped into one and the same set $\mathfrak{S}$, e.g., as

$$\begin{array}{ccccc} \mathfrak{S}_1^{\star}: & \mathbf{x}_a & \mathbf{x}_b & \mathbf{x}_c & \mathbf{x}_d \\ & \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ \mathfrak{S}: & \mathbf{a} & \mathbf{b} & \mathbf{c} & \mathbf{d} \end{array}, \quad \begin{array}{ccccc} \mathfrak{S}_2^{\star}: & \mathbf{y}_c & \mathbf{y}_d & \mathbf{y}_b & \mathbf{y}_a \\ & \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ \mathfrak{S}: & \mathbf{a} & \mathbf{b} & \mathbf{c} & \mathbf{d} \end{array},$$

and $\phi^{\star}$ transforms into ϕ accordingly:

ϕ	**a**	**b**	**c**	**d**
a	$\boxed{0.3}$	0.4	0.6	0.7
b	0.4	$\boxed{0.2}$	0.3	0.5
c	0.5	0.3	$\boxed{1}$	0.2
d	0.8	0.6	0.3	$\boxed{0.1}$

.

To apply canonical transformation to our second example, the psychometric function $\gamma(x, y)$, let us assume that $\gamma(x, y) = 1/2$ holds if and only if $y = h(x)$ for some homeomorphic mapping h (i.e., such that both h and h^{-1} are continuous.) Since $\mathfrak{S}_1^{*} = \mathfrak{S}_1^{\star} = [t_1, u_1[$ and $\mathfrak{S}_2^{*} = \mathfrak{S}_2^{\star} = [t_2, u_2[$, $\mathfrak{S}$ can always be chosen in the form $[t, u[$, by choosing any two homeomorphisms

$$f : [t_1, u_1[\to [t, u[, g : [t_2, u_2[\to [t, u[,$$

such that $g^{-1} \circ f \equiv h$. Note, however, that this only ensures compliance with (3), but not with (2).

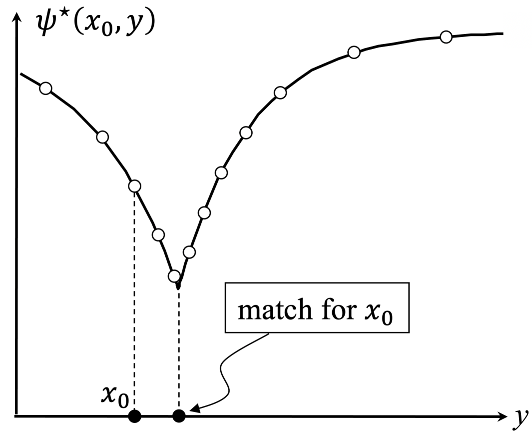

Figure 3: A "same-different" psychometric function $y \mapsto \psi^{\star}(x, y)$ defined on an interval of real numbers. The function shows, for a fixed value of $x = x_0$, the probability with which y is judged to be different from x_0 (generically or with respect to a designated property). The value of y at which $\psi^{\star}(x_0, y)$ reaches its minimum is taken to be a match, or point of subjective equality (PSE) for x_0.

1.5 Same-different judgments

The greater-than comparisons are possible only with respect to a designated characteristic, such as loudness or beauty. It is clear, however, that no such characteristic can reflect all relevant aspects of the stimuli being compared. Moreover, it is not certain that the characteristic's values are always comparable in terms of greater-less, given a sufficiently rich stimulus set. Thus, it may not be clear to an observer which of two given faces is more beautiful, and even loudness may not be semantically unidimensional if the sounds are complex. The same-different comparisons have a greater scope of applicability, and do not have to make use of designated characteristics. The role of the stimulus-space-defining function ϕ^{*} of the previous section in this case is played by

$$\psi^{*}(\mathbf{x}, \mathbf{y}) = \Pr[\mathbf{y} \text{ is judged to be different from } \mathbf{x}], \tag{13}$$

with $\mathbf{x} \in \mathfrak{S}_1^{\star\star}$ and $\mathbf{y} \in \mathfrak{S}_2^{\star\star}$. To be different here means to differ in any respect other than the conspicuous difference between the two observations areas. Thus, if $\mathbf{x}$ is a visual stimulus always presented to the left of $\mathbf{y}$,this difference in spatial locations does not enter in the judgments of whether $\mathbf{x}$ and $\mathbf{y}$ are different or the same. Of course, it is also possible to ask whether the two stimuli differ in a particular respect, such as color or shape.

The reduction of $(\psi^{*}, \mathfrak{S}_1^{\star\star}, \mathfrak{S}_2^{\star\star})$ to $(\psi^{\star}, \mathfrak{S}_1^{\star}, \mathfrak{S}_2^{\star})$, in which psychologically equal stimuli are equal, is effected by assigning an identical label to any $\mathbf{x}, \mathbf{x}' \in \mathfrak{S}_1^{\star\star}$ such that

$$\psi^{\star\star}(\mathbf{x}, \mathbf{y}) = \psi^{\star\star}(\mathbf{x}', \mathbf{y})$$

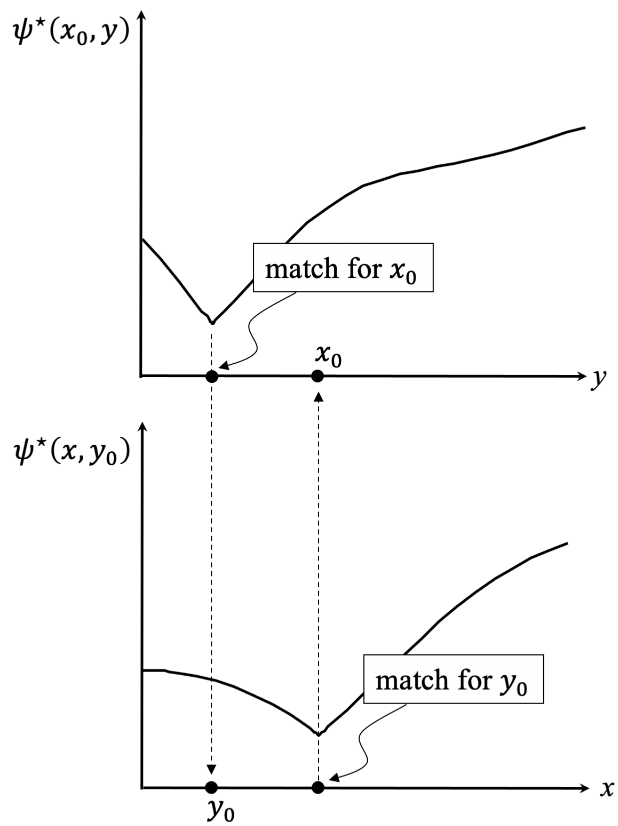

Figure 4: An illustration, for the psychometric function $\psi^\star(x, y)$ in Figure 3, of the symmetry of the relation "to be a PSE for." The upper panel shows the function $y \mapsto \psi^\star(x, y)$ at $x = x_0$, and y_0 denotes the PSE for x_0. The lower panel shows the function $x \mapsto \psi^\star(x, y)$ at $y = y_0$, and x_0 is shown to be PSE for y_0. Conversely, if x_0 denotes the PSE for y_0 in the lower panel, then y_0 is shown to be the PSE for x_0 in the upper panel. Unlike in the case of the "great-less" psychometric function (Figure 2), here the symmetry of the PSE relation is an assumption rather than a consequence of other properties of the function $\psi^\star$.

for all $\mathbf{y} \in \mathfrak{S}_2^{\star\star}$, and similarly for the second observation area.

The PSE relation for the function $\psi^\star$ is defined as follows (see Figure 3): $\mathbf{y} \in \mathfrak{S}_2^\star$ is a PSE for $\mathbf{x} \in \mathfrak{S}_1^\star$ if

$$\psi^\star(\mathbf{x}, \mathbf{y}) < \psi^\star(\mathbf{x}, \mathbf{y}') \text{ for all } \mathbf{y}' \neq \mathbf{y}.$$

Analogously, $\mathbf{x} \in \mathfrak{S}_1^\star$ is a PSE for $\mathbf{y} \in \mathfrak{S}_2^\star$ if

$$\psi^\star(\mathbf{x}, \mathbf{y}) < \psi^\star(\mathbf{x}', \mathbf{y}) \text{ for all } \mathbf{x}' \neq \mathbf{x}.$$

In accordance with the previous section, we assume the existence of a bijection $\mathbf{h} : \mathfrak{S}_1^\star \longrightarrow \mathfrak{S}_2^\star$ such that

$$\begin{array}{c} \psi^\star(\mathbf{x}, \mathbf{h}(\mathbf{x})) < \psi^\star(\mathbf{x}, \mathbf{y}) \text{ for all } \mathbf{y} \neq \mathbf{h}(\mathbf{x}), \\ \psi^\star\left(\mathbf{h}^{-1}(\mathbf{y}), \mathbf{y}\right) < \psi^\star(\mathbf{x}, \mathbf{y}) \text{ for all } \mathbf{x} \neq \mathbf{h}^{-1}(\mathbf{y}). \end{array} \tag{14}$$

That is, we assume that the PSEs in the space $(\psi^\star, \mathfrak{S}_1^\star, \mathfrak{S}_2^\star)$ exist, are unique, and that $\mathbf{y}$ is the PSE for $\mathbf{x}$ if and only if $\mathbf{x}$ is the PSE for $\mathbf{y}$. We refer to this property as the *law of Regular Minimality*. In this chapter it should be taken as part of the definition of the functions we are dealing with rather than an empirical claim.

Now, any canonical transformation, as described above, yields a probability function

$$\psi : \mathfrak{S} \times \mathfrak{S} \longrightarrow [0, 1], \tag{15}$$

such that, for any $\mathbf{a}, \mathbf{x}, \mathbf{y} \in \mathfrak{S}$, if $\mathbf{x} \neq \mathbf{a}$ and $\mathbf{y} \neq \mathbf{a}$, then

$$\psi(\mathbf{a}, \mathbf{a}) < \left\{ \begin{array}{l} \psi(\mathbf{x}, \mathbf{a}) \\ \psi(\mathbf{a}, \mathbf{y}) \end{array} \right. . \tag{16}$$

We will assume in the following that the discrimination probability function ψ is presented in this canonical form. This by no means implies that $\psi(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{y}, \mathbf{x})$, the order of the arguments continues to matter. We will continue to consider the two arguments in $\psi(\mathbf{x}, \mathbf{y})$ as belonging to the first and second observation areas, respectively.

2 Notation conventions

We now introduce notation conventions for the rest of this chapter. They in part codify and in part modify the notation used in the introductory section.

Let us agree that from now on real-valued functions of one or several points of a stimulus set will be indicated by strings without parentheses: $\psi\mathbf{ab}$ in place of $\psi(\mathbf{a}, \mathbf{b})$, $D\mathbf{abc}$ in place of $D(\mathbf{a}, \mathbf{b}, \mathbf{c})$, etc. Boldface lowercase letters denoting stimuli are merely labels, with no implied operations between them, so this notation is unambiguous. (In Section 8, lowercase boldface letters are also used to denote direction vectors, in which case the string convention is not used.) If a stimulus is represented by a real number we may conveniently confuse the two, and write, e.g., $\gamma(x,y)$ instead of the more rigorous $\gamma\mathbf{xy}$ with $\mathbf{x},\mathbf{y}$ represented by (or having values) x,y.

A finite sequence (or *chain*) $(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ of points in stimulus a set will be presented as a string $\mathbf{x}_1\ldots\mathbf{x}_n$. If a chain of stimuli is to be referred to without indicating its elements, then it is indicated by uppercase boldface letters. Thus $\mathbf{X}$ may stand for $\mathbf{abc}$, $\mathbf{Y}$ stand for $\mathbf{y}_1\ldots\mathbf{y}_n$, etc. If $\mathbf{X}=\mathbf{x}_1...\mathbf{x}_k$and $\mathbf{Y}=\mathbf{y}_1...\mathbf{y}_l$ are two chains, then

$$\begin{gathered}\mathbf{XY}=\mathbf{x}_1...\mathbf{x}_k\mathbf{y}_1...\mathbf{y}_l,\\ \mathbf{aXb}=\mathbf{ax}_1...\mathbf{x}_k\mathbf{b},\\ \mathbf{aXbYa}=\mathbf{ax}_1...\mathbf{x}_k\mathbf{by}_1...\mathbf{y}_l\mathbf{a},\\ \text{etc.}\end{gathered}$$

The number of elements in a chain $\mathbf{X}$ is its cardinality $|\mathbf{X}|$. Infinite sequences $\{x_1,\ldots,x_n,\ldots\}$, $\{\mathbf{x}_1,\ldots,\mathbf{x}_n,\ldots\}$, $\{\mathbf{X}_1,\ldots,\mathbf{X}_n,\ldots\}$, etc., are almost always indicated by their generic elements: numerical sequence $\{x_n\}$, stimulus sequence $\{\mathbf{x}_n\}$, sequence of chains $\{\mathbf{X}_n\}$, etc. Convergence of a sequence, such as $\mathbf{x}_n\to\mathbf{x}$, is understood as conditioned on $n\to\infty$. In a sequence of chains, the cardinality $|\mathbf{X}_n|$ is generally changing.

As mentioned earlier, we indicate intervals of reals (closed, open and half-open) by square-brackets: $[a,b]$, $[a,b[$, $]a,b]$, and $]a,b[$. Round-bracketed pairs of numbers of stimuli, (a,b) or $(\mathbf{a},\mathbf{b})$, always indicate an ordered pair.

Sets of stimuli are denoted by Gothic letters, $\mathfrak{S}$, $\mathfrak{S}_1^{\star\star}$, $\mathfrak{s}$, etc. For sets of chains and paths in stimulus spaces we use script letters, $\mathcal{C}$,$\mathcal{P}_a^b$, etc. For other types of sets we use blackboard and sans serif fonts on an ad hoc basis. The set of reals is denoted as usual $\mathbb{R}$.

3 Basics of Fechnerian Scaling

Using our new notation, and considering an at least two-element stimulus space $\mathfrak{S}$ in a canonical form, we have, for any distinct points $\mathbf{x}$ and $\mathbf{y}$ in $\mathfrak{S}$,

$$\begin{aligned}\Psi^{(1)}\mathbf{xy}&=\psi\mathbf{xy}-\psi\mathbf{xx}>0,\\ \Psi^{(2)}\mathbf{xy}&=\psi\mathbf{yx}-\psi\mathbf{xx}>0.\end{aligned}\tag{17}$$

We call the quantities $\Psi^{(1)}\mathbf{xy}$ and $\Psi^{(2)}\mathbf{xy}$ *psychometric increments* of the first and second kind, respectively. Both can be interpreted as ways of quantifying the intuition of a dissimilarity of $\mathbf{y}$ from $\mathbf{x}$. The order "from-to" is important here, as $\Psi^{(i)}\mathbf{yx}\neq\Psi^{(i)}\mathbf{yx}$ $(i=1,2)$.

In Fechnerian Scaling we use the psychometric increments to compute subjective distances in the spirit of Fechner's idea of cumulation of small dissimilarities. We will see that this cumulation can assume different forms, depending on the properties of a stimulus space. However, the general construction, applicable to all spaces, is as follows.

3.1 Step 1

First, we assume that both $\Psi^{(1)}$ or $\Psi^{(2)}$ are *dissimilarity functions*, in accordance with the following definition (to be explained and elaborated later on).

Definition 1. We say that $D : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}$ is a dissimilarity function if it has the following properties:

$\mathcal{D}1$(*positivity*) $D\mathbf{ab} > 0$ for any distinct $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$;

$\mathcal{D}2$ (*zero property*) $D\mathbf{aa} = 0$ for any $\mathbf{a} \in \mathfrak{S}$;

$\mathcal{D}3$ (*uniform continuity*) for any $\varepsilon > 0$ one can find a $\delta > 0$ such that, for any $\mathbf{a}, \mathbf{b}, \mathbf{a}', \mathbf{b}' \in \mathfrak{S}$,

$$\text{if } D\mathbf{aa}' < \delta \text{ and } D\mathbf{bb}' < \delta, \text{ then } |D\mathbf{a}'\mathbf{b}' - D\mathbf{ab}| < \varepsilon;$$

$\mathcal{D}4$ (*chain property*) for any $\varepsilon > 0$ one can find a $\delta > 0$ such that for any chain $\mathbf{aXb}$,

$$\text{if } D\mathbf{aXb} < \delta, \text{ then } D\mathbf{ab} < \varepsilon.$$

For the chain property, we need to define $D\mathbf{aXb}$.

Definition 2. Given a chain $\mathbf{X} = \mathbf{x}_1 ... \mathbf{x}_k$ in $\mathfrak{S}$, its *D-length* (or just *length* once D is specified) is defined as

$$D\mathbf{X} = \begin{cases} D\mathbf{x}_1\mathbf{x}_2 + ... + D\mathbf{x}_{k-1}\mathbf{x}_k & \text{if } |\mathbf{X}| > 1 \\ 0 & \text{if } |\mathbf{X}| \leq 1 \end{cases}.$$

Then, for a given pair of points $\mathbf{a}, \mathbf{b}$, the length of $\mathbf{aXb}$ is

$$D\mathbf{aXb} = \begin{cases} D\mathbf{ax}_1 + D\mathbf{X} + D\mathbf{x}_k\mathbf{b} & \text{if } |\mathbf{X}| > 0 \\ D\mathbf{ab} & \text{if } |\mathbf{X}| = 0 \end{cases}.$$

3.2 Step 2

Next, we consider the set $\mathcal{C}$ of all (finite) chains in $\mathfrak{S}$,

$$\mathcal{C} = \bigcup_{k=0}^{\infty} \mathfrak{S}^k,$$

and define

$$G\mathbf{ab} = \inf_{\mathbf{X} \in \mathcal{C}} D\mathbf{aXb}. \tag{18}$$

We will see below that the function $G : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}$ is a *quasimetric dissimilarity*, in accordance with the following definition.

Definition 3. Function $M : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}$ is a quasimetric dissimilarity function if it has the following properties:

$\mathcal{QM}1$ (*positivity*) $M\mathbf{ab} > 0$ for any distinct $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$;

$\mathcal{QM}2$ (*zero property*) $M\mathbf{aa} = 0$ for any $\mathbf{a} \in \mathfrak{S}$;

$\mathcal{QM}3$ (*triangle inequality*) $M\mathbf{ab} + M\mathbf{bc} \geq M\mathbf{ac}$ for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathfrak{S}$.

$\mathcal{QM}4$ (*symmetry in the small*) for any $\varepsilon > 0$ one can find a $\delta > 0$ such that $M\mathbf{ab} < \delta$ implies $M\mathbf{ba} < \varepsilon$, for any $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$.

To relate quasimetric dissimilarity to two familiar terms, a function satisfying $\mathcal{QM}1$-$\mathcal{QM}3$ is called a *quasimetric*, and a quasimetric is called a *metric* if it satisfies the property

$\mathcal{M}4$ (symmetry) $M\mathbf{ab} = M\mathbf{ba}$, for any $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$.

Quasimetric dissimilarity therefore can be viewed as a concept intermediate between quasimetric and metric. More importantly, however, a quasimetric dissimilarity (hence also a metric), as shown below, is a special form of dissimilarity, whereas quasimetric generally is not (see Figure 5).

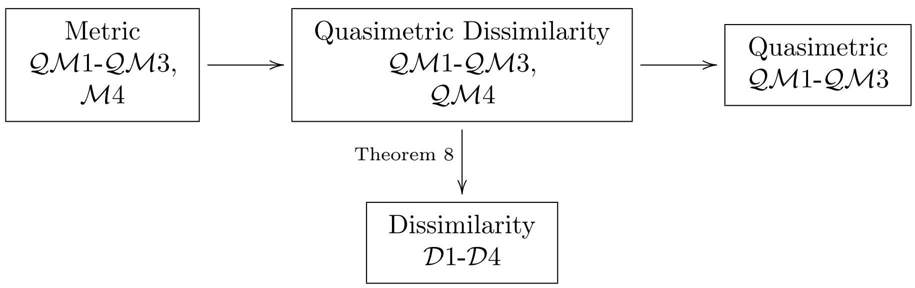

Figure 5: Interrelations between metric-like concepts. Arrows between the boxes stand for "is a special case of."

3.3 Step 3

The quasimetric dissimilarities

$$G^{(1)}\mathbf{ab} = \inf_{\mathbf{X}\in\mathcal{C}} \Psi^{(1)}\mathbf{aXb}$$

and

$$G^{(2)}\mathbf{ab} = \inf_{\mathbf{X}\in\mathcal{C}} \Psi^{(2)}\mathbf{aXb}$$

are generally different. However, we will see below that

$$G^{(1)}\mathbf{ab} + G^{(1)}\mathbf{ba} = G^{(2)}\mathbf{ab} + G^{(2)}\mathbf{ba}, \tag{19}$$

and this quantity is clearly a metric. We will denote it $\overleftrightarrow{G}\mathbf{ab}$, and interpret it as the *Fechnerian distance* between $\mathbf{a}$ and $\mathbf{b}$ in the canonical stimulus space $\mathfrak{S}$. The double-arrow in $\overleftrightarrow{G}$ is suggestive of the following way of presenting this quantity:

$$\overleftrightarrow{G}\mathbf{ab} = \inf_{\mathbf{X},\mathbf{Y}\in\mathcal{C}} \Psi^{(1)}\mathbf{aXbYa} = \inf_{\mathbf{X},\mathbf{Y}\in\mathcal{C}} \Psi^{(2)}\mathbf{aXbYa}, \tag{20}$$

the $\mathbf{aXbYa}$ (equivalently, $\mathbf{bYaXb}$) being a closed chain containing the points $\mathbf{a}$ and $\mathbf{b}$.

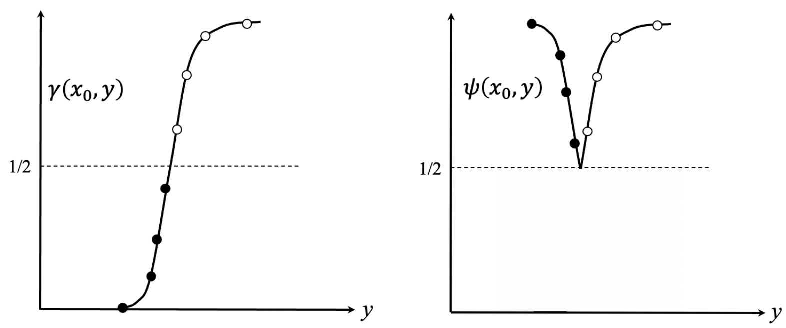

Figure 6: An illustration of how a "greater-less" discrimination probability function (on the left) can be redefined into a "same-different"-type discrimination probability function.

3.4 Subsequent development

The function $\overleftrightarrow{G}$ is, in a sense, the ultimate goal of Fechnerian Scaling. However, the metric structure of a space is part of its geometry, and this is what a full theory of Fechnerian Scaling deals with. In discrete spaces, consisting of isolated points, the general definition of $\overleftrightarrow{G}$ provides the algorithm for computing it. In more structured spaces, however, the Fechnerian metric may be computed in specialized ways. Rather than considering all possible chains, in some spaces one integrates infinitesimal dissimilarities along continuous paths and seeks the shortest paths. In still more structured spaces this leads to a generalized form of Finsler geometry, where computations of distances are based on indicatrices or submetric functions.

The psychometric increments $\Psi^{(1)}$ and $\Psi^{(2)}$ are at the foundation of Fechnerian Scaling. In this chapter they are defined through the psychometric function ψ in (13), which is usually associated with the same-different version of the *method of constant stimuli.* In this method, same-different judgements are recorded for repeatedly presented multiple pairs of stimuli, as indicated, e.g., by the open circles in Figure 3. However, virtually any pairwise comparison procedure can be, in principle, used to define analogues of $\Psi^{(1)}$ and $\Psi^{(2)}$. For instance, if the observer judges pairs of stimuli in terms of "greater-less" with respect to some property, the psychometric function γ of Figure 1 (assuming it is in a canonical form) can be converted into a ψ-like function by putting

$$\psi\mathbf{xy} = \begin{cases} \gamma\mathbf{xy} & \text{if } \gamma\mathbf{xy} \geq \frac{1}{2} \\ 1 - \gamma\mathbf{xy} & \text{if } \gamma\mathbf{xy} < \frac{1}{2} \end{cases} .$$

This is illustrated in Figure 6 for the case $\mathfrak{S}$ is an interval of real numbers. The psychometric increments then are defined as

$$\Psi^{(1)}\mathbf{xy} = \left|\gamma\mathbf{xy} - \frac{1}{2}\right|, \Psi^{(2)}\mathbf{xy} = \left|\gamma\mathbf{yx} - \frac{1}{2}\right|.$$

Some experimental procedures may yield dissimilarity values $D\mathbf{ab}$ "directly." Thus, in one of the procedures of *Multidimensional Scaling* (MDS), observers are presented pairs of stimuli and asked to numerically estimate "how different they are." Then, for every pair of stimuli $\mathbf{a}, \mathbf{b}$, some measure of central tendency of these numerical estimates can be hypothesized to be an efficient estimator of a dissimilarity function

$$\Psi^{(1)}\mathbf{ab} = \Psi^{(2)}\mathbf{ba} = D\mathbf{ab}.$$

If one can establish that $D\mathbf{aa} = 0$ for all stimuli and that $D\mathbf{ab} > 0$ for distinct $\mathbf{a}, \mathbf{b}$, then the stimulus space is in a canonical form, and the hypothesis that D is a dissimilarity function cannot be falsified on any finite set of data. However, given sufficient amount of data, one can usually falsify the hypothesis that $D\mathbf{ab}$ is a quasimetric, by establishing that $D\mathbf{ab}$ violates the triangle inequality. In such situations, MDS seeks a monotone transformation $g \circ D$ that would yield a quasimetric. Dissimilarity cumulation offers an alternative approach, to use D to compute by (18) a quasimetric dissimilarity G and then symmetrize it by (20). We will return to this situation in Section 9.

4 Dissimilarity function

The properties $\mathcal{D}3$ and $\mathcal{D}4$ of Definition 1 are more conveniently presented in terms of convergence of sequences. Let us introduce convergence in a stimulus space.

Definition 4. Given two sequences of points in $\mathfrak{S}$, $\{\mathbf{a}_n\}$ and $\{\mathbf{b}_n\}$, we say that $\mathbf{a}_n$ and $\mathbf{b}_n$ *converge to each other, and write this* $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$, if $D\mathbf{a}_n\mathbf{b}_n \to 0$. In the special case $\mathbf{b}_n \equiv \mathbf{b}$, we say that $\mathbf{a}_n$ *converges to* $\mathbf{b}$,and write $\mathbf{a}_n \to \mathbf{b}$.

The property $\mathcal{D}3$ (uniform continuity) then can be presented as follows:

$$\text{if } \mathbf{a}_n \leftrightarrow \mathbf{a}'_n \text{ and } \mathbf{b}_n \leftrightarrow \mathbf{b}'_n, \text{ then } D\mathbf{a}'_n\mathbf{b}'_n - D\mathbf{a}_n\mathbf{b}_n \to 0.$$

In other words, D is a uniformly continuous function (Figure 7).

It is clear that $\mathbf{a}_n \leftrightarrow \mathbf{a}_n$ is true for any sequence $\{\mathbf{a}_n\}$ (because $D\mathbf{a}_n\mathbf{a}_n = 0$). Assuming that $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$, we can use $\mathcal{D}3$ to observe that

$$\mathbf{a}_n \leftrightarrow \mathbf{b}_n \text{ and } \mathbf{a}_n \leftrightarrow \mathbf{a}_n \Longrightarrow D\mathbf{a}_n\mathbf{a}_n - D\mathbf{b}_n\mathbf{a}_n \to 0 \Longleftrightarrow D\mathbf{b}_n\mathbf{a}_n \to 0.$$

But $D\mathbf{b}_n\mathbf{a}_n \to 0$ means $\mathbf{b}_n \leftrightarrow \mathbf{a}_n$, and we obtain the following proposition.

Theorem 5 (*symmetry in the small*)**.** *For any* $\{\mathbf{a}_n\}, \{\mathbf{b}_n\}$,

$$\mathbf{a}_n \leftrightarrow \mathbf{b}_n \text{ iff } \mathbf{b}_n \leftrightarrow \mathbf{a}_n.$$

This justifies the terminology (convergence to each other) and notation in the definition of $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$.

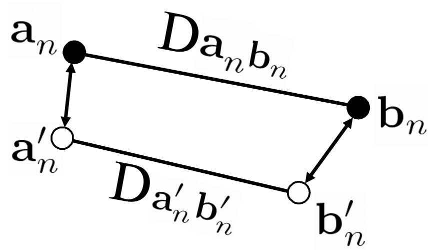

Figure 7: Illustration of the uniform continuity of D. The dissimilarities $D\mathbf{a}_n\mathbf{b}_n$ and $D\mathbf{a}'_n\mathbf{b}'_n$ converge to each other as $\mathbf{a}_n$ with $\mathbf{a}'_n$ converge to each other and $\mathbf{b}_n$ with $\mathbf{b}'_n$ converge to each other.

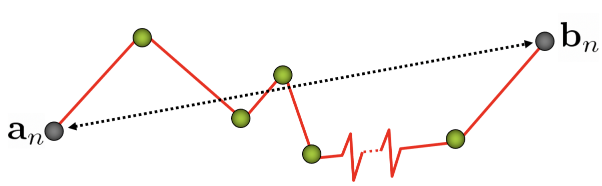

Figure 8: Illustration of the chain property of D. If the overall length of the chains $\mathbf{X}_n$ connecting $\mathbf{b}_n$ to $\mathbf{a}_n$ tends to zero, then $\mathbf{a}_n$ and $\mathbf{b}_n$ converge to each other. This property is nontrivial only if $|\mathbf{X}_n|$, the number of elements in the chains, tends to infinity. If it is bounded, $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$ is a consequence of the transitivity of the $\leftrightarrow$ relation (not discussed in the text, but easily established).

Property $\mathcal{D}4$ (chain property) can be presented as follows: for any sequences $\{\mathbf{a}_n\}, \{\mathbf{b}_n\}$ in $\mathfrak{S}$ and $\{\mathbf{X}_n\}$ in $\mathcal{C}$ (the set of chains),

$$\text{if } D\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n \to 0, \text{ then } \mathbf{a}_n \leftrightarrow \mathbf{b}_n. \tag{21}$$

Figures 8 provides an illustration.

The properties $\mathcal{D}1$-$\mathcal{D}4$ are logically independent: none of them is a consequence of the remaining three. This is proved by constructing examples, for each of these properties, that violate this property while conforming to the others. For example, to prove the independence of $\mathcal{D}4$, consider $\mathfrak{S} = \mathbb{R}$, and let $D\mathbf{xy} = (x - y)^2$ (where x, y are the numerical values representing $\mathbf{x}, \mathbf{y}$, respectively). The function D clearly satisfies $\mathcal{D}1$-$\mathcal{D}3$. However, for any points $\mathbf{a}, \mathbf{b}$, if the elements of a chain $\mathbf{X}_n$ subdivide $[a, b]$ into n equal parts, then

$$D\mathbf{a}\mathbf{X}_n\mathbf{b} = n\left(\frac{b-a}{n}\right)^2 \to 0,$$

while the value of $D\mathbf{ab}$ remains equal to $(b-a)^2$.

5 Quasimetric dissimilarity

We begin by establishing an important fact: the function G defined by 18 and the dissimilarity D are *equivalent in the small.*

Theorem 6. *For any* $\{\mathbf{a}_n\}, \{\mathbf{b}_n\}$,

$$\mathbf{a}_n \leftrightarrow \mathbf{b}_n \text{ iff } G\mathbf{a}_n\mathbf{b}_n \to 0.$$

To prove this, we first observe that $G\mathbf{ab} \geq 0$, as the infimum of nonnegative $D\mathbf{aXb}$. If $D\mathbf{a}_n\mathbf{b}_n \to 0$, we have

$$0 \leq G\mathbf{a}_n\mathbf{b}_n = \inf_{\mathbf{X}\in\mathcal{C}} D\mathbf{a}_n\mathbf{X}\mathbf{b}_n \leq D\mathbf{a}_n\mathbf{b}_n \to 0,$$

and this implies $G\mathbf{a}_n\mathbf{b}_n \to 0$. Conversely, $\inf_{\mathbf{X}\in\mathcal{C}} D\mathbf{a}_n\mathbf{X}\mathbf{b}_n \to 0$ means that for some sequence of chains $\{\mathbf{X}_n\}$, $D\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n \to 0$. By the chain property then, $D\mathbf{a}_n\mathbf{b}_n \to 0$.

Let us now see if G satisfies the properties defining a quasimetric dissimilarity, $\mathcal{QM}1$-$\mathcal{QM}4$ We immediately see that it satisfies the triangle inequality ($\mathcal{QM}3$):

$$G\mathbf{ab} \leq G\mathbf{ac} + G\mathbf{cb},$$

for any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathfrak{S}$. Indeed,

$$G\mathbf{ac} + G\mathbf{cb} = \inf_{\mathbf{X}\in\mathcal{C}} D\mathbf{aXc} + \inf_{\mathbf{Y}\in\mathcal{C}} D\mathbf{cYb} = \inf_{\mathbf{X},\mathbf{Y}\in\mathcal{C}} D\mathbf{aXcYb},$$

and the set of all possible $\mathbf{aXb}$ contains the set of all possible $\mathbf{aXcYb}$ chains. It is also easy to see that the function G is symmetric in the small ($\mathcal{QM}4$). Written in convergence terms, the property is

$$\text{if } G\mathbf{a}_n\mathbf{b}_n \to 0 \text{ then } G\mathbf{b}_n\mathbf{a}_n \to 0.$$

It is proved by observing that, by the previous theorem, if $G\mathbf{a}_n\mathbf{b}_n \to 0$ then $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$, and then $G\mathbf{b}_n\mathbf{a}_n \to 0$. Because we know that $G\mathbf{ab}$ is nonnegative, the properties $\mathcal{QM}1$ and $\mathcal{QM}2$ follow from

$$G\mathbf{ab} = \inf_{\mathbf{X}\in\mathcal{C}} D\mathbf{aXb} = 0 \Longrightarrow D\mathbf{aX}_n\mathbf{b} \to 0,$$

for some sequence of chains $\{\mathbf{X}_n\}$. But this means, by the chain property, $D\mathbf{ab} = 0$, which is true if and only if $\mathbf{a} = \mathbf{b}$. We have established therefore

Theorem 7. *The function G is a quasimetric dissimilarity.*

It is instructive to see why, as mentioned earlier and as its name suggests, any quasimetric dissimilarity, and G in particular, is a dissimilarity function. Let M satisfy the properties $\mathcal{QM}1$-$\mathcal{QM}4$. Then $\mathcal{D}1$ and $\mathcal{D}2$ are satisfied trivially. The property $\mathcal{D}3$ (uniform continuity) follows from the fact that, by the triangle inequality,

$$\begin{cases} M\mathbf{aa}' + M\mathbf{b}'\mathbf{b} & \geq M\mathbf{ab} - M\mathbf{a}'\mathbf{b}', \\ M\mathbf{a}'\mathbf{a} + M\mathbf{bb}' & \geq M\mathbf{a}'\mathbf{b}' - M\mathbf{ab}. \end{cases}$$

By the symmetry in the small property,

$$\begin{array}{c} M\mathbf{a}_n\mathbf{a}'_n \to 0 \Longleftrightarrow M\mathbf{a}'_n\mathbf{a}_n \to 0, \\ M\mathbf{b}'_n\mathbf{b}_n \to 0 \Longleftrightarrow M\mathbf{b}_n\mathbf{b}'_n \to 0, \end{array}$$

so these convergences imply

$$|M\mathbf{ab} - M\mathbf{a}'\mathbf{b}'| \to 0.$$

The chain property, $\mathcal{D}4$, follows from $M\mathbf{aXb} \geq M\mathbf{ab}$, by the triangle inequality. We have established therefore

Theorem 8. *Any quasimetric dissimilarity (hence also any metric) is a dissimilarity function.*

Let us now return to the to the definition of $G^{(1)}$, $G^{(2)}$, and $\overleftrightarrow{G}$. We need to establish (20), from which (19) follows. Given a chain $\mathbf{X} = \mathbf{x}_1\mathbf{x}_2...\mathbf{x}_k$, let us define the opposite chain $\mathbf{X}^\dagger$ as $\mathbf{x}_k\mathbf{x}_{k-1}...\mathbf{x}_1$. By straightforward algebra,

$$\Psi^{(1)}\mathbf{X} = \sum_{i=1}^{k-1} \Psi^{(1)}\mathbf{x}_i\mathbf{x}_{i+1} = \sum_{i=1}^{k-1} \left(\psi\mathbf{x}_i\mathbf{x}_{i+1} - \psi\mathbf{x}_i\mathbf{x}_i\right),$$

$$\Psi^{(2)}\mathbf{X}^\dagger = \sum_{i=1}^{k-1} \Psi^{(2)}\mathbf{x}_{i+1}\mathbf{x}_i = \sum_{i=1}^{k-1} \left(\psi\mathbf{x}_i\mathbf{x}_{i+1} - \psi\mathbf{x}_{i+1}\mathbf{x}_{i+1}\right).$$

It follows that

$$\Psi^{(1)}\mathbf{X} - \Psi^{(2)}\mathbf{X}^\dagger = \psi\mathbf{x}_k\mathbf{x}_k - \psi\mathbf{x}_1\mathbf{x}_1.$$

In particular, if the chain is closed,$\mathbf{x}_k = \mathbf{x}_1$, we have

$$\Psi^{(1)}\mathbf{X} = \Psi^{(2)}\mathbf{X}^\dagger.$$

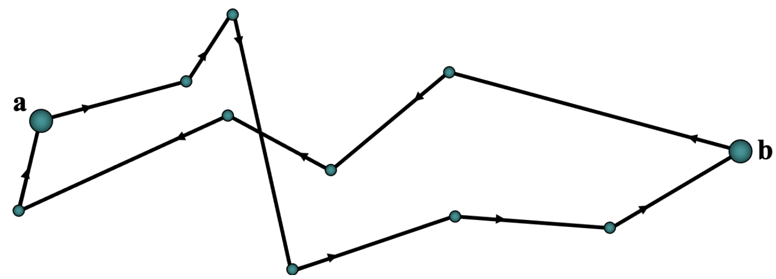

Figure 9: For any closed chain $\mathbf{X}$ containing points $\mathbf{a}, \mathbf{b}$, the value of $\Psi^{(1)}\mathbf{X}$ is the same as the value of $\Psi^{(2)}\mathbf{X}^{\dagger}$, the same chain traversed in the opposite direction.

That is, the $\Psi^{(1)}$-length of a closed chain equals the $\Psi^{(2)}$-length of the same chain traversed in the opposite direction (see Figure 9). Applying this to a chain $\mathbf{aXbYa}$,

$$\Psi^{(1)}\mathbf{aXbYa} = \Psi^{(2)}\mathbf{aY}^{\dagger}\mathbf{bX}^{\dagger}\mathbf{a},$$

whence

$$\inf_{\mathbf{X},\mathbf{Y}\in\mathcal{C}} \Psi^{(1)}\mathbf{aXbYa} = \inf_{\mathbf{Y}^{\dagger},\mathbf{X}^{\dagger}\in\mathcal{C}} \Psi^{(2)}\mathbf{aY}^{\dagger}\mathbf{bX}^{\dagger}\mathbf{a}.$$

Clearly, the set of all possible pairs of chains $(\mathbf{X}, \mathbf{Y})$ is the same as the set of all pairs $\left(\mathbf{Y}^{\dagger}, \mathbf{X}^{\dagger}\right)$, and by simple renaming,

$$\inf_{\mathbf{X},\mathbf{Y}\in\mathcal{C}} \Psi^{(1)}\mathbf{aXbYa} = \inf_{\mathbf{X},\mathbf{Y}\in\mathcal{C}} \Psi^{(2)}\mathbf{aXbYa}.$$

This proves the following

Theorem 9. *For any* $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$*,*

$$G^{(1)}\mathbf{ab} + G^{(1)}\mathbf{ba} = G^{(2)}\mathbf{ab} + G^{(2)}\mathbf{ba} = \overleftrightarrow{G}\mathbf{ab}.$$

The function $\overleftrightarrow{G}$ *is a metric.*

The last statement is an immediate corollary of Theorem 7.

One can think of other ways of combining quasimetric dissimilarities $G^{(1)}\mathbf{ab}$ and $G^{(1)}\mathbf{ba}$ into a metric, such as

$$\max\left(G^{(1)}\mathbf{ab}, G^{(1)}\mathbf{ba}\right), \sqrt{G^{(1)}\mathbf{ab} + G^{(1)}\mathbf{ba}}, \text{etc.}$$

Denoting a combination like this $f\left(G^{(1)}\mathbf{ab}, G^{(1)}\mathbf{ba}\right)$, the natural requirements are that

(i) it should equal $f\left(G^{(2)}\mathbf{ab}, G^{(2)}\mathbf{ba}\right)$, and

(ii) $f(x, x) \propto x$.

The latter requirement ensures that if $G^{(1)}\mathbf{ab}$ always equals $G^{(1)}\mathbf{ba}$ (i.e., it is already a metric), then $f\left(G^{(1)}\mathbf{ab}, G^{(1)}\mathbf{ab}\right)$ is just a multiple of $G^{(1)}\mathbf{ab}$. Clearly, function $\overleftrightarrow{G}$ satisfies these requirements. In fact, up to a scaling coefficient, it is the only such function.

Theorem 10. *Function $f(x,y)$ satisfies **(i)** and **(ii)** above for all stimulus spaces if and only if $f(x,y) = k(x+y)$.*

For a proof, consider a canonical space $(\psi, \mathfrak{S})$ with $\mathfrak{S} = \{\mathbf{a}, \mathbf{b}\}$. It is easy to see that for any $s, z \in (0,1]$ one can find probabilities $\psi\mathbf{aa}$, $\psi\mathbf{ab}$, $\psi\mathbf{ba}$, $\psi\mathbf{bb}$ satisfying

$$\begin{array}{l} G_1\mathbf{ab} = \psi\mathbf{ab} - \psi\mathbf{aa} = s \\ G_1\mathbf{ba} = \psi\mathbf{ba} - \psi\mathbf{bb} = s \\ G_2\mathbf{ab} = \psi\mathbf{ba} - \psi\mathbf{aa} = 2s - z \\ G_2\mathbf{ba} = \psi\mathbf{ab} - \psi\mathbf{bb} = z \end{array}.$$

Then the requirement **(i)** means that

$$f(s,s) = f(2s - z, z)$$

should hold for all $s, z \in (0,1]$. That is, $f(2s - z, z)$ depends on s only, and we have

$$f(x,y) = g(x+y).$$

Putting $x = y = \frac{u}{2}$, it follows from the requirement **(ii)** that

$$g(x+y) = g(u) = ku,$$

for some $k > 0$. So, our definition of $\overleftrightarrow{G}$ is not arbitrary, except for choosing $k = 1$.

6 Dissimilarity cumulation in discrete spaces

6.1 Direct computation of distances

A discrete stimulus space $(\mathfrak{S}, D)$ consists of isolated points, i.e., for every $\mathbf{x} \in \mathfrak{S}$,

$$\inf_{\mathbf{y} \in \mathfrak{S}, \mathbf{y} \neq \mathbf{x}} D\mathbf{xy} > 0. \tag{22}$$

Although genuinely discrete and even finite stimulus spaces exist (e.g., the Morse codes of letters and digits studied for their confusability), this special case is important not so much in its own right as because any set of empirical data forms a discrete (in fact, finite) space. This means, e.g., that even if an observer is asked to compare colors or sounds, the data will form a finite set of pairs associated with some estimate of discriminability. If the data are sufficiently representative, the results of applying to them Fechnerian Scaling of discrete spaces should provide a good approximation to the theoretical Fechnerian Scaling using dissimilarity cumulation along continuous or smooth paths, as described later in this chapter.

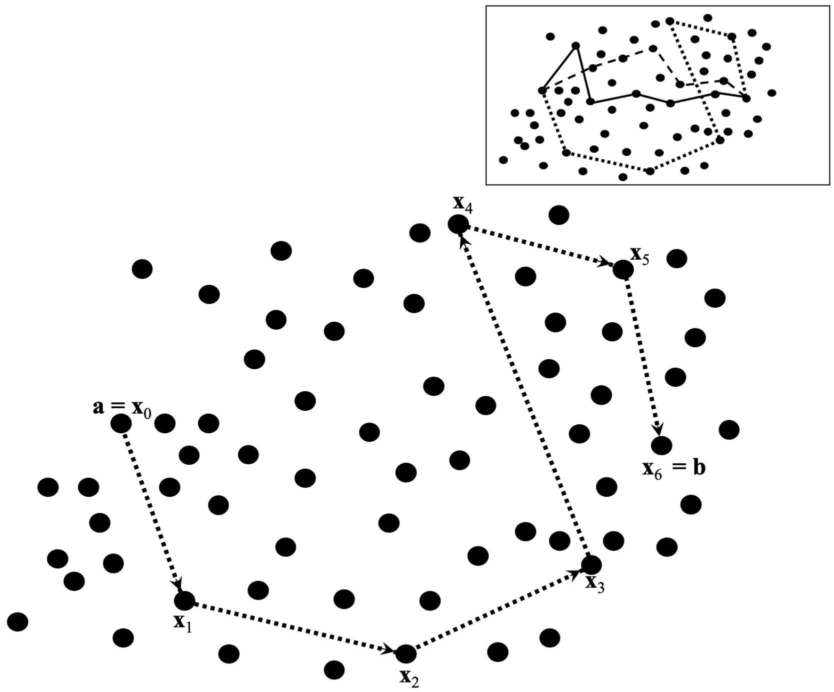

Figure 10: Dissimilarity cumulation is discrete spaces. One considers all possible chains connecting a point **a** to a point **b** and seeks the infimum of their D-lengths. In a finite space this infimum is the smallest among the D-lengths, and it may be attained by more than one chain.

As mentioned earlier, in discrete spaces the general definition of a Fechnerian distance directly determines the algorithm of computing them: one tries all possible chains leading from one point to another (with some obvious heuristics shrinking this set), and finds their infimum or, in special cases, minimum. This is illustrated in Figure 10.

Let us return to the toy example presented in Section 1.4, and assume that the function ϕ there is in fact the discrimination probability function ψ. The canonical space $(\mathfrak{S} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}, \psi)$ is represented by the matrix that we reproduce here for convenience,

ψ	**a**	**b**	**c**	**d**
a	0.3	0.4	0.6	0.7
b	0.4	0.2	0.3	0.5
c	0.5	0.3	0.1	0.2
d	0.8	0.6	0.3	0.1

.

We know that all computations can be performed with either $\Psi^{(1)}$ or $\Psi^{(2)}$, the final result will be the same. Let us therefore compute $\Psi^{(1)}\mathbf{xy}$ by subtracting from each entry $\psi\mathbf{xy}$ the diagonal value in the same row, $\psi\mathbf{xx}$ (because the row labels are representing the stimuli in the first observation area). The result is

$\Psi^{(1)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.3	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	0.7	0.5	0.2	0

. (23)

Let us, e.g., consider next all chains leading from **a** to **d**, and from **d** to **a**. We obviously need not consider chains with loops in them (such as **adcacb**, containing loops **cac** and **adca**).

from **a** to **d**	$\Psi^{(1)}$-ength
ad	0.4
abd	$0.1 + 0.3$
acd	$0.3 + 0.1$
abcd	$0.1 + 0.1 + 0.1$
acbd	$0.3 + 0.2 + 0.3$

,

from **d** to **a**	$\Psi^{(1)}$-ength
da	0.7
dba	$0.5 + 0.2$
dca	$0.2 + 0.4$
dcba	$0.2 + 0.2 + 0.2$
dbca	$0.5 + 0.1 + 0.4$

.

The shortest chains here are **abcd** and either of **dca** and **dcba**, their $\Psi^{(1)}$-lengths being, respectively,

$$G^{(1)}\mathbf{ad} = 0.3, G^{(1)}\mathbf{da} = 0.6.$$

Thence

$$\overleftrightarrow{G}\mathbf{ab} = 0.3 + 0.6 = 0.9.$$

Repeating this procedure for each other pair of stimuli, we obtain the following complete set of $G^{(1)}$-distances,

$G^{(1)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	0.3
b	0.2	0	0.1	0.2
c	0.4	0.2	0	0.1
d	0.6	0.4	0.2	0

, (24)

and, by symmetrization, the complete set of Fechnerian distances,

$\overleftrightarrow{G}$	**a**	**b**	**c**	**d**
a	0	0.3	0.6	0.9
b	0.3	0	0.3	0.6
c	0.6	0.3	0	0.3
d	0.9	0.6	0.3	0

. (25)

The shortest chains are not generally unique, as we have seen in our toy example. However, their infimum for any given pair of points (in the case of finite sets, minimum) is always determined uniquely. (Note that it is only a numerical accident that all $\overset{\leftrightarrow}{G}$ in our example are below 1, there is no general upper bound for $\overset{\leftrightarrow}{G}$ computed from probability values.)

Recall that a label in the canonical stimulus space, say, $\mathbf{a}$, is a representations of two different stimuli in the two observation areas. If one goes back to the original stimulus spaces, the Fechnerian distance 0.6 between points $\mathbf{b}$ and $\mathbf{d}$ in the canonical space $\mathfrak{S}$, is in fact both

(i) the distance between either of the stimuli $\mathbf{x}_2, \mathbf{x}_3$ and either of the stimuli $\mathbf{x}_6, \mathbf{x}_7$ in the stimulus space $\mathfrak{S}_1^*$ (first observation area); and

(ii) the distance between any of the stimuli $\mathbf{y}_4, \mathbf{y}_5, \mathbf{y}_6, \mathbf{y}_7$ and the stimulus $\mathbf{y}_1$ in the stimulus space $\mathfrak{S}_2^*$ (second observation area).

Indeed, any of the stimuli $\mathbf{y}_4, \mathbf{y}_5, \mathbf{y}_6, \mathbf{y}_7$ and either of $\mathbf{x}_2, \mathbf{x}_3$ are each other's PSEs, mapped into $\mathbf{b}$ in the canonical representation. Similarly, either of the stimuli $\mathbf{x}_6, \mathbf{x}_7$ and $\mathbf{y}_1$ are each other's PSEs, mapped into $\mathbf{d}$.

Let us emphasize that Fechnerian distances are always defined *within* observation areas rather than across them. This is the reason Fechnerian distance $\overset{\leftrightarrow}{G}$ is a true metric, with the symmetry property. Within a single observation area the order of two stimuli has no operational meaning, so $\overset{\leftrightarrow}{G}\mathbf{xy}$ cannot be different from $\overset{\leftrightarrow}{G}\mathbf{yx}$. The situation is different when we consider a discrimination probability function ψ or a dissimilarity function D (e.g., $\Psi^{(1)}$ or $\Psi^{(2)}$). In $\psi\mathbf{xy}$ and $D\mathbf{xy}$ the first and second stimuli belong to, respectively, the first and second observation areas, making them meaningfully asymmetric.

The quasimetric dissimilarity G (e.g., $G^{(1)}$or $G^{(2)}$) from which $\overset{\leftrightarrow}{G}$ is computed, strictly speaking, is not interpretable before it is symmetrized. $G\mathbf{xy}$ is merely a component of $\overset{\leftrightarrow}{G}\mathbf{xy}$, the other component being $G\mathbf{yx}$. However, in the rest of this paper we are focusing on G rather than $\overset{\leftrightarrow}{G}$ because the computation of G from D is the nontrivial part of Fechnerian Scaling, leaving one only the trivial step of adding $G\mathbf{yx}$ to $G\mathbf{xy}$.

6.2 Recursive corrections for violations of the triangle inequality

The procedure described in this section is not the only way to compute G from D. Another way, known as the Floyd-Warshall algorithm, is based on the following logic. If one considers in $\mathfrak{S}$ all possible ordered triples $\mathbf{xyz}$ with pairwise distinct elements, and finds out that all of them satisfy the triangle inequality,

$$D\mathbf{xz} \leq D\mathbf{xy} + D\mathbf{yz},$$

then D simply coincides with G. If therefore, in the general case, one could "correct" all ordered triples $\mathbf{xyz}$ for violations of the triangle inequality, one would transform D into G. The following is how this can be done for any finite stimulus space (a generalization to be discussed in Section 9.3).

Let $\mathfrak{S}$ contains k points, and let $\mathfrak{S}_3$ denote the set of $t = k(k-1)(k-2)$ ordered triples of pairwise distinct points of $\mathfrak{S}$. We will call the elements of $\mathfrak{S}_3$ *triangles*. For $n = 0, 1, \ldots$, let $\mathbf{T}^{(n)}$ denote a sequence of the t triangles in $\mathfrak{S}_3$ (in an arbitrary order, as its choice will be shown to be immaterial for the end result). For each n, we index the triangles in $\mathbf{T}^{(n)}$ by double indices $(n,1), (n,2), \ldots, (n,t)$, and we order all such pairs lexicographically: the successor $(n,i)'$ of (n,i) is $(n, i+1)$ if $i < t$ and $(n,t)' = (n+1, 1)$. So the triangle indexed $(n,i)'$ is in $\mathbf{T}^{(n)}$, while the triangle indexed $(n,t)'$ is the first one in $\mathbf{T}^{(n+1)}$.

Definition 11. Given a finite space $(\mathfrak{S}, D)$ and the triangle sequences $\mathbf{T}^{(0)}, \mathbf{T}^{(1)}, \ldots$, the dissimilarity function $M^{(n,i)}$ for $n = 0, 1, \ldots$ and $i = 1, 2, \ldots, t$ is defined by induction as follows.

(i) $M^{(0,i)} \equiv D$ for $i = 1, 2, \ldots, t$.

(ii) Let $M^{(n,i)}$ be defined for some $(n,i) \geq (0,t)$, and let $\mathbf{abc}$ be the triangle indexed by $(n,i)'$. Then $M^{(n,i)'}\mathbf{xy} = M^{(n,i)}\mathbf{xy}$ for all $\mathbf{x}, \mathbf{y} \in \mathfrak{S}$ except, possibly, for $M^{(n,i)'}\mathbf{ac}$, defined as

$$M^{(n,i)'}\mathbf{ac} = \min\left(M^{(n,i)}\mathbf{ac}, M^{(n,i)}\mathbf{ab} + M^{(n,i)}\mathbf{bc}\right).$$

(Note that in every triangle $\mathbf{xyz}$ the triangle inequality is tested only in the form $D\mathbf{xz} \leq D\mathbf{xy} + D\mathbf{yz}$, irrespective of whether any of the remaining five triangles inequalities is violated, $D\mathbf{xy} \leq D\mathbf{xz} + D\mathbf{zy}$, $D\mathbf{zy} \leq D\mathbf{zx} + D\mathbf{xy}$, etc.)

The function $M^{(n,i)}$ for every (n,i) is clearly a dissimilarity function, and it is referred as the *corrected dissimilarity function*. If, at some (n,i), the function $M^{(n,i)}$ is a quasimetric dissimilarity, it is called the *terminal corrected dissimilarity function*.

It follows from Definition 11 that if $(m,j) \geq (n,i)$, then $M^{(m,j)}\mathbf{xy} \leq M^{(n,i)}\mathbf{xy}$ for all $\mathbf{x}, \mathbf{y} \in \mathfrak{S}$. Therefore, if, for some n, $M^{(n+1,t)} \equiv M^{(n,t)}$, then $M^{(n+1,1)} \equiv M^{(n,t)}$, implying that $M^{(n,t)}$ is the terminal dissimilarity function. The converse being obvious, we have

Lemma 12. *$M^{(n,i)}$ is the terminal corrected dissimilarity function if and only if $M^{(n+1,t)} \equiv M^{(n,t)}$.*

The next lemma provides a link between the algorithm being considered and the use of chains in the definition of G. Recall that $\mathcal{C}$ denotes the set of all chains in $\mathfrak{S}$.

Lemma 13. *For any $n = 0, 1, \ldots$, any $i = 1, 2, \ldots, t$, and any $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$, there is a chain $\mathbf{X} \in \mathcal{C}$ such that*

$$M^{(n,i)}\mathbf{ab} = D\mathbf{aXb}.$$

The proof obtains by induction on the lexicographically ordered (n,i). The statement holds for $n = 0$, with $\mathbf{X}$ an empty chain. Let it hold for all double indices up to and including $(n,i) \geq (0,t)$, and let $\mathbf{abc}$ be the triangle indexed $(n,i)'$. Then the statement is clearly true for $M^{(n,i)'}\mathbf{ac}$ whether it equals

$M^{(n,i)}\mathbf{ac}$ or $M^{(n,i)}\mathbf{ab} + M^{(n,i)}\mathbf{bc}$, and it is true for all other $\mathbf{xy}$ because then $M^{(n,i)\prime}\mathbf{xy} = M^{(n,i)}\mathbf{xy}$.

Does a terminal dissimilarity function necessarily exist? Let us assume it does not. Then, by Lemma 12, $M^{(n+1,t)}$ and $M^{(n,t)}$ do not coincide for all $n = 0, 1, \ldots$. Since $\mathfrak{S} \times \mathfrak{S}$ is finite, there should exist distinct points $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$ and an infinite sequence of positive integers $n_1 < n_2 < \ldots$ for which

$$D\mathbf{ab} \neq M^{(n_1,t)}\mathbf{ab} \neq M^{(n_2,t)}\mathbf{ab} \neq \ldots.$$

From Definition 11 it follows then that

$$D\mathbf{ab} > M^{(n_1,t)}\mathbf{ab} > M^{(n_2,t)}\mathbf{ab} > \ldots.$$

By Lemma 13, for every (n_i, t) there should exist a chain $\mathbf{X}_i$ such that

$$M^{(n_i,t)}\mathbf{ab} = D\mathbf{a}\mathbf{X}_i\mathbf{b},\ i = 1, 2, \ldots.$$

But a sequence of inequalities

$$D\mathbf{ab} > D\mathbf{a}\mathbf{X}_{n_1}\mathbf{b} > D\mathbf{a}\mathbf{X}_{n_2}\mathbf{b} > \ldots$$

is impossible in a finite set, because the set of chains with lengths below a given value is finite. This contradiction proves the existence of a terminal dissimilarity function. Let us denote it by M. Observe that for any $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$ and any chain $\mathbf{X} \in \mathcal{C}$,

$$D\mathbf{aXb} \geq M\mathbf{aXb}.$$

But M satisfies the triangle inequality, whence

$$M\mathbf{aXb} \geq M\mathbf{ab},$$

whence

$$M\mathbf{ab} \leq D\mathbf{aXb}.$$

By Lemma 13, this implies

$$M\mathbf{ab} = \min_{\mathbf{X} \in \mathcal{C}} D\mathbf{aXb},$$

which equals $G\mathbf{ab}$ by definition. We have established therefore

Theorem 14. *A terminal corrected dissimilarity function exists, and it coincides with the quasimetric dissimilarity G induced by the initial dissimilarity function D.*

It is worthwhile to emphasize that nowhere in the proof we have used a specific order of the triangles in $\mathbf{T}^{(n)}$.

We see that dissimilarities on finite sets can be viewed as "imperfect" quasimetric dissimilarities, and the dissimilarity cumulation procedure can be recast as a series of recursive corrections of the dissimilarities for the violations of the triangle inequality.

Let us illustrate the procedure on our toy example, starting with the matrix of dissimilarities

$\Psi^{(1)} = D$	**a**	**b**	**c**	**d**
a	0	0.1	0.3	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	0.7	0.5	0.2	0

and using, for each $\mathbf{T}^{(n)}$ the same sequence of $t = 24$ triangles

$$\begin{array}{cccccc} i=1 & 2 & 3 & \dots & 23 & 24 \\ \mathbf{acb} & \mathbf{abc} & \mathbf{adc} & \dots & \mathbf{dac} & \mathbf{dbc} \end{array}. \tag{26}$$

It is obtained by cycling through the first element (4 values), subcycling through the last element (3 values), and sub-subcycling through the middle element (2 values), in the alphabetic order.

Testing the triangles in $\mathbf{T}^{(1)}$ one by one, $M^{(1,1)}$coincides with D because the triangle indexed $(1,1)$ is **acb**, and the triangle inequality in it is not violated. Similarly, $M^{(1,2)} \equiv M^{(1,1)}$ and $M^{(1,3)} \equiv M^{(1,2)}$ because the triangle inequality is not violated in the triangles labeled $(1,2)$ and $(1,3)$. The first violation of the triangle inequality occurs in the triangle indexed $(1,3)$, **abc**:

$$0.3 = D\mathbf{ac} > D\mathbf{ab} + D\mathbf{bc} = 0.1 + 0.1.$$

We "correct" the value of $D\mathbf{ac}$ therefore by replacing 0.3 with 0.2 (shown in parentheses in matrix $M^{(1,3)}$ below):

D	**a**	**b**	**c**	**d**
a	0	0.1	0.3	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	0.7	0.5	0.2	0

$\Rightarrow$

$M^{(1,3)}$	**a**	**b**	**c**	**d**
a	0	0.1	(0.2)	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0.0	0.1
d	0.7	0.5	0.2	0.0

No violations occur until we reach the triangle indexed $(1,20)$, so $M^{(1,19)} \equiv M^{(1,18)} \equiv \ldots \equiv M^{(1,3)}$. In $M^{(1,19)}$,however, we have, for the triangle **dca**:

$$0.7 = M^{(1,19)}\mathbf{da} > M^{(1,19)}\mathbf{dc} + M^{(1,19)}\mathbf{ca} = 0.2 + 0.4,$$

We correct $M^{(1,19)}\mathbf{da}$ from 0.7 to 0.6, as shown in the parentheses in matrix $M^{(1,20)}$.

$M^{(1,19)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	0.7	0.5	0.2	0

$\Rightarrow$

$M^{(1,20)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	(0.6)	0.5	0.2	0

.

We deal analogously with the third violation of the triangle inequality, in the triangle **dcb**, indexed $(1, 22)$:

$$0.5 = M^{(1,21)}\mathbf{db} > M^{(1,21)}\mathbf{dc} + M^{(1,21)}\mathbf{cb} = 0.2 + 0.2.$$

So $M^{(1,21)} \equiv M^{(1,20)} \equiv M^{(1,19)}$, and

$M^{(1,21)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	0.4
b	0.2	0	0.1	0.3
c	0.4	20.	0	0.1
d	0.6	0.5	0.2	0

$\Rightarrow$

$M^{(1,22)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	0.4
b	0.2	0	0.1	0.3
c	0.4	20.	0	0.1
d	0.6	(0.4)	0.2	0

.

With the remaining two triangles before the sequence $\mathbf{T}^{(1)}$ has been exhausted no violations occur, so $M^{(1,24)} \equiv M^{(1,23)} \equiv M^{(1,22)}$ is the matrix with which the second sequence, $\mathbf{T}^{(2)}$, begins. The first and only violation here occurs at the triangle indexed $(2, 5)$, **abd**:

$$0.4 = M^{(2,4)}\mathbf{ad} > M^{(2,4)}\mathbf{ab} + M^{(2,4)}\mathbf{bd} = 0.1 + 0.2,$$

So $M^{(2,4)} \equiv \ldots \equiv M^{(2,1)} \equiv M^{(1,24)}$, and

$M^{(1,24)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	0.4
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	0.6	0.4)	0.2	0

$\Rightarrow$

$M^{(2,5)}$	**a**	**b**	**c**	**d**
a	0	0.1	0.2	(0.3)
b	0.2	0	0.1	0.3
c	0.4	0.2	0	0.1
d	0.6	0.4	0.2	0

.

One can verify that $M^{(2,5)}$ is a quasimetric dissimilarity on $\mathfrak{S} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}$, so that $M^{(2,6)}$ and all higher-indexed matrices remain equal to $M^{(2,5)}$. The latter therefore is the terminal corrected dissimilarity, and its comparison with (24) shows that it coincides with $G = G^{(1)}$, the quasimetric induced by the initial dissimilarity function $D = \Psi^{(1)}$.

7 Dissimilarity cumulation in path-connected spaces

7.1 Chains-on-nets and paths

We now turn to dissimilarity cumulation in stimulus spaces $(\mathfrak{S}, D)$ in which points can be connected by *paths*. A path is a continuous function $\mathbf{f} : [a, b] \rightarrow \mathfrak{S}$. Because $[a, b]$ is a closed interval of reals, this function is also uniformly continuous. The latter means that $\mathbf{f}(x) \leftrightarrow \mathbf{f}(y)$ if $x - y \rightarrow 0$ $(x, y \in [a, b]$. We will present this path more compactly as $\mathbf{f}|\,[a, b]$, and say that it connects $\mathbf{f}(a) = \mathbf{a}$ to $\mathbf{f}(b) = \mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are allowed to coincide.

To introduce the notion of the length of the path $\mathbf{f}|\,[a, b]$, we need the following auxiliary notions. A *net* on $[a, b]$ is defined as a sequence of numbers

$$\mu = (a = x_0 \leq x_1 \leq \ldots \leq x_k \leq x_{k+1} = b),$$

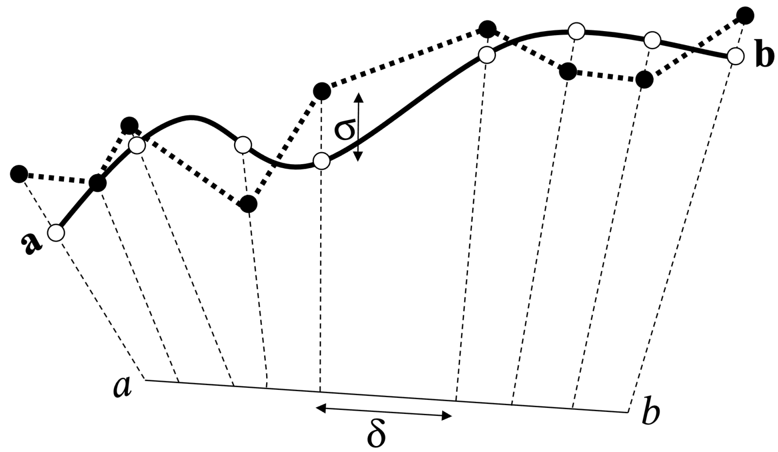

Figure 11: Chains-on-nets $\mathbf{X}^{\mu}$ are converging to a path $\mathbf{f}|[a,b]$ as $\delta = \delta\mu \to 0$ and $\sigma = \sigma(\mathbf{f}, \mathbf{X}^{\mu}) \to 0$. The length $D\mathbf{X}$ of the chains has then the limit inferior that is taken for the length of the path $\mathbf{f}$.

not necessarily pairwise distinct. The quantity

$$\delta\mu = \max_{i=0,1\ldots,k} (x_{i+1} - x_i)$$

is called the net's *mesh*. A net $\mu = (a, x_1, \ldots, x_k, b)$ can be elementwise paired with a chain $\mathbf{X} = \mathbf{x}_0\mathbf{x}_1 \ldots \mathbf{x}_k\mathbf{x}_{k+1}$ to form a *chain-on-net*

$$\mathbf{X}^{\mu} = ((a, \mathbf{x_0}), (x_1, \mathbf{x}_1), \ldots, (x_k, \mathbf{x}_k), (b, \mathbf{x}_{k+1})).$$

Note that the elements of the chain $\mathbf{X}$ need not be pairwise distinct. The *separation* of the chain-on-net $\mathbf{X}^{\mu}$ from the path $\mathbf{f}|[a,b]$ is defined as

$$\sigma(\mathbf{f}, \mathbf{X}^{\mu}) = \max_{x_i \in \mu} D\mathbf{f}(x_i)\,\mathbf{x}_i.$$

Definition 15. The *D-length* of path $\mathbf{f}|[a,b]$ is defined as

$$D\mathbf{f} = \liminf_{\delta\mu \to 0, \sigma(\mathbf{f}, \mathbf{X}^{\mu}) \to 0} D\mathbf{X}.$$

The limit inferior stands here for

$$\sup_{\varepsilon_1 > 0, \varepsilon_2 > 0} \inf \{D\mathbf{X} : \delta\mu < \varepsilon_1, \sigma(\mathbf{f}, \mathbf{X}^{\mu}) < \varepsilon_2\}.$$

Let us agree to say that $\mathbf{X}^{\mu}$ converges *to* $\mathbf{f}$ (and write $\mathbf{X}^{\mu} \to \mathbf{f}$) if $\delta\mu \to 0$ and $\sigma(\mathbf{f}, \mathbf{X}^{\mu}) \to 0$. We can then rewrite the definition above as

$$D\mathbf{f} = \liminf_{\mathbf{X}^{\mu} \to \mathbf{f}} D\mathbf{X}. \tag{27}$$

Using the properties of lim inf, for any path $\mathbf{f}$, there exists a sequence $\{\mathbf{X}_n^{\mu_n}\}$ of chains-on-nets such that $\delta\mu_n \to 0$ and $\sigma(\mathbf{f}, \mathbf{X}_n^{\mu_n}) \to 0$, and $D\mathbf{X}_n \to D\mathbf{f}$.

Let us list some of the most basic properties of the D-length of a path.

Theorem 16. *The length $D\mathbf{f}$ of any path $\mathbf{f}|[a,b]$ has the following properties:*
$\mathcal{L}1$ *(nonnegativity) $D\mathbf{f} \geq 0$;*
$\mathcal{L}2$ *(zero property) $D\mathbf{f} = 0$ if and only if $\mathbf{f}([a,b])$ is a single point;*
$\mathcal{L}3$ *(additivity) for any $c \in [a,b]$, $D\mathbf{f}|[a,b] = D\mathbf{f}|[a,c] + D\mathbf{f}|[c,b]$.*

Proofs of these statements are simple. Thus, to show the additivity of $D\mathbf{f}$, add the point c twice to all nets,

$$\widetilde{\mu} = \left\{ \overbrace{a = x_0 \leq \ldots x_i \leq c}^{\alpha} = \underbrace{c \leq x_{i+1} \leq \ldots \leq x_{k+1} = b}_{\beta} \right\},$$

and two corresponding point $\mathbf{c^1}, \mathbf{c}^2$ to all chains,

$$\widetilde{\mathbf{X}} = \overbrace{\mathbf{x}_0 \ldots \mathbf{x}_i \mathbf{c}^1}^{\mathbf{Y}} \underbrace{\mathbf{c}^2 \mathbf{x}_{i+1} \ldots \mathbf{x}_{k+1}}_{\mathbf{Z}}.$$

Clearly,

$$\liminf_{\widetilde{\mathbf{X}}^{\widetilde{\mu}} \to \mathbf{f}|[a,b]} \widetilde{\mathbf{X}} = \liminf_{\mathbf{Y}^{\alpha} \to \mathbf{f}|[a,c]} \mathbf{Y} + \liminf_{\mathbf{Z}^{\beta} \to \mathbf{f}|[c,b]} \mathbf{Z} = D\mathbf{f}|[a,c] + D\mathbf{f}|[c,b].$$

For any sequence $\{\mathbf{X}_n^{\mu_n}\}$ of chains-on-nets such that $\mathbf{X}_n^{\mu_n} \to \mathbf{f}|[a,c]$, and $D\mathbf{X}_n \to D\mathbf{f}$, we have $\widetilde{\mathbf{X}}_n^{\widetilde{\mu}_n} \to \mathbf{f}|[a,c]$ for the corresponding sequence $\left\{\widetilde{\mathbf{X}}_n^{\widetilde{\mu}_n}\right\}$, assuming $\mathbf{c}_n^1 \to \mathbf{f}(c)$ and $\mathbf{c}_n^2 \to \mathbf{f}(c)$. We also have

$$D\widetilde{\mathbf{X}}_n = D\mathbf{X}_n + \left(D\mathbf{x}_{i_n}\mathbf{c}_n^1 + D\mathbf{c}_n^1\mathbf{c}_n^2 + D\mathbf{c}_n^2\mathbf{x}_{i_n+1} - D\mathbf{x}_{i_n}\mathbf{x}_{i_n+1}\right),$$

where each summand in the parentheses tends to zero by the uniform continuity of $\mathbf{f}$ and D.

Note that $D\mathbf{f}$ is well-defined for any path $\mathbf{f}$, but only on the extended set of nonnegative reals: the value of $D\mathbf{f}$ may very well be equal to ∞. This does not invalidate or complicate any of the results presented in this chapter, but, for brevity sake, we will tacitly assume that $D\mathbf{f}$ is finite.

The reader may wonder why, in the definition of $D\mathbf{f}$, it is not sufficient to deal with the inscribed chains-on-nets, with all elements of the chains belonging to the path $\mathbf{f}$. We will see later that this is indeed sufficient if D is a quasimetric dissimilarity. However, in general, the inscribed chains-on-nets do not reach the infimum of the D-lengths of the "meandering" chains-on-nets. Figure 12 provides an illustration. In this example, the stimuli are points in $\mathbb{R}^2$, and, for $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$,

$$D\mathbf{ab} = |a_1 - b_1| + |a_2 - b_2| + \min(|a_1 - b_1|, |a_2 - b_2|).$$

It is easy to check that D is a dissimilarity function. Thus, $\mathcal{D}3$ follows from the fact

$$D\mathbf{a}_n\mathbf{b}_n \to 0 \Longleftrightarrow |\mathbf{a}-\mathbf{b}| \to 0,$$

where $|\mathbf{a}-\mathbf{b}|$ is the usual Euclidean norm. Also, for any chain $\mathbf{aXb}$,

$$D\mathbf{aXb} \geq |a_1 - b_1| + |a_2 - b_2|,$$

whence $D\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n \to 0$ implies $D\mathbf{a}_n\mathbf{b}_n \to 0$. That is, D satisfies $\mathcal{D}4$. By the same inequality, the length of the line segment $\mathbf{f}$ shown in Figure 12, connecting $\mathbf{a} = (1,0)$ to $\mathbf{b} = (0,1)$, cannot be less than 2. (The domain interval for $\mathbf{f}$ can be chosen arbitrarily, e.g., $[0,1]$) Consider now chains-on-nets $\mathbf{X}^{\mu}$ with the staircase chains, as in the left panel. By decreasing the mesh of μ and the spacing of the elements of $\mathbf{X}$, it can be made to converge to $\mathbf{f}$, and since $D\mathbf{X}$ for all these chains equals 2, $D\mathbf{f} = 2$. At the same time, the inscribed chains, as in the right panel of the figure, are easily checked to have the length 3.

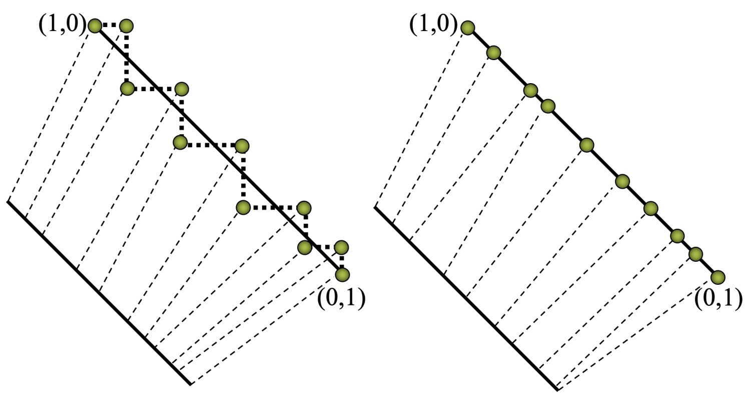

Figure 12: A demonstration of why for D-length computations we need the "meandering" chains like in Figure 11 rather than just inscribed chains. Here, $D\mathbf{ab}$ for $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ is defined as $|a_1 - b_1| + |a_2 - b_2| + \min(|a_1 - b_1|, |a_2 - b_2|)$. All staircase chains $\mathbf{X}$, irrespective of the spacing of their elements, have the cumulated dissimilarity $D\mathbf{X} = 2$, and 2 is the true D-length of the path between $(1,0)$ and (0,1). All inscribed chains, irrespective of the spacing of their elements, have the cumulated dissimilarity 3. Explanations are given in the text.

7.2 Path length through quasimetric dissimilarity

Different dissimilarity functions D lead to different quantifications of path length. We know that the quasimetric dissimilarity G defined by (18) is a dissimilarity function. However, in this case, since G is defined through D by (18), one should expect, for consistency, that the the path-length will remain unchanged on replacing D and with G. This will indeed be established in Section 7.3. We need several preliminary results first, however.

Using G in place of D to define the G-length of paths, we have

$$G\mathbf{f} = \liminf_{\mathbf{X}^\mu \overset{G}{\to} \mathbf{f}} G\mathbf{X}.$$

The condition $\mathbf{X}^\mu \overset{G}{\to} \mathbf{f}$ here means $\delta\mu \to 0$ and

$$\sigma_G\left(\mathbf{f}, \mathbf{X}^\mu\right) = \max_{x_i \in \mu} G\mathbf{f}\left(x_i\right)\mathbf{x}_i \to 0.$$

But by Theorem 6, the latter condition is equivalent to

$$\sigma\left(\mathbf{f}, \mathbf{X}^\mu\right) = \max_{x_i \in \mu} D\mathbf{f}\left(x_i\right)\mathbf{x}_i \to 0.$$

Therefore $\mathbf{X}^\mu \overset{G}{\to} \mathbf{f}$ and $\mathbf{X}^\mu \to \mathbf{f}$ are equivalent, and we can formulate

Definition 17. The *G-length* of path $\mathbf{f}|\left[a,b\right]$ is

$$G\mathbf{f} = \liminf_{\mathbf{X}^\mu \to \mathbf{f}} G\mathbf{X}.$$

Consider now chains-on-nets $\mathbf{Z}^\nu$ *inscribed* in $\mathbf{f}|\left[a,b\right]$, that is, those with

$$\nu = \left\{a = z_0, z_1, \ldots, z_k, z_{k+1} = b\right\}$$

and

$$\mathbf{Z} = \mathbf{f}\left(z_0\right)\ldots\mathbf{f}\left(z_{k+1}\right) = \mathbf{z}_0\ldots\mathbf{z}_{k+1}.$$

Since $\sigma\left(\mathbf{f}, \mathbf{Z}^\nu\right) = 0$, the condition $\mathbf{Z}^\nu \to \mathbf{f}$ here reduces to $\delta\nu \to 0$. Clearly,

$$\liminf_{\delta\nu \to 0} G\mathbf{Z} \geq \liminf_{\mathbf{X}^\mu \to \mathbf{f}} G\mathbf{X} = G\mathbf{f}, \tag{28}$$

because inscribed chains-on-nets converging to $\mathbf{f}$ form a subset of all chains-on-nets converging to $\mathbf{f}$. We will see now that in fact the two quantities in (28) are equal. By the additivity property,

$$G\mathbf{f}|\left[a,b\right] = \sum_{i=0}^{k} G\mathbf{f}|\left[z_i, z_{i+1}\right].$$

Let $\mathbf{X}_i^{\mu_i}$ be an arbitrary chain-in-net with $\mu_i \subset \left[z_i, z_{i+1}\right]$. By the same reasoning as in the proof of the additivity property, if μ_i is changed into

$$\widetilde{\mu}_i = \left\{z_i, \overbrace{\ldots}^{\mu_i}, z_{i+1}\right\},$$

and $\mathbf{X}_i$ into

$$\widetilde{\mathbf{X}}_i = \mathbf{z}_i\mathbf{X}_i\mathbf{z}_{i+1},$$

the conditions $\mathbf{X}_i^{\mu_i} \to \mathbf{f}|\,[z_i, z_{i+1}]$ and $\widetilde{\mathbf{X}}_i^{\widetilde{\mu}_i} \to \mathbf{f}|\,[z_i, z_{i+1}]$ are equivalent. Denoting by $\mathbf{X}^\mu$ the concatenation of $\mathbf{X}_i^{\mu_i}$ for $i = 0, \ldots, k$, and defining $\widetilde{\mathbf{X}}^{\widetilde{\mu}}$ analogously, we have

$$G\mathbf{f} = \liminf_{\mathbf{X}^\mu \to \mathbf{f}} G\mathbf{X} = \liminf_{\widetilde{\mathbf{X}}^{\widetilde{\mu}} \to \mathbf{f}} G\widetilde{\mathbf{X}}.$$

At the same time, by the triangle inequality,

$$G\mathbf{z}_i\mathbf{z}_{i+1} \leq G\mathbf{z}_i\mathbf{X}_i\mathbf{z}_{i+1},$$

whence

$$G\mathbf{Z} \leq G\widetilde{\mathbf{X}}$$

and

$$\liminf_{\delta\nu \to 0} G\mathbf{Z} \geq \liminf_{\widetilde{\mathbf{X}}^{\widetilde{\mu}} \to \mathbf{f}} G\widetilde{\mathbf{X}} = G\mathbf{f}. \tag{29}$$

Together with (28), this establishes

Theorem 18. *For any path* $\mathbf{f}$*,*

$$G\mathbf{f} = \liminf_{\delta\nu \to 0} G\mathbf{Z},$$

where $\mathbf{Z}^\nu$ *are chains-on-nets inscribed in* $\mathbf{f}$*.*

In other words, to approximate $G\mathbf{f}$ by G-lengths of chains-on-nets, one does not need all possible chains converging to $\mathbf{f}$, the inscribed ones only are sufficient. Recall that the analogous statement is not correct for $D\mathbf{f}$. The equality in Theorem 18 critically owes to the fact that G satisfies the triangle inequality.

We can further clarify Theorem 18 as follows.

Theorem 19. *For any path* $\mathbf{f}$*,*

$$G\mathbf{f} = \sup G\mathbf{Z} = \lim_{\delta\nu \to 0} G\mathbf{Z}, \tag{30}$$

where $\mathbf{Z}^\nu$ *are chains-on-nets inscribed in* $\mathbf{f}$*.*

In other words, $G\mathbf{f}$ is the lowest upper bound for the lengths of all inscribed chains-on-nets; and any sequence of the inscribed chains-on-nets converges to $G\mathbf{f}$ as their mesh decreases.

To prove the first equality, $G\mathbf{f} = \sup G\mathbf{Z}$, consider a chain-on-net $\mathbf{Z}^\nu$ with $\sup G\mathbf{Z} - G\mathbf{Z}$ arbitrarily small. For every pair of successive z_i, z_{i+1} in ν, one can find an inscribed chain-on-net $\mathbf{V}_i^{\mu_i}$ such that $\mu_i = \{z_{i_n}, \ldots, z_{i_n+1}\}$ and $|G\mathbf{V}_i - G\mathbf{f}|\,[z_i, z_{i+1}]|$ is arbitrarily small. By the additivity of G-length, denoting by $\mathbf{V}^\mu$ the concatenation of all $\mathbf{V}_i^{\mu_i}$, we can make $|G\mathbf{V} - G\mathbf{f}|\,[a, b]|$ arbitrarily small. From the triangle inequality it follows that $G\mathbf{V} \geq G\mathbf{Z}$, whence $G\mathbf{f} \geq \sup G\mathbf{Z}$. But $G\mathbf{V} \leq \sup G\mathbf{Z}$, whence we also have $G\mathbf{f} \leq \sup G\mathbf{Z}$.

To prove that $G\mathbf{f} = \lim_{\delta\nu \to 0} G\mathbf{Z}$, deny it, and assume that there is a sequence of inscribed chains-on-nets $\mathbf{V}_n^{\mu_n}$ such that $\delta\mu_n \to 0$ but $G\mathbf{V}_n \not\to D\mathbf{f}$. Since $D\mathbf{f} = \sup D\mathbf{Z}$ across all possible inscribed chains-on-nets, $D\mathbf{V}_n \leq D\mathbf{f}$ for all n.

Then one can find a $\Delta > 0$ and a subsequence of $\mathbf{V}_n^{\mu_n}$ (which, with no loss of generality, we can assume to be $\mathbf{V}_n^{\mu_n}$ itself) such that

$$D\mathbf{V}_n \to D\mathbf{f} - \Delta.$$

Let $\mathbf{Z}^\nu$ be an inscribed chain-on-net with

$$D\mathbf{Z} > D\mathbf{f} - \Delta/2.$$

For every z_i in ν and every n, let $v_{k_{i,n}}^n, v_{k_{i,n}+1}^n$ be two successive elements of μ_n such that $v_{k_{i,n}}^n \le z_i \le v_{k_{i,n}+1}^n$. For a sufficiently large n, $\delta\mu_n$ is sufficiently small to ensure that z_i is the only member of ν falling between $v_{k_{i,n}}^n$ and $v_{k_{i,n}+1}^n$ (without loss of generality, we can assume that ν contains no identical elements). Denote by $\nu \uplus \mu_n$ the nets formed by the elements of ν inserted into μ_n. Consider the inscribed chains-on-nets $\mathbf{U}^{\nu \uplus \mu_n}$. We have (denoting by l the cardinality of ν),

$$\begin{aligned} G\mathbf{U} &= G\mathbf{V}_n \\ &+ \sum_{i=0}^{l} \left\{ G\mathbf{f}\left(v_{k_{i,n}}^n\right)\mathbf{f}\left(z_i\right) + G\mathbf{f}\left(z_i\right)\mathbf{f}\left(v_{k_{i,n}+1}^n\right) - G\mathbf{f}\left(v_{k_{i,n}}^n\right)\mathbf{f}\left(v_{k_{i,n}+1}^n\right)\right\}. \end{aligned}$$

By the uniform continuity of $\mathbf{f}$, the expression under the summation operator tends to zero, whence

$$G\mathbf{U} - G\mathbf{V}_n \to 0,$$

and then

$$G\mathbf{U} \to G\mathbf{f} - \Delta.$$

But by the triangle inequality, for all n,

$$D\mathbf{U} \ge D\mathbf{Z} > D\mathbf{f} - \Delta/2.$$

This contradiction completes the proof.

7.3 The equality of the D-length and G-length of paths

As mentioned previously, one can expect that path length should not depend on whether one chooses dissimilarity D or the quasimetric dissimilarity G induced by D.

Theorem 20. *For any path* $\mathbf{f}$*,*

$$D\mathbf{f} = G\mathbf{f}.$$

Comparing Definitions 15 and 17, since $D\mathbf{X} \ge G\mathbf{X}$ for any chain, we have $D\mathbf{f} \ge G\mathbf{f}$. To see that $D\mathbf{f} \le G\mathbf{f}$, we form a sequence of inscribed chains-on-nets $\mathbf{Z}_n^{\nu_n}$ such that $\delta\nu_n \to 0$, and $G\mathbf{Z}_n \to G\mathbf{f}$. By the definition of G, one can insert chains $\mathbf{X}_i^n$ between pairs of successive elements $\mathbf{z}_i^n, \mathbf{z}_{i+1}^n$ of $\mathbf{Z}_n$, so that

$$D\mathbf{U}_n - G\mathbf{Z}_n \le \frac{1}{n},$$

where

$$\mathbf{U}_n = \mathbf{z}_0^n \mathbf{X}_0^n \mathbf{z}_1^n \ldots \mathbf{z}_{k_n}^n \mathbf{X}_{k_n}^n \mathbf{z}_{k_n+1}^n.$$

In other words, $D\mathbf{U}_n \to G\mathbf{f}$. Let us now create a net μ_n for every $\mathbf{U}_n$ as follows: if $z_i^n \in \nu_n$ is associated with $\mathbf{z}_i^n \in \mathbf{Z}_n$, we associate z_i^n with every element of $\mathbf{X}_i^n$. The resulting chain-on-net is

$$\mathbf{U}_n^{\mu_n} = \left(\ldots, (z_i^n, \mathbf{z}_i^n), \left(z_i^n, \mathbf{x}_1^{i,n}\right), \ldots, \left(z_i^n, \mathbf{x}_{l_{i,n}}^{i,n}\right)(z_{i+1}^n, \mathbf{z}_{i+1}^n), \ldots\right).$$

We will show now that $\mathbf{U}_n^{\mu_n} \to \mathbf{f}$. Since $\delta\mu_n = \delta\nu_n \to 0$, we have to show that $\sigma(\mathbf{f}, \mathbf{U}_n^{\mu_n}) \to 0$. Let $\left(z_{i_n}^n, \mathbf{m}_{i_n}^n\right)$ be an element of $\mathbf{U}_n^{\mu_n}$ such that

$$\sigma(\mathbf{f}, \mathbf{U}_n^{\mu_n}) = D\mathbf{f}\left(z_{i_n}^n\right)\mathbf{m}_{i_n}^n = D\mathbf{z}_{i_n}^n \mathbf{m}_{i_n}^n.$$

By the uniform continuity of $\mathbf{f}$ and G,

$$G\mathbf{z}_{i_n}^n \mathbf{z}_{i_n+1}^n = G\mathbf{f}\left(z_{i_n}^n\right)\mathbf{f}\left(z_{i_n+1}^n\right) \to 0$$

as $\delta\mu_n = \delta\nu_n \to 0$. By the construction of $\mathbf{U}_n$,

$$D\mathbf{z}_{i_n}^n = D\mathbf{z}_{i_n}^n \overbrace{\mathbf{x}_1^{i_n,n} \ldots \mathbf{m}_{i_n}^n \ldots \mathbf{x}_{l_{i_n,n}}^{i_n,n}}^{\mathbf{X}_{i_n}^n} \mathbf{z}_{i_n+1}^n \to 0,$$

implying

$$D\mathbf{z}_{i_n}^n \mathbf{x}_1^{i_n,n} \ldots \mathbf{m}_{i_n}^n \to 0.$$

By the chain property of dissimilarity functions,

$$\sigma(\mathbf{f}, \mathbf{U}_n^{\mu_n}) = D\mathbf{z}_{i_n}^n \mathbf{m}_{i_n}^n \to 0.$$

We have therefore a sequence of chains-on-nets $\mathbf{U}_n^{\mu_n} \to \mathbf{f}$ with $G\mathbf{f}$ as the limit point of $D\mathbf{U}_n$, and then $G\mathbf{f} \geq D\mathbf{f}$ because $D\mathbf{f}$ is the infimum of all such limit points. This completes the proof.

We see that although $D\mathbf{xy}$ and $G\mathbf{xy}$ are generally distinct for points $\mathbf{x}, \mathbf{y}$, when it comes to paths $\mathbf{f}$, the quantities $D\mathbf{f}$ and $G\mathbf{f}$ can be used interchangeably. One consequence of this result is that the properties of the D-length of paths can now be established by replacing it with the G-length, the advantage of this being that we acquire the powerful triangle inequality to use, and also restrict chains-on-nets to the inscribed ones, more familiar than the "meandering" chains in Figure 11. However, the general definition of $D\mathbf{f}$ remains convenient in many situations. We illustrate this on the important property of *lower semicontinuity* of the D-length.

Definition 21. A sequence of paths $\mathbf{f}_n | [a, b]$ *converges* to a path $\mathbf{f} | [a, b]$ (in symbols, $\mathbf{f}_n \to \mathbf{f}$) if

$$\sigma(\mathbf{f}, \mathbf{f}_n) = \max_{x \in [a,b]} D\mathbf{f}(x)\mathbf{f}_n(x) \to 0.$$

Consider any sequence of chains-on-nets $\mathbf{X}_n^{\mu_n} \to \mathbf{f}_n$ such that $|D\mathbf{X}_n - D\mathbf{f}_n| \to 0$. By the uniform continuity of D,

$$[\sigma(\mathbf{f}_n, \mathbf{X}_n^{\mu_n}) \to 0] \text{ and } [\sigma(\mathbf{f}, \mathbf{f}_n) \to 0] \Longrightarrow \sigma(\mathbf{f}, \mathbf{X}_n^{\mu_n}) \to 0.$$

Then $\mathbf{X}_n^{\mu_n} \to \mathbf{f}$, whence $\liminf_{n\to\infty} D\mathbf{X}_n \geq D\mathbf{f}$. But $\liminf_{n\to\infty} D\mathbf{X}_n = \liminf_{n\to\infty} D\mathbf{f}_n$. This proves

Theorem 22 (Lower semicontinuity)**.** *For any sequence of paths $\mathbf{f}_n|[a,b] \to \mathbf{f}|[a,b]$,*

$$\liminf_{n\to\infty} D\mathbf{f}_n \geq D\mathbf{f}.$$

7.4 Intrinsic metrics and spaces with intermediate points

In a path-connected space, a metric is traditionally called *intrinsic* if the distance between two points is the greatest lower bound for the length of all paths connecting the two points. For instance, in $\mathbb{R}^n$ endowed with the Euclidean geometry, the Euclidean distance

$$D\mathbf{ab} = |\mathbf{a} - \mathbf{b}|$$

between points $\mathbf{a}$ and $\mathbf{b}$ is intrinsic, because it is also the length of the shortest path connecting these points, a straight line segment. By contrast,

$$D\mathbf{ab} = \sqrt{|\mathbf{a} - \mathbf{b}|}$$

is also a metric, but it is not intrinsic: the path length $D\mathbf{f}$ induced by this metric is infinitely large for every path $\mathbf{f}$. As an example of a non-intrinsic metric with a finite path length function, consider

$$D\mathbf{ab} = \tan|a - b|$$

on the interval $\left[0, \frac{\pi}{2}\right[$, where a, b are the values of $\mathbf{a}, \mathbf{b}$, respectively. The length of the (only) path connecting $\mathbf{a}$ to $\mathbf{b}$ here is $|a - b| \neq \tan|a - b|$.

In this section we consider a generalization of the notion of intrinsic metric to quasimetric dissimilarities.

Definition 23. The quasimetric dissimilarity G defined in a space $(\mathfrak{S}, D)$ by (18) is called intrinsic if, for any $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$,

$$G\mathbf{ab} = \inf_{\mathbf{f}\in\mathcal{P}_{\mathbf{a}}^{\mathbf{b}}} D\mathbf{f},$$

where $\mathcal{P}_{\mathbf{a}}^{\mathbf{b}}$ is the class of all *paths* connecting $\mathbf{a}$ to $\mathbf{b}$.

Figure 13 provides an illustration.

We know that in Definition 23 $D\mathbf{f}$ can be replaced with $G\mathbf{f}$. We also know that $G\mathbf{f}$ for any $\mathbf{f} \in \mathcal{P}_{\mathbf{a}}^{\mathbf{b}}$ can be arbitrarily closely approximated by $G\mathbf{aXb}$ for

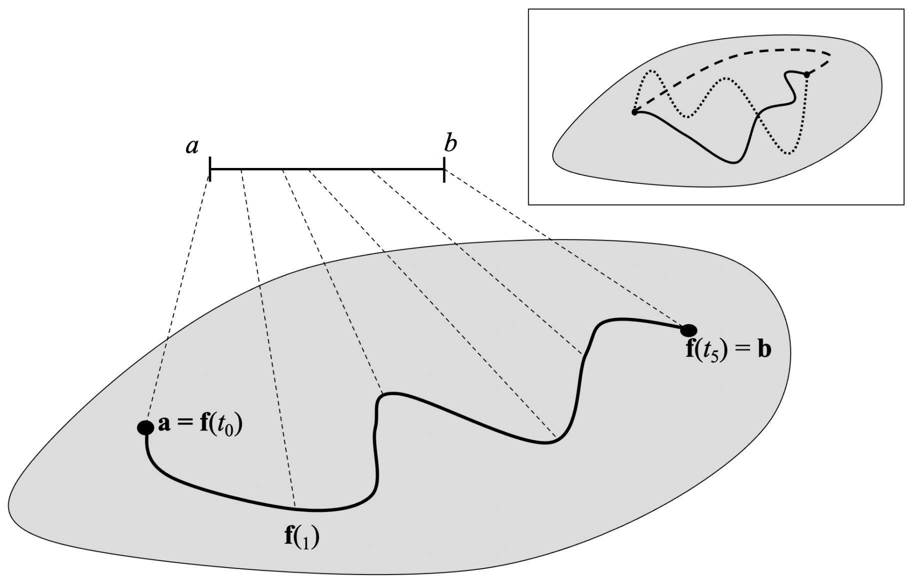

Figure 13: The metric G induced by dissimilarity D is intrinsic if the G-distance from **a** to **b** equal the infimum of D-lengths (equivalently, G-lengths) of all paths connecting **a** to **b**.

some inscribed chain-on-net $\mathbf{X}^\mu$. By the triangle inequality, $G\mathbf{ab} \leq G\mathbf{aXb}$. Therefore, in any space $(\mathfrak{S}, D)$,

$$G\mathbf{ab} \leq \inf_{\mathbf{f} \in \mathcal{P}_{\mathbf{a}}^{\mathbf{b}}} D\mathbf{f}. \tag{31}$$

We need now to consider a special class of spaces in which this inequality can be reversed.

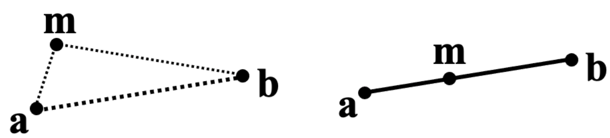

Figure 14: If $D\mathbf{amb} \leq D\mathbf{ab}$, the point $\mathbf{m}$ is said to be intermediate to $\mathbf{a}$ and $\mathbf{b}$. As a special case, if D is Euclidean distance (right picture), any $\mathbf{m}$ on the straight line segment connecting $\mathbf{a}$ and $\mathbf{b}$ is intermediate to $\mathbf{a}$ and $\mathbf{b}$.

Definition 24. A stimulus space $(\mathfrak{S}, D)$ is said to be a space *with intermediate points* if, for any distinct $\mathbf{a}, \mathbf{b}$, one can find an $\mathbf{m}$ such that $\mathbf{m} \notin \{\mathbf{a}, \mathbf{b}\}$ and $D\mathbf{amb} \leq D\mathbf{ab}$.

Fig. 14 provides an illustration. If D is a metric (or quasimetric dissimilarity), the inequality $D\mathbf{amb} \leq D\mathbf{ab}$ can only have the form

$$D\mathbf{amb} = D\mathbf{ab}.$$

In this form the notion is know as *Menger convexity*.

A sequence $\mathbf{x_1}, \mathbf{x}_2, \ldots$ in $(\mathfrak{S}, D)$ is called a *Cauchy sequence* if

$$\lim_{\substack{k \to \infty \\ l \to \infty}} D\mathbf{x}_k\mathbf{x}_l = 0,$$

that is, if for any $\varepsilon > 0$ one can find an n such that $D\mathbf{x}_k\mathbf{x}_l < \varepsilon$ whenever $k, l > n$.

Definition 25. A space $(\mathfrak{S}, D)$ is called *D-complete* (or simply, complete) if every Cauchy sequence in it converges to a point.

That is, in a complete space, for any Cauchy sequence $\mathbf{x_1}, \mathbf{x}_2, \ldots$, there is a point $\mathbf{x} \in \mathfrak{S}$ such that $\mathbf{x}_n \to \mathbf{x}$. For example, if stimuli are represented by points in a closed region of $\mathbb{R}^n$, and the convergence $\mathbf{x}_n \to \mathbf{x}$ coincides with the usual convergence of n-element vectors, then the space is complete.

The main mathematical fact we are interested in is as follows.

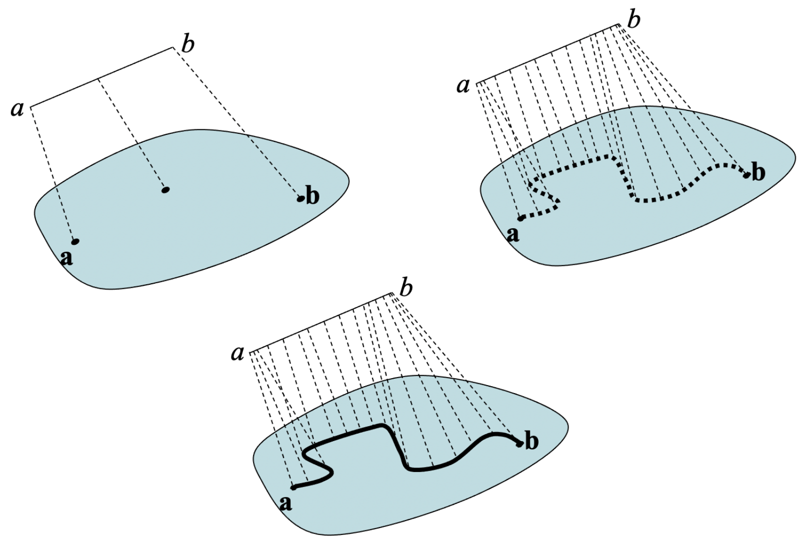

Figure 15: An informal illustration of Theorems 26 and 27: by adding intermediate points for every pair of successive points one can create at the limit a path connecting $\mathbf{a}$ to $\mathbf{b}$, with its D-length not exceeding $D\mathbf{ab}$. The infimum of then D-lengths of all such paths equals $G\mathbf{ab}$.

Theorem 26. *In a complete space* $(\mathfrak{S}, D)$ *with intermediate points, any point* $\mathbf{a}$ *can be connected to any point* $\mathbf{b}$ *by a path* $\mathbf{f}$ *with*

$$D\mathbf{f} \leq D\mathbf{ab}.$$

A proof of this statement known to us is rather involved (see Section 10 for a reference), and we will omit it here. Figure 15 provides an intuitive illustration.

A consequence of this theorem that is of special importance for us is as follows. In any sequence of chains-on-nets $\mathbf{X}_n$ connecting $\mathbf{a}$ to $\mathbf{b}$, with $D\mathbf{X}_n \to G\mathbf{ab}$, each link $\mathbf{x}_{i_n}\mathbf{x}_{i_n+1}$ in each chain $\mathbf{X}_n$ can be replaced with a path $\mathbf{f}_{i_n}$ connecting $\mathbf{x}_{i_n}$ to $\mathbf{x}_{i_n+1}$, such that $D\mathbf{f}_{i_n} \leq D\mathbf{x}_{i_n}\mathbf{x}_{i_n+1}$. This would create a path $\mathbf{f}_n$ connecting $\mathbf{a}$ to $\mathbf{b}$, with $D\mathbf{f}_n \leq D\mathbf{X}_n$. Hence

$$\inf_{\mathbf{f}\in\mathcal{P}_{\mathbf{a}}^{\mathbf{b}}} D\mathbf{f} \leq \liminf_{n\to\infty} D\mathbf{f} \leq \lim_{n\to\infty} D\mathbf{X}_n = G\mathbf{ab}. \tag{32}$$

Combining this with (31), we establish

Theorem 27. *In a complete space* $(\mathfrak{S}, D)$ *with intermediate points, the quasimetric dissimilarity* G *is intrinsic:*

$$G\mathbf{ab} = \inf_{\mathbf{f}\in\mathcal{P}_{\mathbf{a}}^{\mathbf{b}}} D\mathbf{f}.$$

8 Dissimilarity Cumulation in Euclidean spaces

8.1 Introduction

We are now prepared to see how the general theory of path length can be specialized to a variant of (Finsler) differential geometry. We assume that in the canonical space of stimuli $(\mathfrak{S}, D)$, the set $\mathfrak{S}$ is an an *open connected* region of the Euclidean n-space $\mathbb{R}^n$. The Euclidean n-space is endowed with the global coordinate system,

$$\mathbf{x} = \left(x^1, \ldots, x^n\right),$$

and the conventional metric

$$E\mathbf{ab} = |\mathbf{a} - \mathbf{b}|\,. \tag{33}$$

Recall that the connectedness of $\mathfrak{S}$ means that it cannot be presented as a union of two open nonempty sets. In the Euclidean space this notion is equivalent to *path-connectedness*: any two points can be connected by a path.

Among all paths we focus on continuously differentiable ones. We develop a way of measuring the value $F\left(\mathbf{f}(x), \dot{\mathbf{f}}(x)\right)$ of the tangent vector $\dot{\mathbf{f}}(x)$ to the path $\mathbf{f}|\,[a,b]$ at point x, by showing (under certain assumptions) that

$$\widehat{F}\left(\mathbf{f}(x), \dot{\mathbf{f}}(x)\right) = \lim_{s \to 0+} \frac{G\mathbf{f}(x)\,\mathbf{f}(x+s)}{s}.$$

The D-length of the path is then computed as

$$\int_a^b \widehat{F}\left(\mathbf{f}(x), \dot{\mathbf{f}}(x)\right) \mathrm{d}x.$$

The idea is illustrated in Figure 16.

We begin now a systematic development.

Definition 28. The *tangent space* $\mathbb{T}_\mathbf{p}$ at a point $\mathbf{p}$ of $\mathfrak{S}$ is the set $\{\mathbf{p}\} \times \mathbb{U}^n$, where $\mathbb{U}^n$ is the *vector space*

$$\{\mathbf{u} = \mathbf{x} - \mathbf{p} : \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{p}\}$$

endowed with the Euclidean vector norm $|\mathbf{u}|$ and the standard topology. The n-vectors $\mathbf{u} \in \mathbb{U}^n$ are referred to as *directions*, and the elements $(\mathbf{p}, \mathbf{u})$ of $\mathbb{T}_\mathbf{p}$ as *line elements*. The set of all line elements

$$\mathbb{T} = \mathfrak{S} \times \mathbb{U}^n = \bigcup_{\mathbf{p} \in \mathfrak{S}} \mathbb{T}_\mathbf{p}$$

is called the *tangent bundle* of the space $\mathfrak{S}$.

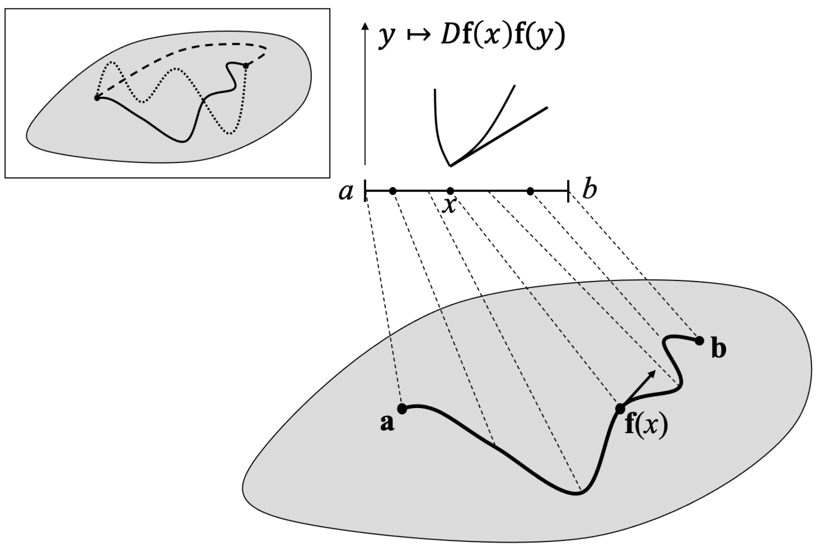

Figure 16: As the point on the path moves away from a position $\mathbf{f}(x)$, the dissimilarity $D\mathbf{f}(x)\mathbf{f}(y)$ increases from zero, and the rate of this increase, $\mathrm{d}D\mathbf{f}(x)\mathbf{f}(x+s)/\mathrm{d}y|_{s=0}$ is shown by the slope of the tangent line in the graph of $y \mapsto D\mathbf{f}(x)\mathbf{f}(y)$. This derivative then is integrated with respect to x from a to b to obtain the length of the path $\mathbf{f}$. If this derivative only depends on $\mathbf{f}(x)$ and $\mathrm{d}\mathbf{f}(x)/\mathrm{d}x$ (assuming the path is continuously differentiable), then it can be viewed as a way of measuring the tangent vector to the path as a point moves along it, $F(\mathbf{f}(x), \mathrm{d}\mathbf{f}(x)/\mathrm{d}x)$. The infimum of the lengths of all such smooth paths connecting $\mathbf{a}$ to $\mathbf{b}$ is then taken for the value of $G\mathbf{ab}$.

This definition deviates from the traditional one, which does not include the point $\mathbf{p}$ explicitly, but it is more convenient for our purposes. In the more general case of a *differentiable manifold* the vector space $\mathbb{U}^n$ should be redefined. Note that the vectors in $\mathbb{U}^n$ do not represent stimuli, but we still use boldface letters to denote them. In the context of Euclidean spaces the boldface notation for both stimuli and directions can simply be taken as indicating vectors.

For any $\mathbf{u} \in \mathbb{U}^n$ the notation $\overline{\mathbf{u}}$ will be used for the unit vector codirectional with $\mathbf{u}$:

$$\overline{\mathbf{u}} = \frac{\mathbf{u}}{|\mathbf{u}|}, \qquad |\overline{\mathbf{u}}| = 1. \tag{34}$$

8.2 Submetric Function

We make the following two assumptions about the space $(\mathfrak{S}, D)$ and its relation to $(\mathfrak{S}, E)$.

($\mathcal{E}1$) The topologies of $(\mathfrak{S}, D)$ and $(\mathfrak{S}, E)$ coincide.

The coincidence of the D-topology and the Euclidean topology means that the notion of convergence,

$$\mathbf{a}_n \to \mathbf{a}, \tag{35}$$

means simultaneously $D\mathbf{a}_n\mathbf{a} \to 0$ and $|\mathbf{a}_n - \mathbf{a}| \to 0$. As a result, all topological concepts (openness, continuity, compactness, etc.) can be used without the prefixes D, G, or E. In particular, dissimilarity $D\mathbf{xy}$ and metric $G\mathbf{xy}$ are continuous in $(\mathbf{x}, \mathbf{y})$ with respect to the usual Euclidean topology.

Note, however, that the notions of uniform convergence in $(\mathfrak{S}, D)$ and $(\mathfrak{S}, E)$ are not assumed to coincide. Thus, it is possible that $D\mathbf{a}_n\mathbf{b}_n \to 0$ but $|\mathbf{a}_n - \mathbf{b}_n| \not\to 0$, or vice versa. In particular, dissimilarity $D\mathbf{xy}$ and metric $G\mathbf{xy}$ are not generally uniformly continuous in the Euclidean sense.

($\mathcal{E}2$) For any $\mathbf{x}, \mathbf{a}_n, \mathbf{b}_n \in \mathfrak{S}$ ($\mathbf{a}_n \neq \mathbf{b}_n$) and any unit vector $\overline{\mathbf{u}}$, if $\mathbf{a}_n \to \mathbf{x}$, $\mathbf{b}_n \to \mathbf{x}$, and $\overline{\mathbf{b}_n - \mathbf{a}_n} \to \overline{\mathbf{u}}$ (see Figure 17), then

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{|\mathbf{b}_n - \mathbf{a}_n|}$$

tends to a positive limit, denoted $F(\mathbf{x}, \overline{\mathbf{u}})$.

Putting $\mathbf{a}_n = \mathbf{x}$ and $\overline{\mathbf{b}_n - \mathbf{a}_n} = \overline{\mathbf{u}}$ in Assumption $\mathcal{E}2$, and denoting $\mathbf{b}_n = \mathbf{x} + \overline{\mathbf{u}}s$, the function $F(\mathbf{x}, \overline{\mathbf{u}})$ can be presented as

$$F(\mathbf{x}, \overline{\mathbf{u}}) = \lim_{s \to 0+} \frac{D\mathbf{x}\,[\mathbf{x} + \overline{\mathbf{u}}s]}{s}. \tag{36}$$

We now generalize this function to apply to any vector $\mathbf{u}$, not just the unit one.

Definition 29. The function

$$F : \mathbb{T} \cup \{(\mathbf{x}, \mathbf{0}) : \mathbf{x} \in \mathfrak{S}\} \to \mathbb{R}$$

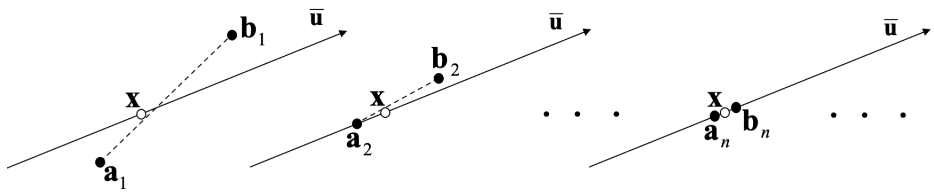

Figure 17: An illustration for Assumption $\mathcal{E}2$. Shown are a point $\mathbf{x}$ (open circle), a direction $\overline{\mathbf{u}}$ attached to it, and (in successive panels from left to right) pairs of points $(\mathbf{a}_1, \mathbf{b}_1)$, $(\mathbf{a}_2, \mathbf{b}_2)$, ..., $(\mathbf{a}_n, \mathbf{b}_n)$, ... gradually converging to $\mathbf{x}$ so that the dashed line connecting them (and directed from $\mathbf{a}_n$ to $\mathbf{b}_n$) gradually aligns with the the direction $\overline{\mathbf{u}}$. The assumption says that in this situation the dissimilarity $D\mathbf{a}_n\mathbf{b}_n$ and the Euclidean distance $|\mathbf{b}_n - \mathbf{a}_n|$ are comeasurable in the small: neither of them tends to zero infinitely faster than the other.

defined as

$$F(\mathbf{x}, \mathbf{u}) = \begin{cases} \lim_{s\to 0+} \frac{D\mathbf{x}[\mathbf{x}+\mathbf{u}s]}{s} & \text{if } \mathbf{u} \neq \mathbf{0} \\ 0 & \text{if } \mathbf{u} = \mathbf{0} \end{cases}, \tag{37}$$

is called a *submetric function.*

The standard term for $F(\mathbf{x}, \mathbf{u})$ in differential geometry is "metric function." It can, however, be easily confused with a metric on the space of stimuli, such as $G\mathbf{ab}$. To prevent this confusion, we use the non-standard term "submetric function."

Theorem 30. *$F(\mathbf{x}, \mathbf{u})$ is well-defined for any $(\mathbf{x}, \mathbf{u}) \in \mathbb{T} \cup \{(\mathbf{x}, \mathbf{0}) : \mathbf{x} \in \mathfrak{S}\}$. It is positive for $\mathbf{u} \neq \mathbf{0}$, continuous in $(\mathbf{x}, \mathbf{u})$, and Euler homogeneous in $\mathbf{u}$.*

Euler homogeneity in $\mathbf{u}$ means that for any $k > 0$, $F(\mathbf{x}, k\mathbf{u}) = kF(\mathbf{x}, \mathbf{u})$. See Appendix for a proof.

Assumption $\mathcal{E}2$ can now be strengthened as follows.

Theorem 31. *For any $\mathbf{a}_n, \mathbf{b}_n \in \mathfrak{s} \subset \mathfrak{S}$, if $\mathfrak{s}$ is compact and $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$ ($\mathbf{a}_n \neq \mathbf{b}_n$) then*

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{F(\mathbf{a}_n, \mathbf{b}_n - \mathbf{a}_n)} \to 1.$$

Indeed, rewrite

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{F(\mathbf{a}_n, \mathbf{b}_n - \mathbf{a}_n)} = \frac{D\mathbf{a}_n\mathbf{b}_n}{F\left(\mathbf{a}_n, \overline{\mathbf{b}_n - \mathbf{a}_n}\right)|\mathbf{b}_n - \mathbf{a}_n|},$$

and denote either lim inf or lim sup of this ratio by l. There is an infinite subsequence of $(\mathbf{a}_n, \mathbf{b}_n)$ (without loss of generality, the sequence itself) for which

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{F\left(\mathbf{a}_n, \overline{\mathbf{b}_n - \mathbf{a}_n}\right)|\mathbf{b}_n - \mathbf{a}_n|} \to l.$$

But within a compact set $\mathfrak{s}$ one can always select from this sequence $(\mathbf{a}_n, \mathbf{b}_n)$ a subsequence with $\mathbf{a}_n \leftrightarrow \mathbf{x}$, $\mathbf{b}_n \leftrightarrow \mathbf{x}$, for some $\mathbf{x}$; and due to the compactness of the set $\overline{\mathfrak{u}}$ of all unit directions, one can always select a subsequence of this subsequence with $\overline{\mathbf{b}_n - \mathbf{a}_n} \to \overline{\mathbf{u}}$, for some $\overline{\mathbf{u}}$. In this resulting subsequence (again, without changing the indexation for convenience),

$$F\left(\mathbf{a}_n, \overline{\mathbf{b}_n - \mathbf{a}_n}\right) \to F\left(\mathbf{a}, \overline{\mathbf{u}}\right),$$

whence

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{|\mathbf{b}_n - \mathbf{a}_n|} \to lF\left(\mathbf{a}, \overline{\mathbf{u}}\right).$$

By Assumption $\mathcal{E}2$ then, $l = 1$. Since this result holds for both lim inf and lim sup of the original ratio, the statement of the theorem follows.

8.3 Indicatrices

Definition 32. The function

$$\mathbf{1} : \mathbb{T} \to \mathbb{U}^n$$

defined by

$$\mathbf{1}\left(\mathbf{a}, \mathbf{u}\right) = \frac{\mathbf{u}}{F\left(\mathbf{a}, \mathbf{u}\right)}$$

is called the *radius-vector function* associated with (or corresponding to) the submetric function $F(\mathbf{a}, \mathbf{u})$. The values of this function are referred to as *radius-vectors*. For a fixed $\mathbf{a} \in \mathfrak{S}$, the function $\mathbf{u} \mapsto \mathbf{1}(\mathbf{a}, \mathbf{u})$ is called the *indicatrix centered at* (or *attached to*) the point $\mathbf{a}$. The set

$$\mathbb{I}_\mathbf{a} = \{\mathbf{u} \in \mathbb{U}^n : F(\mathbf{a}, \mathbf{u}) \leq 1\}$$

is called the *body* of this indicatrix, and the set

$$\delta\mathbb{I}_\mathbf{a} = \{\mathbf{u} \in \mathbb{U}^n : F(\mathbf{a}, \mathbf{u}) = 1\}$$

is called its *boundary*.

Figure 18 provides an illustration for the relationship between $F(\mathbf{a}, \mathbf{u})$ and $\mathbf{1}(\mathbf{a}, \mathbf{u})$.

Note that $\{\mathbf{a}\} \times \mathbb{I}_\mathbf{a}$ is a subset of the tangent space $\mathbb{T}_\mathbf{a}$. Note also that the body (or the boundary) of an indicatrix is a set of vectors in $\mathbb{U}^n$ emanating from a common origin. The boundary should not be thought of as the set of the endpoints of the radius-vectors: the latter set does not determine the

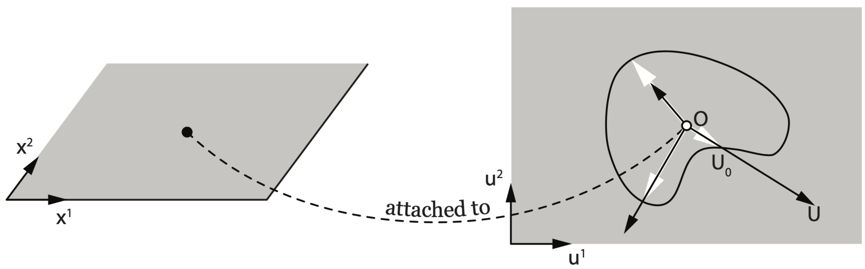

Figure 18: An indicatrix (right) attached to a point in plane (left). The value of the submetric function F at this point and any vector $\overrightarrow{OU}$ is computed as the ratio of $\overrightarrow{OU}$ to the codirectional radius-vector of the indicatrix, $\overrightarrow{OU_0}$ (shown in white).

indicatrix uniquely, as one should also know the position of the origin within the boundary (see Figure 19). Not all points within a given set of endpoints may serve as points of origin: by definition, there can be no endpoint A on the boundary which is not connected to the origin O by a vector $\overrightarrow{OA} \in \delta\mathbb{I}_{\mathbf{a}}$, and the boundary cannot have two codirectional but non-identical vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ (see Figure 20): indeed, if

$$\frac{\overrightarrow{OA}}{\overrightarrow{OB}} = k \neq 1,$$

then

$$\frac{F\left(\mathbf{a}, \overrightarrow{OA}\right)}{F\left(\mathbf{a}, \overrightarrow{OB}\right)} = k,$$

so one of the vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ does not belong to $\delta\mathbb{I}_{\mathbf{a}}$.

Figure 21 offers a geometric interpretation for measuring the length of a smooth path, to be rigorously justified later.

We now list basic, almost obvious, properties of the unit vector function and the corresponding indicatrices.

Theorem 33. *The following statements hold true:*

(i) $\mathbf{1}(\mathbf{a}, \mathbf{u})$ *is continuous;*

(ii) $\mathbf{1}(\mathbf{a}, k\mathbf{u}) = \mathbf{1}(\mathbf{a}, \mathbf{u})$ for all $(\mathbf{a}, \mathbf{u}) \in \mathbb{T}$ and all $k > 0$ (Euler homogeneity in $\mathbf{u}$ of order zero);

(iii) for any $\mathbf{a} \in \mathfrak{S}$, *the mapping* $\overline{\mathbf{u}} \mapsto \mathbf{1}(\mathbf{a}, \overline{\mathbf{u}})$ *is a homeomorphism;*

(iv) $\mathbb{I}_{\mathbf{a}}$ *is a compact set in* $\mathbb{U}^n$*;*

(v) $\delta\mathbb{I}_{\mathbf{a}}$ *is a compact set in* $\mathbb{U}^n$*;*

(vi) for any $\mathbf{a} \in \mathfrak{S}$, *there are two positive reals* $k_{\mathbf{a}}, K_{\mathbf{a}}$ *such that*

$$k_{\mathbf{a}} \leq |\mathbf{1}(\mathbf{a}, \mathbf{u})| \leq K_{\mathbf{a}}$$

for all $\mathbf{u} \in \mathbb{U}$, and the values $k_{\mathbf{a}}, K_{\mathbf{a}}$ *are attained by* $\mathbf{1}(\mathbf{a}, \mathbf{u})$ *at some* $\mathbf{u}$.

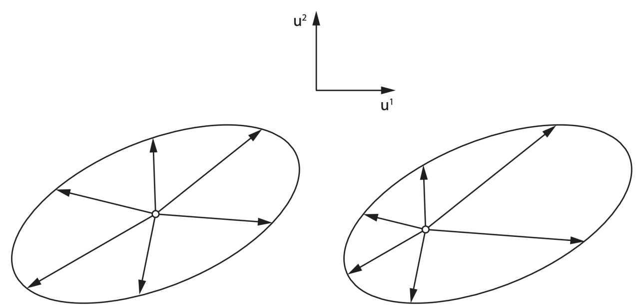

Figure 19: The two indicatrices are different (consist of different vectors) although they have identical sets of endpoints.

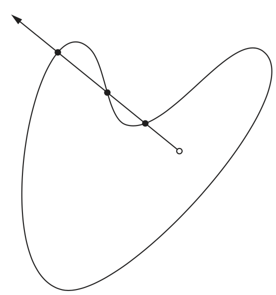

Figure 20: This combination of a set of endpoints with a position of the origin does not form an indicatrix, because a radius-vector from the origin (shown by the open circle) intersects the boundary at more than one point.

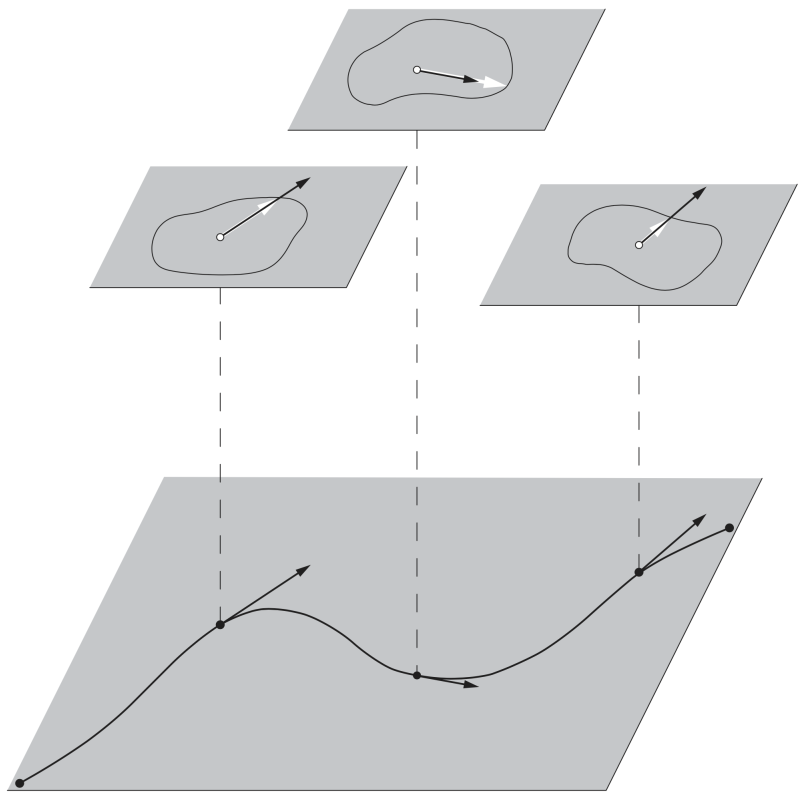

Figure 21: Geometric interpretation of how indicatrices measure tangents to a smooth path: by centering the indicatrix $\mathbb{I}_{\mathbf{f}(x)}$ at each point $\mathbf{f}\,(x)$, one measures the magnitude of the tangent at this point by relating it to the codirectional radius-vector of the indicatrix, as explained in Figure 18. The length of the path $\mathbf{f}|[a,b]$ then is obtained by integrating this magnitude from a to b. For the conventional Euclidean length all indicatrices should be unit-radius circles.

The proof of Propositions (i) and (ii) follow from the continuity and Euler homogeneity of $F(\mathbf{a},\mathbf{u})$. Denoting $\mathbf{1}(\mathbf{a},\overline{\mathbf{u}})$ by $\widetilde{\mathbf{u}}$, Proposition (iii) follows from the relations

$$\frac{\widetilde{\mathbf{u}}}{|\widetilde{\mathbf{u}}|}=\overline{\mathbf{u}}$$

and

$$\widetilde{\mathbf{u}}=\frac{\overline{\mathbf{u}}}{F(\mathbf{a},\overline{\mathbf{u}})},$$

because both these functions are injective and continuous. The continuous function $\overline{\mathbf{u}}\mapsto\widetilde{\mathbf{u}}$ induces the continuous function $k\overline{\mathbf{u}}\mapsto k\widetilde{\mathbf{u}}$ for all $k\in[0,1]$, and (iv)-(v) then follow from the compactness of the unit Euclidean ball $\{k\overline{\mathbf{u}}:k\in[0,1]\}$ and the unit Euclidean sphere $\{\overline{\mathbf{u}}\}$. The continuous mapping $\overline{\mathbf{u}}\mapsto\mathbf{1}(\mathbf{a},\overline{\mathbf{u}})$ of the compact unit Euclidean sphere should attain a maximum value $K_{\mathbf{a}}$and a minimum value $k_{\mathbf{a}}$, and we get (vi) due to (ii).

Based on Theorem 33, we can think of an indicatrix boundary as a homeomorphically "deformed" Euclidean $(n-1)$-sphere "sandwiched" between two concentric Euclidean $(n-1)$-spheres of radii $k_{\mathbf{a}}>0$ and $K_{\mathbf{a}}\geq k_{\mathbf{a}}$. Figure 22 illustrates this for $n=2$.

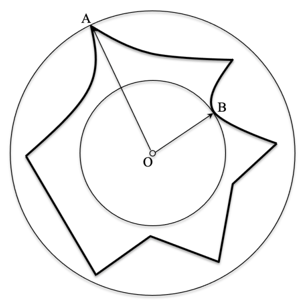

Figure 22: A planar indicatrix (whose origin point O is attached to a point $\mathbf{a}$ in $\mathfrak{S}$) is sandwiched between two concentric circles of radii $\left|\overrightarrow{OA}\right|=K_{\mathbf{a}}$ and $\left|\overrightarrow{OB}\right|=k_{\mathbf{a}}$.

8.4 Convex combinations and hulls

To further investigate the properties of indicatrices, we need to recall certain notions from linear algebra. In the vector space $\mathbb{U}^n$, a linear combination

$$\mathbf{u}=\lambda_1\mathbf{v}_1+\ldots+\lambda_m\mathbf{v}_m,\ m\geq 1, \tag{38}$$

is called a *convex!combination* of $\mathbf{v}_1,\ldots,\mathbf{v}_m$ if $\lambda_i\geq 0$ for $i=1,\ldots,m$, and

$$\lambda_1+\ldots+\lambda_m=1.$$

From a geometric point of view, the set of convex combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ forms an $(m-1)$-dimensional facet with vertices $\mathbf{v}_1, \ldots, \mathbf{v}_m$. The following therefore is obviously true.

Lemma 34. *If $\alpha\mathbf{u}$ is a convex combination of $a_1\mathbf{v}_1, \ldots, a_m\mathbf{v}_m$ and $\beta\mathbf{u}$ is a convex combination of $b_1\mathbf{v}_1, \ldots, b_m\mathbf{v}_m$, with $a_i \geq b_i$ for $i = 1, \ldots, m$ and at least one inequality being strict, then $\alpha > \beta$.*

Figure 23 provides an illustration.

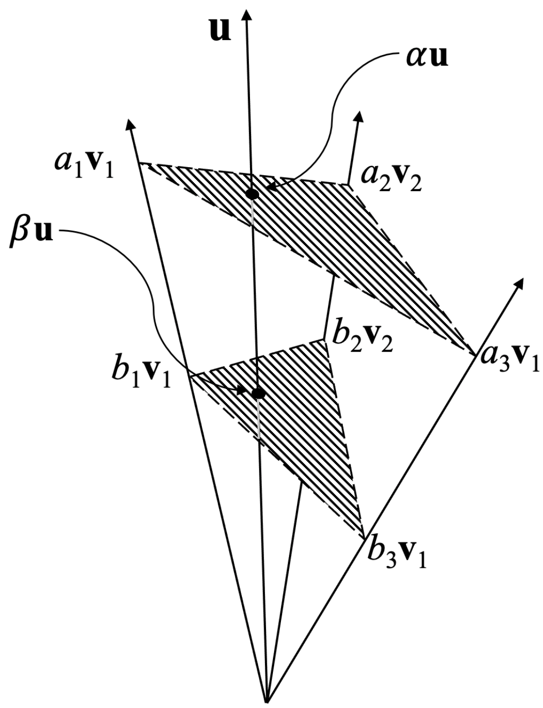

Figure 23: Illustration for Lemma 34: a direction within the cone formed by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ first crosses the lower facet and then the higher facet.

Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are called *affinely dependent* if, for some $\gamma_1, \ldots, \gamma_m$, not all zero,

$$\begin{array}{c} \gamma_1\mathbf{v}_1 + \ldots + \gamma_m\mathbf{v}_m = \mathbf{0} \\ \gamma_1 + \ldots + \gamma_m = 0 \end{array} . \tag{39}$$

If $\mathbf{u}$ is a convex combination of affinely dependent vectors, we have simultaneously

$$\left\{ \begin{array}{l} \lambda_1\mathbf{v}_1 + \ldots + \lambda_m\mathbf{v}_m = \mathbf{u} \\ \gamma_1\mathbf{v}_1 + \ldots + \gamma_m\mathbf{v}_m = \mathbf{0} \end{array} \right. ,$$

where

$$\begin{cases} \lambda_1 + \ldots, +\lambda_m = 1 \\ \gamma_1 + \ldots, +\gamma_m = 0 \end{cases},$$

all λ's are nonnegative and some γ's are nonzero (which means that at least one of them is positive and at least one negative). To exclude trivial cases, let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be pairwise distinct and let $\lambda_i > 0$ for $i = 1, \ldots, m$. Let c be the minimum $\left|\frac{\lambda_i}{\gamma_i}\right|$ among all negative ratios $\frac{\lambda_i}{\gamma_i}$. Then at least one of the coefficients in the representation

$$\mathbf{u} = (\lambda_1 + c\gamma_1)\,\mathbf{v}_1 + \ldots, + (\lambda + c\gamma_m)\,\mathbf{v}_m$$

is zero, while all other coefficients are nonnegative and sum to 1. This means that $\mathbf{u}$ is a convex combination of at most $m-1$ elements of $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, and we have

Lemma 35. *If $\mathbf{u} \in \mathbb{U}^n$ is a convex combination of affinely dependent $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{U}^n$, then $\mathbf{u}$ is a convex combination of some $m' < m$ elements of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.*

The following corollary of the lemma is known as a Carathéodory theorem.

Corollary 36. *If $\mathbf{u}, \mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{U}^n$, $m > n+1$, and $\mathbf{u}$ is a convex combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$, then $\mathbf{u}$ is a convex combination of at most $n+1$ elements of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.*

This follows from the fact that if $m > n+1$, any $\mathbf{v}_1, \ldots, \mathbf{v}_m$ in $\mathbb{U}^n$ are affinely dependent. Indeed, since $\operatorname{rank}(\mathbf{v}_1, \ldots, \mathbf{v}_m) \leq n$, there should exist reals $\alpha_1, \ldots, \alpha_m$, not all zero, such that the system of $n+1$ linear equations

$$\begin{cases} \alpha_1\mathbf{v}_1 + \ldots + \alpha_m\mathbf{v}_m = \mathbf{0} \\ \alpha_1 + \ldots + \alpha_m = 0 \end{cases}$$

is satisfied.

A subset $\mathbb{V}$ of $\mathbb{U}^n$ is said to be *convex* if it contains any convex combination

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y}, \quad \lambda \in [0,1],$$

of any two of its elements $\mathbf{x}, \mathbf{y}$. By induction from 2 to $(n+1)$-element subsets of $\mathbb{V}$ (which is sufficient by Corollary 36), we see that a convex set $\mathfrak{X} \subset \mathbb{U}^n$contains all convex combinations of *all* finite subsets of $\mathbb{V}$.

For any $\mathfrak{X} \subset \mathbb{U}^n$ the set of all convex combinations of all $(n+1)$-tuples of elements of $\mathbb{V}$ is called the *convex hull* of $\mathbb{V}$ and is denoted conv$\mathbb{V}$. Again, conv$\mathbb{V}$ is, clearly, the set of all convex combinations of all finite subsets of $\mathbb{V}$, and it is the smallest convex subset of $\mathbb{U}^n$ containing $\mathbb{V}$.

Consider now an indicatrix $\mathbb{I}_\mathbf{a}$ and its convex hull. The following is obvious.

Lemma 37. *For any indicatrix $\mathbb{I}_\mathbf{a}$, conv$\mathbb{I}_\mathbf{a}$ is compact in $\mathbb{U}^n$.*

Let now $\mathbf{u} \in \text{conv}\mathbb{I}_{\mathbf{a}}$. Then, for some $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{I}_{\mathbf{a}}$ and some nonnegative reals $\lambda_1, \ldots, \lambda_m$ that sum to 1,

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \ldots + \lambda_m \mathbf{v}_m.$$

But then

$$|\mathbf{u}| = |\lambda_1 \mathbf{v}_1 + \ldots + \lambda_m \mathbf{v}_m| \leq \lambda_1 |\mathbf{v}_1| + \ldots + \lambda_m |\mathbf{v}_m| \\ \leq (\lambda_1 + \ldots, +\lambda) K_{\mathbf{a}} = K_{\mathbf{a}},$$

where $K_{\mathbf{a}}$ denotes $\max_{\mathbf{u} \in \mathbb{I}_{\mathbf{a}}} |\mathbf{u}|$ (whose existence is stated in Theorem 33, v). We have therefore

Lemma 38. *For any* $\mathbf{a} \in \mathfrak{S}$,

$$\max_{\mathbf{u} \in \text{conv}\mathbb{I}_{\mathbf{a}}} |\mathbf{u}| = \max_{\mathbf{u} \in \mathbb{I}_{\mathbf{a}}} |\mathbf{u}|.$$

Definition 39. For any $(\mathbf{a}, \mathbf{u}) \in \mathbb{T}$, the quantity

$$\kappa(\mathbf{a}, \mathbf{u}) = \max\{\alpha > 0 : \alpha \mathbf{1}(\mathbf{a}, \mathbf{u}) \in \text{conv}\mathbb{I}_{\mathbf{a}}\}$$

is called the *maximal production factor* for $\mathbf{u}$ in $\mathbb{I}_{\mathbf{a}}$, and the vector $\kappa(\mathbf{a}, \mathbf{u}) \mathbf{1}(\mathbf{a}, \mathbf{u})$ is called the *maximal production* of (or *maximally produced*) $\mathbf{u}$ in $\mathbb{I}_{\mathbf{a}}$.

This is clearly a well-defined function, because it follows from the compactness of $\text{conv}\mathbb{I}_{\mathbf{a}}$ that

Lemma 40. *For any* $\mathbf{a} \in \mathfrak{S}$, *every* $\mathbf{u} \in \mathbb{U}^n$ *has its maximal production in* $\mathbb{I}_{\mathbf{a}}$.

The following statement holds because $\alpha \mathbf{u}$ and $\mathbf{u}$ have one and the same maximal production in $\mathbb{I}_{\mathbf{a}}$.

Lemma 41. *The function* $\kappa(\mathbf{a}, \mathbf{u})$ *is Euler homogeneous of zero order,*

$$\kappa(\mathbf{a}, \alpha \mathbf{u}) = \kappa(\mathbf{a}, \mathbf{u}).$$

Finally, we need to observe the following.

Lemma 42. *For any* $(\mathbf{a}, \mathbf{u}) \in \mathbb{T}$, *the maximal production* of $\mathbf{u}$ in $\mathbb{I}_{\mathbf{a}}$ *can be presented as a convex combination of* n *(not necessarily distinct) radius-vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \delta \mathbb{I}_{\mathbf{a}}$.

See Appendix for a proof.

Figure 24 provides an illustration for this lemma on three-dimensional indicatrices. (It also illustrates the useful notion of the degree of flatness for a radius vector within the body of the indicatrix.)

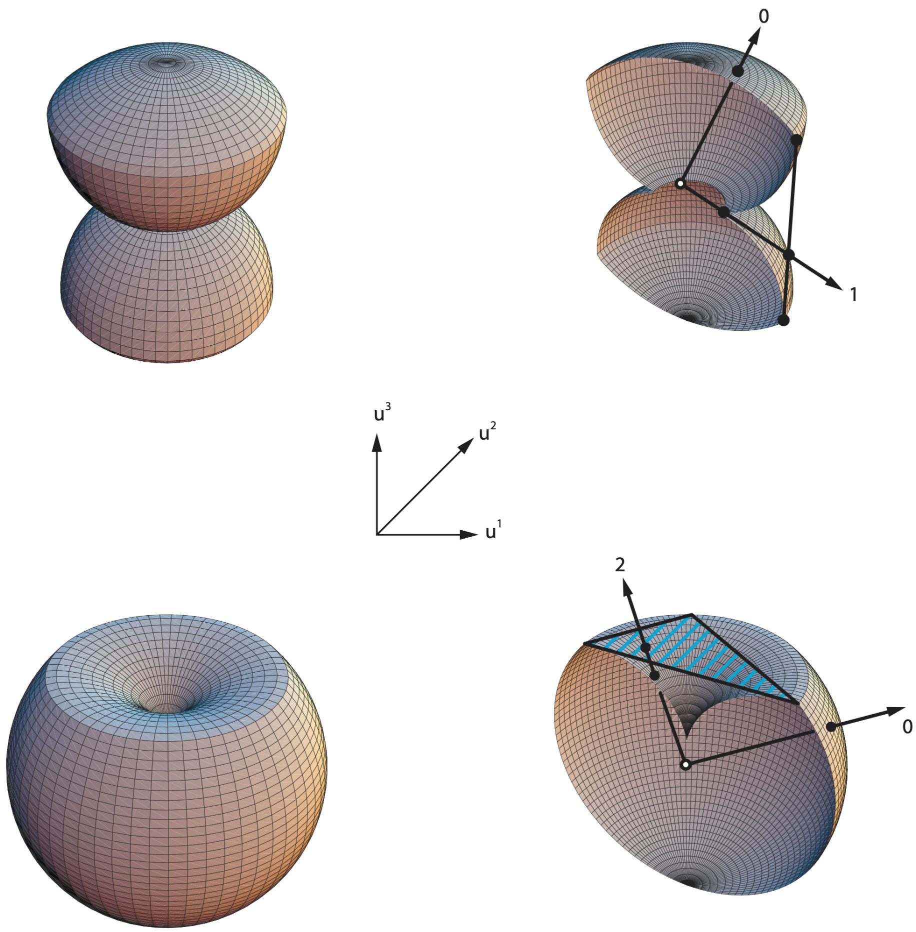

Figure 24: Two indicatrices in $\mathbb{U}^3$ (left) and their cross-sections (right) showing the position of the origin (white dots). The maximal productions of two vectors are shown in each of the indicatrices, as parts of the vectors between the origin to the farthest black dot. The number attached to a vector $\mathbf{v}$ shows the degree of flatness $r-1$ of the indicatrix in the direction $\mathbf{v}$, where r is the maximum number of linearly independent radius-vectors whose convex combination equals the maximum production of $\mathbf{v}$ in the body of the indicatrix.

8.5 Minimal submetric function and convex hulls of indicatrices

In this section we consider the problem of finding a *geodesic in the small*, a shortest path connecting stimuli $\mathbf{a}$ and $\mathbf{a}+\mathbf{u}s$ as $s \to 0$. It will be established later (Section 8.6) that $G\mathbf{a}(\mathbf{a}+\mathbf{u}s)$ in $\mathfrak{S} \subseteq \mathbb{R}^n$ can be approximated by concatenation of $m \leq n$ straight line segments with lengths $F(\mathbf{a},\mathbf{u}_i)\,s$ for some vectors $\mathbf{u}_1,\ldots,\mathbf{u}_m$ summing to $\mathbf{u}$. So we begin with investigating the minimal value for certain sums of $F(\mathbf{a},\mathbf{u}_i)$.

Definition 43. A sequence of vectors $(\mathbf{u}_1,\ldots,\mathbf{u}_m)$ in $\mathbb{U}^n$, $m \geq 1$, is said to form a *minimizing vector chain* for a line element $(\mathbf{a},\mathbf{u}) \in \mathbb{T}$,

$$\mathbf{u} = \mathbf{u}_1 + \ldots + \mathbf{u}_m$$

and

$$F(\mathbf{a},\mathbf{u}_1) + \ldots + F(\mathbf{a},\mathbf{u}_m) = \min\left\{F(\mathbf{a},\mathbf{v}_1) + \ldots + F(\mathbf{a},\mathbf{v}_k)\right\},$$

where the minimum is taken over all $k \geq 1$ and all finite sequences $(\mathbf{v}_1,\ldots,\mathbf{v}_k)$ in $\mathbb{U}^n$ such that

$$\mathbf{u} = \mathbf{v}_1 + \ldots + \mathbf{v}_k.$$

Note that this definition does not require that $\mathbf{u}_1,\ldots,\mathbf{u}_m$ be pairwise distinct, so a minimizing chain for $(\mathbf{a},\mathbf{u})$ may, e.g., be $\left\{\frac{1}{n}\mathbf{u},\ldots,\frac{1}{n}\mathbf{u}\right\}$(which is equivalent to $\mathbf{u}$ alone being a minimizing vector chain for $(\mathbf{a},\mathbf{u})$ too). Note also, that if $(\mathbf{u}_1,\ldots,\mathbf{u}_m)$ is a minimizing chain, then so is any permutation thereof.

Theorem 44. *A minimizing chain for any* $(\mathbf{a},\mathbf{u}) \in \mathbb{T}$ *exists and consists of* n *(not necessarily distinct) nonzero vectors* $\mathbf{u}_1,\ldots,\mathbf{u}_m$*, such that*

$$F(\mathbf{a},\mathbf{u}_1) + \ldots + F(\mathbf{a},\mathbf{u}_n) = \frac{F(\mathbf{a},\mathbf{u})}{\kappa(\mathbf{a},\mathbf{u})},$$

where $\kappa(\mathbf{a},\mathbf{u})$ *is the maximal production factor for* $\mathbf{u}$ *in* $\mathbb{I}_{\mathbf{a}}$.

To prove this, we fix $\kappa(\mathbf{a},\mathbf{u}) = \kappa$ as we deal with a fixed $(\mathbf{a},\mathbf{u})$. Consider the maximal production $\kappa\mathbf{1}(\mathbf{a},\mathbf{u})$ of $\mathbf{u}$. By Lemma 42, it can be presented as a convex combination of some n radius-vectors $\widetilde{\mathbf{v}}_1,\ldots,\widetilde{\mathbf{v}}_n$ in $\delta\mathbb{I}_{\mathbf{a}}$,

$$\kappa\mathbf{1}(\mathbf{a},\mathbf{u}) = \lambda_1\widetilde{\mathbf{v}}_1 + \ldots + \lambda_n\widetilde{\mathbf{v}}_n,$$

where all coefficients are nonnegative and sum to 1. Then, denoting

$$\mathbf{v}_i = \frac{\lambda_i}{\kappa}\widetilde{\mathbf{v}}_i, \qquad i = 1,\ldots,n,$$

we have

$$\mathbf{1}(\mathbf{a},\mathbf{u}) = \mathbf{v}_1 + \ldots + \mathbf{v}_n$$

and

$$F(\mathbf{a},\mathbf{v}_1)+\ldots+F(\mathbf{a},\mathbf{v}_n)=\frac{1}{\kappa}.$$

We prove now that for any $\mathbf{w}_1,\ldots,\mathbf{w}_m$ in $\mathbb{U}^n$, if

$$\mathbf{1}(\mathbf{a},\mathbf{u})=\mathbf{w}_1+\ldots+\mathbf{w}_m,$$

then

$$F(\mathbf{a},\mathbf{w}_1)+\ldots+F(\mathbf{a},\mathbf{w}_m)=\delta\geq\frac{1}{\kappa}.$$

Indeed, we have

$$\mathbf{1}(\mathbf{a},\mathbf{u})=F(\mathbf{a},\mathbf{w}_1)\,\mathbf{1}(\mathbf{a},\mathbf{w}_1)+\ldots+F(\mathbf{a},\mathbf{w}_m)\,\mathbf{1}(\mathbf{a},\mathbf{w}_m)$$

and

$$\frac{1}{\delta}\mathbf{1}(\mathbf{a},\mathbf{u})=\frac{F(\mathbf{a},\mathbf{w}_1)}{\delta}\mathbf{1}(\mathbf{a},\mathbf{w}_1)+\ldots+\frac{F(\mathbf{a},\mathbf{w}_{n+1})}{\delta}\mathbf{1}(\mathbf{a},\mathbf{w}_m).$$

That is, $\frac{1}{\delta}\mathbf{1}(\mathbf{a},\mathbf{u})$ is a convex combination of m radius-vectors of $\delta\mathbb{I}_{\mathbf{a}}$. But then

$$\frac{1}{\delta}\leq\kappa.$$

It follows that $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ is a minimizing vector chain for $(\mathbf{a},\mathbf{1}(\mathbf{a},\mathbf{u}))$, with

$$F(\mathbf{a},\mathbf{v}_1)+\ldots+F(\mathbf{a},\mathbf{v}_m)=\frac{1}{\kappa}.$$

The statement of the theorem obtains by putting $\mathbf{u}_i=F(\mathbf{a},\mathbf{u})\,\mathbf{v}_i$, $i=1,\ldots,n$.

We introduce now one of the central notions of the theory.

Definition 45. For any $(\mathbf{a},\mathbf{u})\in\mathbb{T}\cup\{(\mathbf{x},\mathbf{0}):\mathbf{x}\in\mathfrak{S}\}$, the function

$$\widehat{F}(\mathbf{a},\mathbf{u})=\begin{cases}\frac{F(\mathbf{a},\mathbf{u})}{\kappa(\mathbf{a},\mathbf{u})} & \text{if } \mathbf{u}\neq\mathbf{0}\\ 0 & \text{if } \mathbf{u}=\mathbf{0}\end{cases}$$

is called the *minimal submetric function.*

Clearly,

$$\widehat{F}(\mathbf{a},\mathbf{u})\leq F(\mathbf{a},\mathbf{u}).$$

Theorem 46. *The minimal submetric function $\widehat{F}(\mathbf{a},\mathbf{u})$ has all the properties of a submetric function: it is positive for $\mathbf{u}\neq\mathbf{0}$, Euler homogeneous, and continuous.*

See Appendix for a proof.

Theorem 47. *The indicatrix at $\mathbf{a}\in\mathfrak{S}$ associated with $\widehat{F}(\mathbf{a},\mathbf{u})$,*

$$\mathbf{u}\mapsto\widehat{\mathbf{1}}(\mathbf{a},\mathbf{u})=\frac{\mathbf{u}}{\widehat{F}(\mathbf{a},\mathbf{u})},$$

has the body

$$\widehat{\mathbb{I}}_{\mathbf{a}} = \left\{ \mathbf{u} \in \mathbb{U}^n : \widehat{F}(\mathbf{a}, \mathbf{u}) \le 1 \right\} = \mathrm{conv}\mathbb{I}_{\mathbf{a}},$$

where $\mathbb{I}_{\mathbf{a}}$ *is the body of the indicatrix* $\mathbf{u} \mapsto \mathbf{1}(\mathbf{a}, \mathbf{u})$ *associated with* $F(\mathbf{a}, \mathbf{u})$. *The boundary*

$$\delta\widehat{\mathbb{I}}_{\mathbf{a}} = \left\{ \mathbf{u} \in \mathbb{U}^n : \widehat{F}(\mathbf{a}, \mathbf{u}) = 1 \right\}$$

of the indicatrix $\mathbf{u} \mapsto \widehat{\mathbf{1}}(\mathbf{a}, \mathbf{u})$ *is the set of all maximally produced radius-vectors of the indicatrix* $\mathbf{u} \mapsto \mathbf{1}(\mathbf{a}, \mathbf{u})$.

This is essentially a summary of the results established so far. To prove the second statement of the theorem, by Lemma 40 and Theorem 44, the maximal production $\kappa(\mathbf{a}, \mathbf{u}) \mathbf{1}(\mathbf{a}, \mathbf{u})$ of $\mathbf{u}$ in $\mathbb{I}_{\mathbf{a}}$ exists for every $\mathbf{u}$, and

$$\widehat{F}(\mathbf{a}, \mathbf{1}(\mathbf{a}, \mathbf{u})) = \frac{1}{\kappa(\mathbf{a}, \mathbf{u})}.$$

It follows that $\widehat{F}(\mathbf{a}, \mathbf{u}) = 1$ if and only if

$$\mathbf{u} = \kappa(\mathbf{a}, \mathbf{u}) \mathbf{1}(\mathbf{a}, \mathbf{u}).$$

To prove the first statement of the theorem, by Lemma 42, $\kappa(\mathbf{a}, \mathbf{u}) \mathbf{1}(\mathbf{a}, \mathbf{u})$ is a convex combination of some vectors $\mathbf{v}_1, ..., \mathbf{v}_n$ in $\mathbb{I}_{\mathbf{a}}$. But then $c\kappa(\mathbf{a}, \mathbf{u}) \mathbf{1}(\mathbf{a}, \mathbf{u})$ is a convex combination of $c\mathbf{v}_1, ..., c\mathbf{v}_n \in \mathbb{I}_{\mathbf{a}}$ for any $c \in [0, 1]$. It is clear then that $\mathrm{conv}\mathbb{I}_{\mathbf{a}}$ consists of all vectors

$$\mathbf{u} = c\kappa(\mathbf{a}, \mathbf{u}) \mathbf{1}(\mathbf{a}, \mathbf{u}), \qquad c \in [0, 1].$$

But these are precisely the vectors satisfying $\widehat{F}(\mathbf{a}, \mathbf{u}) \le 1$. This completes the proof.

It follows from this theorem that $\widehat{\mathbf{1}}(\mathbf{a}, \mathbf{u})$, $\widehat{\mathbb{I}}_{\mathbf{a}}$, and $\delta\widehat{\mathbb{I}}_{\mathbf{a}}$ have all the properties listed in Theorem 33. If $\delta\mathbb{I}_{\mathbf{a}}$ is a homeomorphically deformed Euclidean sphere sandwiched between two Euclidean spheres of radii $k_{\mathbf{a}}$ and $K_{\mathbf{a}}$, then $\widehat{\delta\mathbb{I}}_{\mathbf{a}}$ is a a homeomorphically deformed (but convex) Euclidean sphere sandwiched between two Euclidean spheres of radii $k^*_{\mathbf{a}}$ and $K_{\mathbf{a}}$ (where $k^*_{\mathbf{a}} \ge k_{\mathbf{a}}$ and $K_{\mathbf{a}}$ is the same for $\delta\mathbb{I}_{\mathbf{a}}$ and $\widehat{\delta\mathbb{I}}_{\mathbf{a}}$, as stated in Lemma 38). Figure 25 illustrates this using the indicatrix shown in Figure 22. Figure 26 shows the convex hulls of the indicatrices shown in Figure 24.

8.6 Length and Metric in Euclidean spaces

Definition 48. A submetric function $F(\mathbf{a}, \mathbf{u})$ is called *convex* if for any $\mathbf{a} \in \mathfrak{S}$ and $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{U}^n$,

$$F(\mathbf{a}, \mathbf{u}_1 + \mathbf{u}_2) \le F(\mathbf{a}, \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2).$$

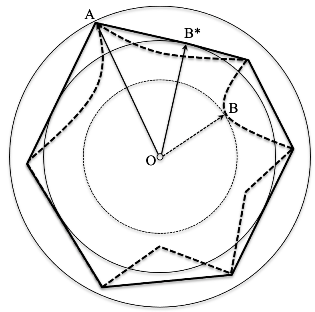

Figure 25: The convex hull of the indicatrix body shown in Figure 22 is sandwiched between $\left|\overrightarrow{OA}\right| = K_{\mathbf{a}}$(the same as for the indicatrix itself) and $\left|\overrightarrow{OB^*}\right| = k^*_{\mathbf{a}}$ which is greater than $\left|\overrightarrow{OB}\right| = k_{\mathbf{a}}$.

Assume, excluding the trivial case, that $\mathbf{u}_1, \mathbf{u}_2$ are not both zero. If $\mathbb{I}_{\mathbf{a}}$ is convex, then the vector

$$\frac{F(\mathbf{a}, \mathbf{u}_1)}{F(\mathbf{a}, \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2)} \mathbf{1}(\mathbf{a}, \mathbf{u}_1) + \frac{F(\mathbf{a}, \mathbf{u}_2)}{F(\mathbf{a}, \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2)} \mathbf{1}(\mathbf{a}, \mathbf{u}_2) \in \mathbb{I}_{\mathbf{a}}. \tag{40}$$

This is equivalent to

$$F\left(\mathbf{a}, \frac{F(\mathbf{a}, \mathbf{u}_1)}{F(\mathbf{a}, \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2)} \mathbf{1}(\mathbf{a}, \mathbf{u}_1) + \frac{F(\mathbf{a}, \mathbf{u}_2)}{F(\mathbf{a}, \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2)} \mathbf{1}(\mathbf{a}, \mathbf{u}_2)\right) \leq 1.$$

But the lefthand side expression equals

$$\frac{F(\mathbf{a}, \mathbf{u}_1 + \mathbf{u}_2)}{F(\mathbf{a}, \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2)},$$

whence we see that $F(\mathbf{a}, \mathbf{u})$ is convex. Conversely, if the expression above is ≤ 1, then (40) holds. Since it holds for any $\mathbf{u}_1, \mathbf{u}_2$, it also holds for $\lambda \mathbf{u}_1, (1-\lambda) \mathbf{u}_2$ for $0 \leq \lambda \leq 1$. But, as λ changes from 0 to 1, the expression

$$\frac{F(\mathbf{a}, \lambda \mathbf{u}_1)}{F(\mathbf{a}, \lambda \mathbf{u}_1) + F(\mathbf{a}, \mathbf{u}_2)} = \frac{\lambda F(\mathbf{a}, \mathbf{u}_1)}{\lambda F(\mathbf{a}, \lambda \mathbf{u}_1) + (1-\lambda) F(\mathbf{a}, (\mathbf{1}-\lambda) \mathbf{u}_2)}$$

runs through all values from 0 to 1 too. Since

$$\mathbf{1}(\mathbf{a}, \lambda \mathbf{u}_1) = \mathbf{1}(\mathbf{a}, \mathbf{u}_1), \mathbf{1}(\mathbf{a}, (1-\lambda) \mathbf{u}_2) = \mathbf{1}(\mathbf{a}, \mathbf{u}_2),$$

we have

$$\theta \mathbf{1}(\mathbf{a}, \mathbf{u}_1) + (1-\theta) \mathbf{1}(\mathbf{a}, \mathbf{u}_2) \in \mathbb{I}_{\mathbf{a}},$$

for any $0 \leq \theta \leq 1$. This means that $\mathbb{I}_{\mathbf{a}}$ is convex, and we have proved

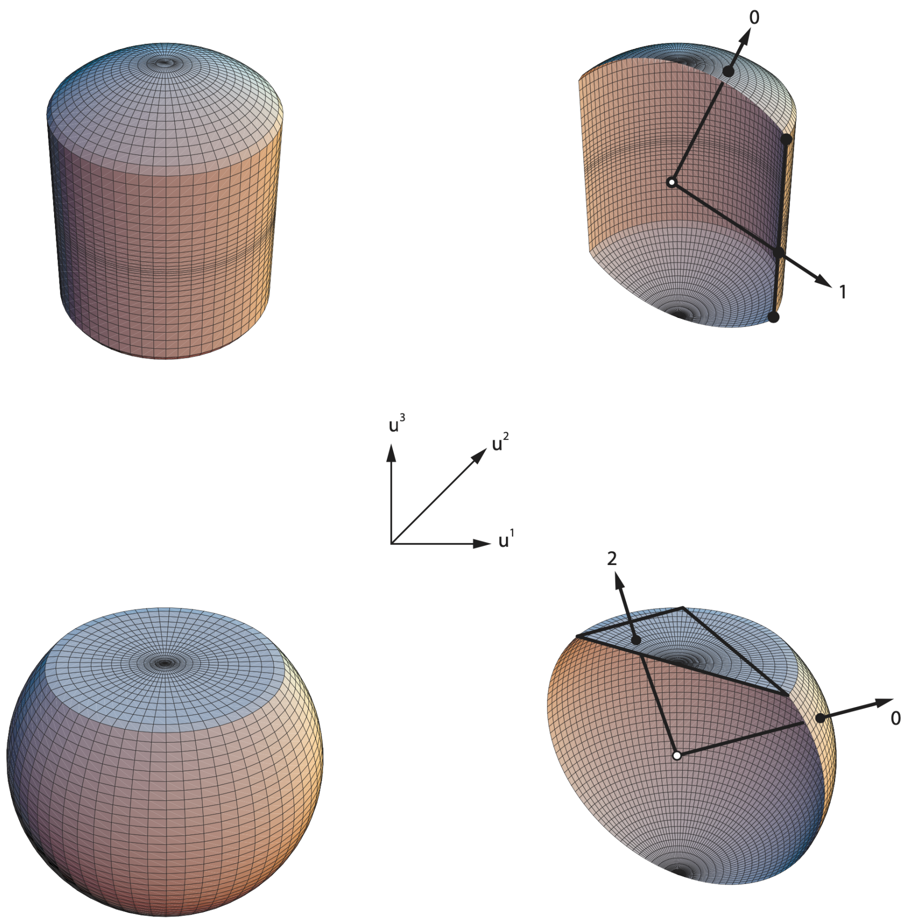

Figure 26: The convex hulls of the indicatrices shown in Figure 24. The degree of flatness of codirectional radius-vectors remains unchanged.

Theorem 49. *$F(\mathbf{a},\mathbf{u})$ is convex if and only if the body of the associated indicatrix $\mathbb{I}_{\mathbf{a}}$ at any point $\mathbf{a}$ is convex.*

From this and Theorem 47 we immediately have

Corollary 50. *For every submetric function F,*
(i) the corresponding minimal submetric function $\widehat{F}$ is convex,
(ii) $F \equiv \widehat{F}$ if and only if F is convex.

We also have

Corollary 51. *If a submetric function F is convex, then $\{\mathbf{u}\}$ is a minimizing vector chain for any line element $(\mathbf{a},\mathbf{u}) \in \mathbb{T}$.*

This follows from $F(\mathbf{a},\mathbf{u}) = \widehat{F}(\mathbf{a},\mathbf{u})$.

Of course, if F is convex, the following are also minimizing vector chains for $\mathbf{u} \in \mathbb{U}^n$: $\left\{\frac{1}{2}\mathbf{u}, \frac{1}{2}\mathbf{u}\right\}$, $\left\{\frac{1}{3}\mathbf{u}, \frac{2}{3}\mathbf{u}\right\}$,$\left\{\frac{1}{n}\mathbf{u}, \ldots, \frac{1}{n}\mathbf{u}\right\}$, etc. Moreover, if F is not strictly convex (i.e., the inequality in Definition 48 may be equality for some $\mathbf{u}_1, \mathbf{u}_2$), there may very well be minimizing chains involving vectors that are not collinear with $\mathbf{u}$.

We have now arrived at one of the central theorems in the theory.

Theorem 52. *The distance $G(\mathbf{x},\mathbf{x}+\mathbf{u}s)$ is differentiable at $s = 0+$ for any $(\mathbf{x},\mathbf{u}) \in \mathbb{T}$, and*

$$\left.\frac{\mathrm{d}G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{\mathrm{d}s+}\right|_{s=0} = \lim_{s\to 0+} \frac{G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{s} = \widehat{F}(\mathbf{x},\mathbf{u}).$$

See Appendix for a proof.

An important corollary to this theorem is as follows. Let $\mathbf{f}|\,[a,b]$ be a continuously differentiable path. Consider

$$\frac{G\mathbf{f}(t)\,\mathbf{f}(\tau)}{\widehat{F}(\mathbf{f}(t),\mathbf{f}(\tau)-\mathbf{f}(t))}, t<\tau.$$

By presenting it

$$\frac{G\mathbf{x}(t)\left(\mathbf{f}(t)+\frac{\mathbf{f}(\tau)-\mathbf{f}(t)}{\tau-t}(\tau-t)\right)}{\widehat{F}\left(\mathbf{f}(t),\frac{\mathbf{f}(\tau)-\mathbf{x}(t)}{\tau-t}\mathbf{f}(\tau-t)\right)} = \frac{G\mathbf{x}(t)\left(\mathbf{f}(t)+\dot{\mathbf{f}}(\theta)(\tau-t)\right)}{\widehat{F}\left(\mathbf{f}(t),\dot{\mathbf{f}}(\theta)\,\mathbf{f}(\tau-t)\right)},$$

with $t \le \theta \le \tau$, we see that if $\tau - t \to 0+$ on $[a,b]$, the ratio tends to 1 (by Theorem 52 and because all functions involved are uniformly continuous on $[a,b]$). This establishes

Corollary 53. *For any smooth path $\mathbf{f}|\,[a,b]$ and $[t,\tau] \subset [a,b]$,*

$$\lim_{\tau-t\to 0+} \frac{G\mathbf{f}(t)\,\mathbf{f}(\tau)}{\widehat{F}(\mathbf{f}(t),\mathbf{f}(\tau)-\mathbf{f}(t))} = 1.$$

We are ready now to formulate the standard differential-geometric computation of the length of a continuously differentiable path by integration of the submetric function applied to its points and tangents.

Theorem 54. *For any continuously differentiable path* $\mathbf{f}|\,[a,b]$*,*

$$D\mathbf{f}|\,[a,b]=\int_a^b \widehat{F}\left(\mathbf{f}\left(t\right),\dot{\mathbf{f}}\left(t\right)\right)\mathrm{d}t.$$

Indeed, by definition,

$$D\mathbf{f}|\,[a,b]=\lim_{\delta\mu\to 0}\sum G\mathbf{f}\left(t_i\right)\mathbf{f}\left(t_{i+1}\right)$$

across all nets $\mu=\{...,t_i,t_{i+1}...\}$ partitioning $[a,b]$. This limit can be presented as

$$\lim_{\delta\mu\to 0}\sum \widehat{F}\left(\mathbf{f}\left(t_i\right),\mathbf{f}\left(t_{i+1}\right)-\mathbf{f}\left(t_i\right)\right)\frac{G\mathbf{f}\left(t_i\right)\mathbf{f}\left(t_{i+1}\right)}{\widehat{F}\left(\mathbf{f}\left(t_i\right),\mathbf{f}\left(t_{i+1}\right)-\mathbf{f}\left(t_i\right)\right)}.$$

By Corollary 53,

$$\lim_{\delta\mu\to 0}\frac{G\mathbf{f}\left(t_i\right)\mathbf{f}\left(t_{i+1}\right)}{\widehat{F}\left(\mathbf{f}\left(t_i\right),\mathbf{f}\left(t_{i+1}\right)-\mathbf{f}\left(t_i\right)\right)}=1.$$

Then

$$\begin{aligned}D\mathbf{f}|\,[a,b]&=\lim_{\delta\mu\to 0}\sum \widehat{F}\left(\mathbf{f}\left(t_i\right),\mathbf{f}\left(t_{i+1}\right)-\mathbf{f}\left(t_i\right)\right)\\&=\lim_{\delta\mu\to 0}\sum \widehat{F}\left(\mathbf{f}\left(t_i\right),\frac{\mathbf{f}\left(t_{i+1}\right)-\mathbf{f}\left(t_i\right)}{t_{i+1}-t_i}\right)\left(t_{i+1}-t_i\right).\end{aligned}$$

But

$$\lim_{\delta\mu\to 0}\widehat{F}\left(\mathbf{f}\left(t_i\right),\frac{\mathbf{f}\left(t_{i+1}\right)-\mathbf{f}\left(t_i\right)}{t_{i+1}-t_i}\right)=\widehat{F}\left(\mathbf{f}\left(t\right),\dot{\mathbf{f}}\left(t\right)\right)$$

and $\widehat{F}\left(\mathbf{f}\left(t\right),\dot{\mathbf{f}}\left(t\right)\right)$ is uniformly continuous on $[a,b]$. Hence

$$D\mathbf{f}|\,[a,b]=\lim_{\delta\mu\to 0}\sum \widehat{F}\left(\mathbf{f}\left(t_i\right),\dot{\mathbf{f}}\left(t_i\right)\right)\left(t_{i+1}-t_i\right)=\int_a^b \widehat{F}\left(\mathbf{f}\left(t\right),\dot{\mathbf{f}}\left(t\right)\right)\mathrm{d}t,$$

completing the proof.

Since

$$\lim_{\tau-t\to 0+}\frac{\int_t^\tau \widehat{F}\left(\mathbf{f}\left(x\right),\dot{\mathbf{f}}\left(x\right)\right)\mathrm{d}x}{\widehat{F}\left(\mathbf{f}\left(t\right),\frac{\mathbf{f}(\tau)-\mathbf{f}(t)}{\tau-t}\right)\left(\tau-t\right)}=1,$$

we also have

Corollary 55. *For any continuously differentiable path* $\mathbf{f}|\,[a,b]$*, and* $[t,\tau]\subset[a,b]$*,*

$$\lim_{\tau-t\to 0+}\frac{G\mathbf{f}\left(t\right)\mathbf{f}\left(\tau\right)}{D\mathbf{f}|\,[t,\tau]}=1.$$

8.7 Continuously differentiable paths and intrinsic metric G

Before proceeding, we need an auxiliary observation. The space $(\mathfrak{S}, E)$ being open, each point $\mathbf{p}$ in $\mathfrak{S}$ can be enclosed in a compact Euclidean ball

$$\mathfrak{B}(\mathbf{p}, r) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{p}| \leq r\} \subseteq \mathfrak{S},$$

and we can associate with any $\mathbf{p}$ the ball $\mathfrak{B}(\mathbf{p}, r)$ with the supremal value of $r_{\text{sup}}(\mathbf{p})$ (including ∞). The observation is that, given any compact subset $\mathfrak{s}$ of $\mathfrak{S}$,

$$\inf_{\mathbf{p} \in \mathfrak{s}} r_{\text{sup}}(\mathbf{p}) = \min_{\mathbf{p} \in \mathfrak{s}} r_{\text{sup}}(\mathbf{p}) > 0.$$

A *straight line segment* is defined as

$$\mathbf{s}(x) = \mathbf{a} + \mathbf{u}x, \qquad x \in [a, b], (\mathbf{a}, \mathbf{u}) \in \mathbb{T}.$$

If $\mathbf{x}$ and $\mathbf{y}$ are within any ball $\mathfrak{B}(\mathbf{p}, r)$ they can be connected by the straight line segment

$$\mathbf{s}(x) = \mathbf{x} + \frac{\mathbf{y} - \mathbf{x}}{b - a}(x - a), \qquad x \in [a, b].$$

Concatenations of straight line segments forms piecewise linear paths, about which we have the following result.

Theorem 56. *For every path* $\mathbf{h}|[a, b]$ *connecting* $\mathbf{a}$ *to* $\mathbf{b}$ *one can find a piecewise linear path from* $\mathbf{a}$ *to* $\mathbf{b}$ *which is arbitrarily close to* $\mathbf{h}|[a, b]$ *pointwise and in its length.*

See Appendix for a proof.

The straight-line segments are not indispensable in such an approximation. In fact, we can use the following "corner-rounding" procedure to replace any piecewise linear path with a continuously differentiable path. It is illustrated in Figure 27.

Let two adjacent straight line segments be presented as

$$\mathbf{p}(t) = \begin{cases} \mathbf{a} + \overline{\mathbf{u}}_1 t & \text{if } t \in [-a, 0] \\ \mathbf{a} + \overline{\mathbf{u}}_2 t & \text{if } t \in [0, b] \end{cases},$$

with $a, b > 0$. On a small interval $[-s, s]$,

$$D\mathbf{p}|[-s, s] = \int_{-s}^{0} \widehat{F}(\mathbf{a} + \overline{\mathbf{u}}_1 t, \overline{\mathbf{u}}_1)\,\mathrm{d}t + \int_{0}^{s} \widehat{F}(\mathbf{a} + \overline{\mathbf{u}}_2 t, \overline{\mathbf{u}}_2)\,\mathrm{d}t.$$

Corner-rounding consists in replacing $\mathbf{p}|[-s, s]$ with a continuously differentiable path

$$\mathbf{q}(t) = \mathbf{a} + \mathbf{u}(t)\,t, \quad t \in [-s, s], \tag{41}$$

such that

$$\begin{aligned} &\mathbf{u}(-s) = \overline{\mathbf{u}}_1, \mathbf{u}(s) = \overline{\mathbf{u}}_2 \\ &\dot{\mathbf{u}}(-s) = \dot{\mathbf{u}}(s) = \mathbf{0} \end{aligned} \tag{42}$$

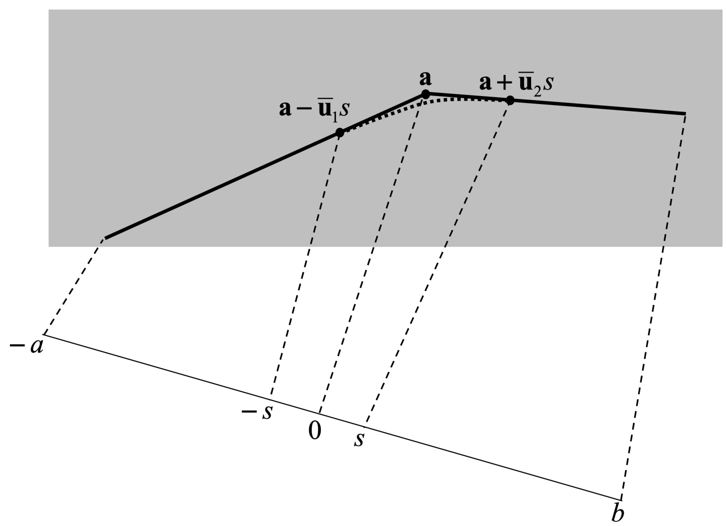

Figure 27: An illustration for the corner-rounding procedure. The piecewise linear path is shown as a mapping of the interval $[-a, b]$ into Euclidean plane (gray area). At 0 the two segments meet, and around this point they are replaced by the path shown by the dotted line of an arbitrarily close length.

and

$$\lim_{s\to 0+} D\mathbf{x}|\,[-s, s] = 0. \tag{43}$$

The requirements (42) ensure that the modified path $\mathbf{r}|\,[-a, b]$ defined by

$$\mathbf{r}\,(t) = \begin{cases} \mathbf{p}\,(t) & \text{if } t \notin [-s, s] \\ \mathbf{q}\,(t) & \text{if } t \in [-s, s] \end{cases}$$

is continuously differentiable. The requirement (43) ensures that the difference

$$|D\mathbf{p}|\,[-a, b] - D\mathbf{r}|\,[-a, b]|$$

can be made arbitrarily small by choosing s sufficiently small. One example of (41) is given by

$$\mathbf{u}\,(t) = \frac{\overline{\mathbf{u}}_1 + \overline{\mathbf{u}}_2}{2} + \frac{\left(\frac{t}{s}\right)^3 - 3\left(\frac{t}{s}\right)}{4}\left(\overline{\mathbf{u}}_1 - \overline{\mathbf{u}}_2\right).$$

We can now reformulate Theorem 56 as follows.

Theorem 57. *For every path* $\mathbf{h}|\,[a, b]$ *connecting* $\mathbf{a}$ *to* $\mathbf{b}$ *one can find a continuously differentiable path from* $\mathbf{a}$ *to* $\mathbf{b}$ *which is arbitrarily close to* $\mathbf{h}|\,[a, b]$ *pointwise and in its length.*

As an immediate consequence, we have the following.

Theorem 58. *If G in $(\mathfrak{S}, D)$ is an intrinsic metric, then, for any $\mathbf{a}, \mathbf{b}$ in $\mathfrak{S}$,*

$$G\mathbf{ab} = \inf \int_a^b \widehat{F}\left(\mathbf{f}(t), \dot{\mathbf{f}}(t)\right) \mathrm{d}t,$$

where the infimum is taken across all continuously differentiable paths (or piecewise continuously differentiable, if more convenient) connecting $\mathbf{a}$ to $\mathbf{b}$.

Recall that G is defined as intrinsic $G\mathbf{ab}$ is an infimum of the length of all paths connecting $\mathbf{a}$ to $\mathbf{b}$. This property is not derivable from the assumptions $\mathcal{E}1$ and $\mathcal{E}2$ we made about the relationship between $(\mathfrak{S}, D)$ and $(\mathfrak{S}, E)$. It should therefore be stipulated as an additional assumption or derived from other additional assumptions, e.g., that $(\mathfrak{S}, D)$ is a complete space with intermediate points.

9 Dissimilarity cumulation: Extensions and applications

In this section we give a few examples of extensions of the dissimilarity cumulation theory aimed at broadening the scope of its applicability.

9.1 Example 1: Observational sorites "paradox"

The issue of pairwise discrimination is the main application of Fechnerian Scaling and the original motivation for its development. As we know from Sections 1.4 and 1.5, it is a fundamental fact that two stimuli being compared must belong to distinct observation areas, say, one being on the left and the other on the right in visual field, or one being first and the other second in time. Without this one would not be able to speak, e.g., of a stimulus with value $\mathbf{x}$ being compared to a stimulus with the same value, because then we would simply have a single stimulus. Similarly, without the distinct observation areas there would be no operational meaning in distinguishing $(\mathbf{x}, \mathbf{y})$ from $(\mathbf{y}, \mathbf{x})$. Throughout this chapter the observation areas in our notation were implicit: e.g., we assumed that the stimulus written first in $(\mathbf{x}, \mathbf{y})$ belongs to the first observation area, or that $\mathbf{x}$ always denotes a stimulus in the first observation area. Here, however, we will need to indicate observation areas explicitly: $\mathbf{v}^{(o)}$ means a stimulus with value $\mathbf{v}$ in observation area o. If we assume that the observation areas are fixed, we can denote them 1 and 2, so that every value $\mathbf{v}$ may be part of the stimuli $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$. Note that with this notation any pair $\left\{\mathbf{x}^{(1)}, \mathbf{y}^{(2)}\right\}$ can be considered unordered, because $\left\{\mathbf{y}^{(2)}, \mathbf{x}^{(1)}\right\}$ represents the same pair.

There is an apparent "paradox" related to pairwise comparisons that seems so compelling that many describe it as a well-known empirical fact. Quoting from R. Duncan Luce (1956):

> It is certainly well known from psychophysics that if "preference" is taken to mean which of two weights a person believes to be heavier

> after hefting them, and if "adjacent" weights are properly chosen, say a gram difference in a total weight of many grams, then a subject will be indifferent between any two "adjacent" weights. If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false.

In other words, one can have a sequence of weights in which every two successive weights subjectively match each other, but the first and the last one do not. In philosophy, this seemingly paradoxical situation is referred to as *observational sorites*. The term "sorites" means "heap" in Greek, and the paradox is traced back to the Greek philosopher Eubulides (4th century BCE). In fact, Eubulides dealt with another form of the paradox, one in which stimuli are mapped into one of two categories one at a time. This form of sorites requires a different analysis. In our case, we have pairs of stimuli mapped into categories "match" or "do not match." The resolution of this paradox is based on two considerations:

1. The relationship "$\mathbf{x}^{(1)}$ matches $\mathbf{y}^{(2)}$" (or vice versa) is computed from an ensemble of responses rather than observed as an individual response. Individual responses to the same pair $\left\{\mathbf{x}^{(1)}, \mathbf{y}^{(2)}\right\}$ vary, and the pair can only be associated to a probability of a response, say,

$$\psi^{*}\left(\mathbf{x}^{(1)}, \mathbf{y}^{(2)}\right) = \Pr\left[\mathbf{x}^{(1)} \text{ is judged to be different from } \mathbf{y}^{(2)}\right]. \qquad (44)$$

2. Stimuli $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ have the same value $\mathbf{v}$ but they are different. To repeat the same stimulus, it should be presented in the same observation area in addition to having the same value.

Applying these considerations to the above quotation from Luce, let

$$w_1, w_2, w_3, w_4, \ldots, w_n$$

be the sequence of weights about which Luce (and many others) think as one in which w_{k-1} and w_k match (for $k = 2, \ldots, n$) but w_1 and w_n do not. Such a sequence is called a *(comparative) soritical sequence*. Let us, however, assign the weights to observation areas, as they should be. One can, e.g., place one weight in an observer's left hand and another weight in her right hand to be hefted simultaneously, in which case $w^{(1)} = w^{(left)}$ and $w^{(2)} = w^{(right)}$. Or the observer can heft one weight first and the other weight after a short interval, in which case $w^{(1)} = w^{(first)}$ and $w^{(2)} = w^{(second)}$. Whichever the case, since two adjacent weights in our sequence are to be compared, they should belong to different observation areas,

$$w_1^{(1)}, w_2^{(2)}, w_3^{(1)}, w_4^{(2)}, \ldots, w_n^{(2)}.$$

The last and the first stimuli also should belong to different observation areas if they are to be compared, so n must be an even number. Assuming that the discrimination here is of the "greater-less" variety, we have a function

$$\gamma\left(x^{(1)}, y^{(2)}\right) = \Pr\left[x^{(1)} \text{ is judged to be lighter than } y^{(2)}\right],$$

and the match is determined by

$$\gamma\left(x^{(1)}, y^{(2)}\right) = \frac{1}{2}.$$

So we have

$$\gamma\left(w_1^{(1)}, w_2^{(2)}\right) = \gamma\left(w_2^{(2)}, w_3^{(1)}\right) = \gamma\left(w_3^{(1)}, w_4^{(2)}\right) = \ldots = \gamma\left(w_{n-1}^{(1)}, w_n^{(2)}\right) = \frac{1}{2}.$$

It is not obvious now that we can have $w_1 < w_2 < w_3 < w_4 < \ldots < w_n$. In fact, if we accept the usual model of a psychometric function γ, as in Figure 1 and 2, w_k is uniquely determined as a match for w_{k-1}, and, moreover,

$$\begin{array}{l} w_1^{(1)} = w_3^{(1)} = \ldots = w_{n-1}^{(1)}, \\ w_2^{(2)} = w_4^{(2)} = \ldots = w_n^{(2)}. \end{array}$$

The sequence clearly is not soritical, because $w_1^{(1)}$ and $w_n^{(2)}$ (for an even n) necessarily match.

Generalizing, if one explicitly considers observation areas as part of stimuli's identity, the idea of soritical sequences becomes unfounded. If one further accepts the principles stipulated in Section 1.4, enabling one to construct a canonical space $(\mathfrak{S}, D)$, then soritical sequences become impossible. Essentially we are dealing with the problem of a reasonable definition of a match (PSE). We outline below an axiomatic scheme that defines stimulus spaces in which soritical sequences are impossible.

Not to be constrained to just two fixed observation areas, we consider a union of stimulus spaces indexed by observation areas:

$$\mathcal{S} = \bigcup_{\alpha\in\Omega} \mathfrak{S}_\omega^*.$$

We indicate the elements of $\mathfrak{S}_\omega^*$ by the corresponding superscript, say $\mathbf{x}^{(\omega)}$. The set $\mathcal{S}$ is endowed with a binary relation $\mathbf{x}^{(\alpha)}\mathrm{M}\mathbf{y}^{(\beta)}$ (read as "$\mathbf{x}$ in α is matched by $\mathbf{y}$ in β"). The most basic property of M is

$$\mathbf{x}^{(\alpha)}\mathrm{M}\mathbf{y}^{(\beta)} \Longrightarrow \alpha \neq \beta. \tag{45}$$

Definition 59. Given a space $(\mathcal{S}, \mathrm{M})$, we call a sequence $\mathbf{x}_1^{(\omega_1)}, \ldots, \mathbf{x}_n^{(\omega_n)}$ *well-matched* if

$$\omega_i \neq \omega_j \Longrightarrow \mathbf{x}_i^{(\omega_i)}\mathrm{M}\mathbf{x}_j^{(\omega_j)} \tag{46}$$

for all $i, j \in \{1, \ldots n\}$. The stimulus space $(\mathcal{S}, \mathrm{M})$ is *well-matched* if, for any sequence $\alpha, \beta, \gamma \in \Omega$ and any $\mathbf{a}^{(\alpha)} \in \mathcal{S}$, there is a well-matched sequence $\mathbf{a}^{(\alpha)}, \mathbf{b}^{(\beta)}, \mathbf{c}^{(\gamma)}$.

In particular, in a well-matched space, for any $\mathbf{a}^{(\alpha)}$ and any $\beta \in \Omega$, one can find a $\mathbf{b}^{(\beta)} \in \mathcal{S}$ such that $\mathbf{a}^{(\alpha)}\mathrm{M}\mathbf{b}^{(\beta)}$ and $\mathbf{b}^{(\beta)}\mathrm{M}\mathbf{a}^{(\alpha)}$.

Definition 60. Two stimuli $\mathbf{a}^{(\omega)}, \mathbf{b}^{(\omega)}$ in $(\mathcal{S}, \mathrm{M})$ are called *equivalent*, in symbols $\mathbf{a}^{(\omega)}\mathrm{E}\mathbf{b}^{(\omega)}$, if for any $\mathbf{c}^{(\iota)} \in \mathcal{S}$,

$$\mathbf{c}^{(\iota)}\mathrm{M}\mathbf{a}^{(\omega)} \iff \mathbf{c}^{(\iota)}\mathrm{M}\mathbf{b}^{(\omega)}. \tag{47}$$

$(\mathcal{S}, \mathrm{M})$ is a *regular space* if, for any $\mathbf{a}^{(\omega)}, \mathbf{b}^{(\omega)}, \mathbf{c}^{(\omega')} \in \mathcal{S}$ with $\omega \neq \omega'$,

$$\mathbf{a}^{(\omega)}\mathrm{M}\mathbf{c}^{(\omega')} \wedge \mathbf{b}^{(\omega)}\mathrm{M}\mathbf{c}^{(\omega')} \Longrightarrow \mathbf{a}^{(\omega)}\mathrm{E}\mathbf{b}^{(\omega)}. \tag{48}$$

This is a generalization of the notion of psychological equality introduced in Section 1.4.

Definition 61. Given a space $(\mathcal{S}, \mathrm{M})$, a sequence $\mathbf{x}_1^{(\omega_1)}, \ldots, \mathbf{x}_n^{(\omega_n)}$ with $\mathbf{x}_i^{(\omega_i)} \in \mathcal{S}$ for $i = 1, \ldots, n$, is called *soritical* if

1. $\mathbf{x}_i^{(\omega_i)}\mathrm{M}\mathbf{x}_{i+1}^{(\omega_{i+1})}$ for $i = 1, \ldots, n-1$,
2. $\omega_1 \neq \omega_n$,
3. but it is not true that $\mathbf{x}_1^{(\omega_1)}\mathrm{M}\mathbf{x}_n^{(\omega_n)}$.

Well-matchedness and regularity can be shown to be independent properties. Our interest is in the spaces that are both regular and well-matched. It can be proved that

Theorem 62. *In a regular well-matched space it is impossible to form a soritical sequence.*

9.2 Example 2: Thurstonian-type representations

Consider now the special case of the regular well-matched spaces, when the matching (PSE) relation is defined through minima of a same-different discrimination probability function $\psi^\star : \mathfrak{S}_1^\star \times \mathfrak{S}_2^\star \to [0,1]$ in (44). The issue discussed in this example is $\psi^\star$ can be "explained" by a *random-utility* (or *Thurstonian*) *model*, according to which each stimulus is mapped into a random variable in some perceptual space, and the decision "same" or "different" is determined by the values of these random variables for the stimuli $\mathbf{x}^{(1)}$ and $\mathbf{y}^{(2)}$.

Let us assume that both $\mathfrak{S}_1^\star \times \mathfrak{S}_2^\star$ are open connected regions of $\mathbb{R}^n$, and we present the property of Regular Minimality (14) in the following special form: there is a homeomorphism $\mathbf{h} : \mathfrak{S}_1^\star \to \mathfrak{S}_2^\star$ (a continuous function with a continuous $\mathbf{h}^{-1}$) such that

$$\begin{cases} \arg\min_{\mathbf{y}} \psi^\star(\mathbf{x}, \mathbf{y}) = \mathbf{h}(\mathbf{x}), \\ \arg\min_{\mathbf{x}} \psi^\star(\mathbf{x}, \mathbf{y}) = \mathbf{h}^{-1}(\mathbf{y}). \end{cases} \tag{49}$$

Here we once again drop the superscripts in $\mathbf{x}^{(1)}$ and $\mathbf{y}^{(2)}$. The function $\arg\min_{a_i} f(a_1, \ldots, a_n)$ indicate the value of the argument a_i at which f reaches its minimum (at fixed values of the remaining arguments). Empirical studies show that generally the *minimum-level function* $\psi^\star(\mathbf{x},\mathbf{h}(\mathbf{x}))$ *varies with* $\mathbf{x}$,

$$\psi^\star(\mathbf{x},\mathbf{h}(\mathbf{x})) \neq const. \tag{50}$$

Equivalently written,

$$\psi^\star\left(\mathbf{h^{-1}}(\mathbf{y}),\mathbf{y}\right) \neq const.$$

We call this property *nonconstant self-dissimilarity* of $\psi^\star$.

Rather than using Regular Minimality (49) to bring the stimulus space to a canonical form, we will consider the following construction. Consider a point $(\mathbf{p},\mathbf{h}(\mathbf{p}))$ in $\mathfrak{S}_1^\star \times \mathfrak{S}_2^\star$ and a direction $\mathbf{u}$ in

$$\mathbb{U}^n = \{\mathbf{u} = \mathbf{x} - \mathbf{p} : \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{p}\}.$$

For $(x,y) \in [-a,a]^2$, where a is a small positive number, the function

$$\lambda(x,y) = \psi^\star(\mathbf{p}+\mathbf{u}x, \mathbf{h}(\mathbf{p}+\mathbf{u}y))$$

is called a *patch* of the function $\psi^\star(\mathbf{x},\mathbf{y})$ at $(\mathbf{p},\mathbf{h}(\mathbf{p}))$. Note that the $(\mathbf{p},\mathbf{h}(\mathbf{p}))$ itself corresponds to $(x=0,y=0)$, and the graph of the PSE function $(\mathbf{x},\mathbf{h}(\mathbf{x}))$ in the vicinity of $\mathbf{x}=\mathbf{p}$ is mapped into the diagonal $\{(x,y): x=y\}$. We have therefore the following "patch-wise" version of the Regular Minimality and nonconstant self-dissimilarity:

$$\begin{cases} \arg\min_y \lambda(x,y) = x, \\ \arg\min_x \lambda(x,y) = y, \end{cases}$$

and

$$\lambda(x,x) \neq const$$

for $(x,y) \in [-a,a]^2$. We will call a patch *typical* if $\lambda(x,x)$ is nonconstant for all sufficiently small positive a. Figure 28 illustrates the notion.

In a Thurstonian-type model (called so in honor of Leon Thurstone who introduced such models in psychology in the 1920s), there is some internal space of images P, and each stimulus $\mathbf{x} \in \mathfrak{S}_1^\star$ (hence also any x representing $\mathbf{x}$ in a patch) is mapped into a random variable A with values in P, and, similarly, $\mathbf{y} \in \mathfrak{S}_2^\star$ (hence also any y representing $\mathbf{y}$ in a patch) is mapped into a random variables B with values in P. We will denote these random variables $A(\mathbf{x})$ and $B(\mathbf{y})$, and their sets of possible values a and b, respectively. We will consider first the case when $A(\mathbf{x})$ and $B(\mathbf{y})$ are stochastically independent. According to the model, there is a function

$$d : \mathsf{a} \times \mathsf{b} \to \{\text{same, different}\},$$

determining which response will be given in a given presentation of the stimuli. In complete generality, with no constraints imposed, such a model is not falsifiable.

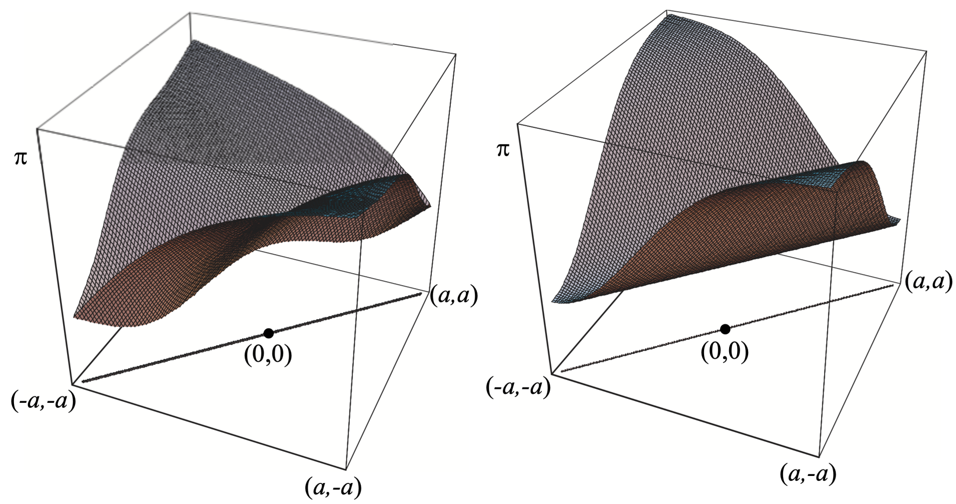

Figure 28: A typical patch (left) and an atypical patch (right) on a small square $[-a, a]^2$.

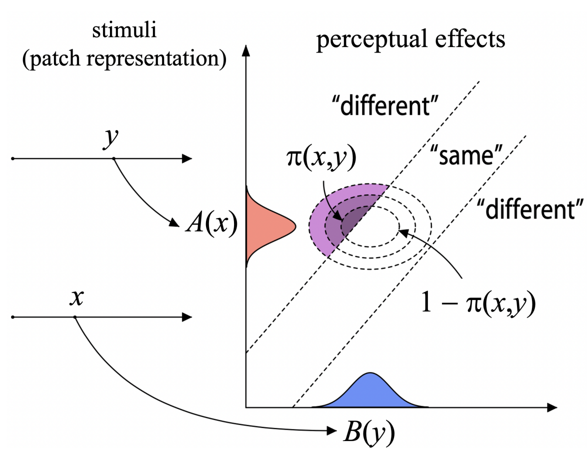

Figure 29: A schematic representation of a Thurstonian-type model. The stimuli are represented by their patch variables x and y, and their perceptual effects are points in an interval of reals. The response "same" is given if and only if both random variables $A(x)$ and $B(y)$ fall within the area between the two dashed lines.

Theorem 63. *Any psychometric function* $\psi^{\star} : \mathfrak{S}_1^{\star} \times \mathfrak{S}_2^{\star} \to [0,1]$ *can be generated by a Thurstonian-type model with stochastically independent random variables* $A(\mathbf{x})$ *and* $B(\mathbf{y})$.

This is not, however, very interesting, because one normally would want to deal only with sufficiently "well-behaved" Thurstonian-type models. The intuition here is that, as $\mathbf{x}$ and $\mathbf{y}$ continuously change, the random variables $A(\mathbf{x})$ and $B(\mathbf{y})$ change sufficiently smoothly. Consider, e.g., Figure 29, depicting a common way of modeling same-different comparisons. If the patch variables x and y change by a small amount, on should expect that the shapes of the probability density functions not change in an abrupt way. To formalize this intuition, denote, for any A-measurable set a in the perceptual space,

$$A_x(\mathsf{a}) = \Pr[A(x) \in \mathsf{a}],$$

and analogously, for any B-measurable set b in the perceptual space,

$$B_y(\mathsf{b}) = \Pr[B(y) \in \mathsf{b}].$$

Definition 64. Given a patch $\lambda(x,y)$, a Thurstonian type model generating it is said to be *well-behaved* if, for every A-measurable set a and B-measurable set b, the left-hand and right-hand derivatives

$$\frac{\mathrm{d}A_x(\mathsf{a})}{\mathrm{d}x\pm}, \frac{\mathrm{d}B_y(\mathsf{b})}{\mathrm{d}y\pm}$$

exist, and are bounded across all measurable sets.

The latter means that there is a constant c such that

$$\left|\frac{\mathrm{d}A_x(\mathsf{a})}{\mathrm{d}x\pm}\right| < c, \left|\frac{\mathrm{d}B_y(\mathsf{b})}{\mathrm{d}y\pm}\right| < c$$

for all measurable a and b. The "textbook" distributions (such as normal, Weibull, etc.) with parameters depending on x and y in a piecewise differentiable way will always satisfy this definition.

Definition 65. A patch $\lambda(x,y)$ is called *near-smooth* if he left-hand and right-hand derivatives

$$\frac{\partial\lambda(x,y)}{\partial x\pm}$$

exist and are continuous in y; and similarly,

$$\frac{\partial\lambda(x,y)}{\partial y\pm}$$

exist and are continuous in x.

It turns out that, perhaps not surprisingly,

Theorem 66. *A well-behaved Thurstonian representation can only generate near-smooth patches.*

A critical point in the development is created by the following fact.

Theorem 67. *No near-smooth patch can be typical, i.e. satisfy simultaneously the Regular Minimality and nonconstant self-dissimilarity properties.*

This means that for Thurstonian-type modeling of discrimination probabilities one cannot use well-behaved models, which in turn means the models should be quite complex mathematically (or else one should reject either Regular Minimality or nonconstant self-dissimilarity). With appropriate modifications of the definitions, this conclusion has been extended to Thurstonian models with *stochastically interdependent* (but *selectively influenced*) random variables, and to Thurstonian models in which the mapping of perceptual effects into responses is probabilistic too.

9.3 Example 3: Universality of corrections for violations of the triangle inequality.

In Section 6 we described the Floyd-Warshall algorithm for finite stimulus spaces. It turns out that it can be extended to arbitrary sets, generally infinite and not necessarily discrete. This is done by using the Axiom of Choice of the set theory to index all triangles in a stimulus set by *ordinals*. An *ordinal* is a set α such that each $\beta \in \alpha$ is a set, and $\beta \subseteq \alpha$. Thus,

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \ldots \tag{51}$$

are (finite) ordinals. For any two ordinals α and β, one and only one of the following is true: $\alpha = \beta$, $\alpha \in \beta$, or $\beta \in \alpha$. The ordinals are ordered in the following way: if $\alpha \in \beta$, we write $\alpha < \beta$; if either $\alpha \in \beta$ or $\alpha = \beta$, we write $\alpha \leq \beta$. For each ordinal α, $\alpha \cup \{\alpha\}$ is also an ordinal, called the *successor* of α and denoted $\alpha + 1$. There are two types of ordinals:

1. *successor* ordinals α, such that α is the successor of another ordinal,

2. *limit* ordinals, those that do not succeed other ordinals.

Thus, we can identify $\emptyset$ in (51) with 0, and identify $n \cup \{n\}$ with $n + 1$ for any ordinal identified with n. We have then that 0 is a limit ordinal, and each of $1, 2, 3, \ldots$ is a successor ordinal. The ordinal

$$\omega = \{0, 1, 2, 3, \ldots\}$$

is the smallest limit ordinal after 0, and the smallest infinite ordinal. The ordinals $\omega+1, \omega+2$, etc. are again successor ordinals, $\omega+\omega$ is a limit ordinal, and so on. Theorems involving ordinals are often proved by *transfinite induction*: if a certain property holds for 0, and it holds for any ordinal α whenever it holds for all ordinals $\beta < \alpha$, then this property holds for all ordinals. Similarly, definitions of a property of ordinals can be given by means of *transfinite recursion*: if it is defined for 0,and if, having defined it for all $\beta < \alpha$, we can use our definition to define it for α, then we define it for all ordinals. Thus, in Definition 11, the procedure of correcting dissimilarity functions for violations of the triangle inequalities is described by means of the usual mathematical induction. It can be replaced with transfinite recursion as follows. We index the triangles $\mathbf{xyz}$ with pairwise distinct elements by ordinals, so that for for every ordinal α there is an ordinal $\beta > \alpha$ indexing the same triangle. In other words, each triangle occurs an infinite number of times.

Definition 68. Define for each ordinal α a function $M^{(\alpha)} : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}$ as follows:
(i) $M^{(0)} \equiv D$;
(ii) for any successor ordinal $\alpha = \beta + 1$, and for all $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$,

$$M^{(\alpha)}\mathbf{ab} = \begin{cases} \min\{M^{(\beta)}\mathbf{ab}, M^{(\beta)}\mathbf{ax} + M^{(\beta)}\mathbf{xb}\} & \text{if } \mathbf{axb} \text{ is indexed by } \beta, \\ M^{(\beta)}\mathbf{ab} & \text{otherwise;} \end{cases}$$

(iii) if α is a limit ordinal, then, for all $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$,

$$M^{(\alpha)}\mathbf{ab} = \inf_{\beta<\alpha} M^{(\beta)}\mathbf{ab}.$$

It turns out that all results presented in Section 6 have their transfinite analogous in this generalization. In particular, "eventually" (i.e., at some ordinal α) the procedure is terminated with $M^{(\alpha)}$ coinciding with the quasimetric dissimilarity G, as defined in (18).

9.4 Example 4: Data Analysis

Multidimensional Scaling (MDS) and clustering are among the widely used tools of data analysis and data visualization. The departure point of MDS is a matrix

$$\{d_{ij} : i, j = 1, 2, \ldots, n\}$$

whose entries are values of a dissimilarity function on the set of objects $\mathfrak{S} = \{1, 2, \ldots, n\}$. This requires that, for all $i \neq j$,

$$d_{ii} = 0, \text{and } d_{ij} > 0.$$

If this is not the case, but Regular Minimality is satisfied, the matrix can be brought first to a canonical form, so that d_{ii} is the smallest value both in the ith raw and in the ith columns. Then one can replace d_{ij} with

$$\delta_{ij}^{(1)} = d_{ij} - d_{ii},$$

or with

$$\delta_{ij}^{(2)} = d_{ji} - d_{ii}.$$

The choice between the two corresponds to the choice between psychometric increments of the first and second kind. We know that this choice is immaterial in Fechnerian Scaling, but in MDS it is immaterial only if the matrix is symmetrical,

$$d_{ij} = d_{ji}.$$

If this is not the case, one usually uses in MDS some symmetrization procedure: e.g., one can replace each d_{ij} with

$$\delta_{ij} = d_{ij} + d_{ji} - d_{ii} - d_{jj} = \begin{cases} \delta_{ij}^{(1)} + \delta_{ji}^{(1)} \\ \delta_{ij}^{(2)} + \delta_{ji}^{(2)} \end{cases},$$

proposed by Roger Shepard in the 1950s for so-called *confusion matrices* (we will refer to it as *Shepard symmetrization*, SS). Following these or similar modifications, the matrix δ_{ij} can be viewed as a symmetric dissimilarity function.

If in addition the entries of the matrix satisfy the triangle inequality, the matrix represents a true metric on the set $\mathfrak{S} = \{1, 2, \ldots, n\}$. In such a case one can apply a procedure of *metric* MDS (mMDS), that consists in embedding the n elements of $\mathfrak{S}$ in an $\mathbb{R}^k$ so that the distances Δ_{ij} between the points are as close as possible to the corresponding δ_{ij}. The quality of approximation is usually estimated by a measure called *stress*, one variant of which is

$$\left(\frac{\sum_{i,j} \left(\Delta_{ij} - \delta_{ij}\right)^2}{\sum_{i,j} \delta_{ij}^2} \right)^{1/2}.$$

Since one of the goals of MDS is to help one to visualize the data, the distance in $\mathbb{R}^k$ is usually chosen to be Euclidean, and k chosen as small as possible (preferably 2 or 3).

However, in most applications δ_{ij} does not satisfy the triangle inequality, because of which MDS is used in its *nonmetric* version (nmMDS): here one seeks an embedding into a low-dimensional $\mathbb{R}^k$ in which the Euclidean distances match as close as possible not δ_{ij} but some monotonically increasing transformation of δ_{ij}. The stress measure then has the form

$$\left(\frac{\sum_{i,j} \left(\Delta_{ij} - g\left(\delta_{ij}\right)\right)^2}{\sum_{i,j} g\left(\delta_{ij}^2\right)} \right)^{1/2},$$

minimized across all possible monotone functions g.

Dissimilarity cumulation offers a different approach to the same problem, one that does not require any transformations. Once the original matrix d_{ij} is brought to a canonical form and replaced with $\delta_{ij}^{(1)}$ or $\delta_{ij}^{(2)}$, one computes from either of them the Fechnerian distances $\overleftrightarrow{G}_{ij}$. Since these are true distances, one can apply to them the metric version of MDS to seek a low-dimensional Euclidean embedding. For illustration, consider an experiment reported in Dzhafarov and Paramei (2010). Images of faces shown Figure 30 were presented

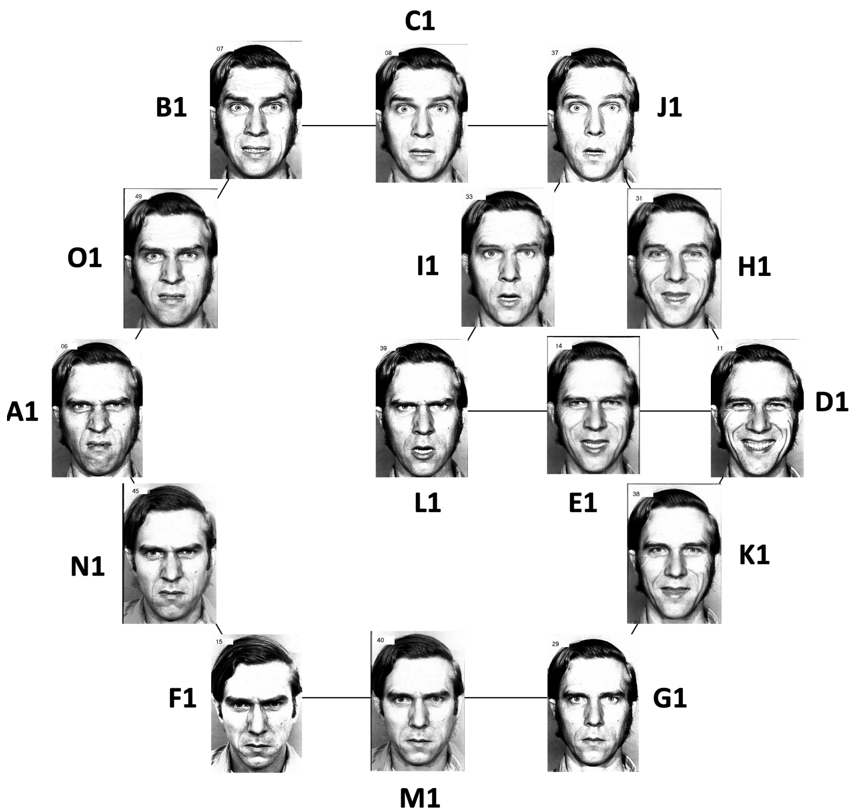

Figure 30: A sample of faces presented two at a time with the question whether they represent the same emotion or different emotions.

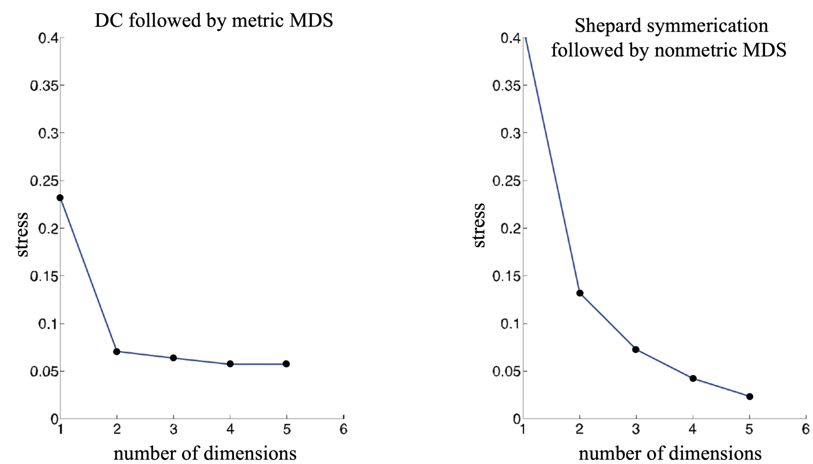

Figure 31: Scree plots of mMDS following Fechnerian Scaling (left) and nm-MDS following Shepard's symmetrization. The optimal number of dimensions is usually chosen as one at which the scree plot visibly decelerates (exhibits a "knee").

two at a time, and the observer was asked to determine whether they exhibited the same emotion or different emotions. The data d_{ij} were estimates of the probabilities of the response "different emotions." Figure 31 shows the value of stress as a function of k in the embedding space $\mathbb{R}^k$ (so-called *scree plots*). The comparison of the two procedures,

(DC-mMDS) metric MDS applied to the results of dissimilarity cumulation, and

(SS-nmMDS) non-metric MDS applied to Shepard-symmetrized data,

shows that the former seems to better identify the minimal dimensionality of the embedding space. In DC-mMDS, acceptably small value of stress is achieved at $k = 2$ or 3, and stress drops very slowly afterwards, whereas in SS-nmMDS, the deceleration of the scree plot is less pronounced. Having chosen, say, $k = 3$, the results of both procedures can be further subjected to cluster analysis, which groups the points in $\mathbb{R}^3$ into a designated number of *clusters* (the K-means procedure) or constructs their *dendrogram* (hierarchical cluster analysis). We do not discuss these procedure, as our goal is to merely point out that Fechnerian Scaling allows one to base all of them on true distances, without resorting to an unconstrained search of a monotone transformation. Moreover, the example in the next section describes an alternative to the dissimilarity cumulation approach that results in a cluster analysis representation.

There are two public-domain programs that perform MDS and clustering of the results of dissimilarity cumulation. One of them is the Matlab-based software package FSCAMDS (stands for Fechnerian Scaling – Clustering – and

– Multi$d$imensional Scaling), the other is the R-language package fechner (see the next section for references). These data-analytic programs have a variety of options of which we will mention the following.

It is sometimes the case, especially if the data are probabilities, or if they are sampled from a path-connected space, that large values of dissimilarity are unreliable, and the cumulation is to be restricted only to smaller values. The software packages allow one to set a value above which a dissimilarity $D\mathbf{ab}$ is replaced with infinity, removing thereby the link $\mathbf{ab}$ from the cumulation process (because it seeks the smallest cumulated value).

It is sometimes the case that Regular Minimality in the original data set is violated. The software packages allow one to choose between the following options:

1. to "doctor" the data by designating the pairs of PSE and, following the canonical transformation, to replace negative values of $d_{ij} - d_{ii}$ with zero;

2. to perform Fechnerian Scaling separately for the two observation areas, obtaining thereby $\overleftrightarrow{G_1}$ and $\overleftrightarrow{G_2}$ distances, not equal to each other.

The justifiability of the second option depends on one's position with respect to the empirical status of the Regular Minimality law. As mentioned in Section 1.5, Regular Minimality in this chapter is not taken as an empirical claim. Rather it has been part of the definition of the functions we have dealt with in our mathematical theory.

9.5 Example 5: Ultrametric Fechnerian Scaling

There is a more direct way to obtain a representation of dissimilarities by hierarchical clusters (dendogram or *rooted tree*). The basic idea consists in replacing "dissimilarity cumulation" by a "dissimilarity maximization" procedure.

Given a chain $\mathbf{X} = \mathbf{x}_1 \ldots \mathbf{x}_n$ and a binary (real-valued) function F, the notation $\Delta_F \mathbf{X}$ stands for

$$\max_{i=1,\ldots,n-1} F\mathbf{x_i x_{i+1}},$$

again with the obvious convention that the quantity is zero if n is 1 or 0. A dissimilarity function M on a finite set $\mathfrak{S}$ is called a *quasi-ultrametric* if it satisfies the *ultrametric inequality*,

$$\max\{M\mathbf{ab}, M\mathbf{bc}\} \geq M\mathbf{ac} \tag{52}$$

for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathfrak{S}$.

The ultrametric inequality is rather restrictive: it is equivalent to postulating that, for any triple of elements, two dissimilarities have to be equal and not smaller than the third.

Definition 69. Given a dissimilarity D on a finite set $\mathfrak{S}$, the *quasi-ultrametric* G^{∞} *induced by* D is defined as

$$G^{\infty}\mathbf{ab} = \min_{\mathbf{X}\in\mathcal{C}} \Delta_D \mathbf{aXb}, \tag{53}$$

for all $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$.

Thus, the value of $G^{\infty}\mathbf{ab}$ is obtained by taking the minimum, across all chains $\mathbf{X}$ from $\mathbf{a}$ to $\mathbf{b}$, of the maximum dissimilarity value of the chain. That G^{∞} is a quasi-ultrametric is easy to prove. A reasonable symmetrization procedure, yielding a metric is

$$G^{\infty^*}\mathbf{ab} = \max\{G^{\infty}\mathbf{ab}, G^{\infty}\mathbf{ba}\} \tag{54}$$

called the *overall Fechnerian ultrametric* on $\mathfrak{S}$.

The ultrametric inequality is often violated in empirical data. However, in analogy to recursive corrections for violations of the triangle inequality, it can be shown that a corresponding series of recursive corrections on the dissimilarity values for violations of the ultrametric inequality would yield the induced quasi-ultrametric distances. This is in contrast to applying the different standard hierarchical cluster algorithms (like single-link, combined-link, etc.) to one and the same data set: when violations exist, these algorithms will typically result in rather different ultrametrics.

One can consider procedures intermediate between cumulation and maximization of dissimilarities by defining, for any dissimilarity function D, the length of a chain $\mathbf{X} = \mathbf{x}_1, \ldots \mathbf{x}_n$ by

$$D\mathbf{X} = ((D\mathbf{x_1x_2})^k + \ldots + (D\mathbf{x}_{n-1}\mathbf{x}_n)^k)^{1/k}. \tag{55}$$

For $k \to \infty$ this would result in the ultrametric approach outlined above. For finite k, the procedure is generalizable to arbitrary dissimilarity spaces. This follows from the fact the use of (55) is equivalent to the use of the original dissimilarity cumulation procedure in which one, first, redefines D into D^k (which yields another dissimilarity function), and then redefines the quasimetric G induced by D^k into $G^{1/k}$ (which yields another quasimetric).

10 Related Literature

Fechner's original theory is presented in the *Elemente der Psychophysik* (Fechner, 1860), but important additions and clarifications can be found in a later book (Fechner, 1877), and in a paper written shortly before Fechner's death (Fechner, 1887). A detailed modern account of Fechner's original theory, especially the ways he derived his logarithmic psychophysical law, can be found in Dzhafarov and Colonius (2011). For related interpretations of Fechner's theory, see Pfanzagl (1962), Creelman (1967), Krantz (1971), and Falmagne (1971). A different interpretation of Fechner's theory, one that finds it lacking in mathematical coherence and with which we disagree, is presented in Luce & Edwards (1958) and Luce and Galanter (1963).

The theory of dissimilarity cumulation is presented in Dzhafarov and Colonius (2007) and elaborated in Dzhafarov (2008a). The geometric aspects of this theory are close to those of the distance and geodesics theory developed in Blumenthal (1953), Blumenthal and Menger (1970), and Busemann (2005). To better understand the topology and uniformity aspects of dissimilarity cumulation, one can consult, e.g., Kelly (1955) and Hocking and Young (1961). A proof of Theorem 10 can be found in Dzhafarov and Colonius (2007). A proof of Theorem 26 is presented in Dzhafarov (2008a).

For stimuli spaces defined on regions of $\mathbb{R}^n$, the mathematical theory essentially becomes a generalized form of Finsler geometry, as presented in Dzhafarov (2008b). A more detailed presentation, however, and one closer to this chapter, is found in earlier work (Dzhafarov & Colonius, 1999, 2001). This part of the theory has its precursors in Helmholtz (1891) and Schrödinger (1920/1970, 1926/1970), both of whom, in different ways, used Fechner's cumulation of infinitesimal differences to construct a Riemannian geometry (a special case of Finsler geometry) of color space.

In this chapter we have entirely omitted the important topic of invariance of length and distance under homeomorphic (for general path-connected spaces) and diffeomorphic (for $\mathbb{R}^n$-based spaces) transformations of space and reparametrizations of paths. These topics are discussed in Dzhafarov (2008b, c) and Dzhafarov & Colonius (2001). We have also ignored the difference between paths and arcs, discussed in detail in Dzhafarov (2008b).

Dissimilarity cumulation in discrete stimulus spaces is described in Dzhafarov and Colonius (2006a, c) and Dzhafarov (2010a). The generalization of the Floyd-Warshall algorithm to arbitrary spaces (Section 9.3) is described in Dzhafarov and Dzhafarov (2011).

The notion of separate observation area in stimulus comparisons, as well as the Regular Minimality law have been initially formulated in Dzhafarov (2002) and elaborated in Dzhafarov (2006b), Dzhafarov and Colonius (2006b), and Kujala and Dzhafarov (2008, 2009a). The application of the regularity and well-matchedness principles to the comparative sorites "paradox" is presented in Dzhafarov and Dzhafarov (2010, 2012), with a proof of Theorem 62, and in Dzhafarov and Perry (2014).

The application of these principles together with nonconstant self-dissimilarity to Thurstonian-type modeling is presented in Dzhafarov (2003a, b), where one can find proofs of the theorems in Section 9.2. This part of the theory has been generalized and greatly extended in Kujala and Dzhafarov (2008, 2009a, b).

For Multidimensional Scaling see, e.g., Borg and Groenen (1997). Clustering procedures, hierarchical and K-means, are described in standard textbooks of multivariate statistics, e.g. Everitt et. al. (2011). The ultrametric Fechnerian Scaling approach is presented in Colonius & Dzhafarov (2012).

The link and instructions to the R language software package fechner mentioned in Section 9.4 is available in Ünlü, Kiefer, and Dzhafarov (2009). The link and instructions to the software package FSCAMDS are available in Dzhafarov (2010b).

Appendix: Select proofs

Theorem 30. *$F(\mathbf{x},\mathbf{u})$ is well-defined for any $(\mathbf{x},\mathbf{u}) \in \mathbb{T} \cup \{(\mathbf{x},\mathbf{0}) : \mathbf{x} \in \mathfrak{S}\}$. It is positive for $\mathbf{u} \neq \mathbf{0}$, continuous in $(\mathbf{x},\mathbf{u})$, and Euler homogeneous in $\mathbf{u}$.*

Proof. We first show that $F(\mathbf{x},\overline{\mathbf{u}})$ is continuous in $(\mathbf{x},\overline{\mathbf{u}})$. By Assumptions $\mathcal{E}2$, for any $\varepsilon > 0$ there is a $\delta = \delta(\mathbf{x},\overline{\mathbf{u}},\varepsilon) > 0$ such that

$$\max\left\{|\mathbf{a}-\mathbf{x}|, |\mathbf{b}-\mathbf{x}|, \left|\overline{\mathbf{b}-\mathbf{a}}-\overline{\mathbf{u}}\right|\right\} < \delta(\mathbf{x},\overline{\mathbf{u}},\varepsilon) \Longrightarrow \left|\frac{D\mathbf{ab}}{|\mathbf{b}-\mathbf{a}|} - F(\mathbf{x},\overline{\mathbf{u}})\right| < \varepsilon.$$

Consider a sequence $(\mathbf{x}_n,\overline{\mathbf{u}}_n) \to (\mathbf{x},\overline{\mathbf{u}})$, and let $(\mathbf{a}_n,\mathbf{b}_n)$, $\mathbf{a}_n \neq \mathbf{b}_n$, be any sequence satisfying

$$\max\left\{|\mathbf{a}_n-\mathbf{x}_n|, |\mathbf{b}_n-\mathbf{x}_n|, \left|\overline{\mathbf{b}_n-\mathbf{a}_n}-\overline{\mathbf{u}}_n\right|\right\} < \min\left\{\delta\left(\mathbf{x}_n,\overline{\mathbf{u}}_n,\frac{1}{n}\right), \frac{1}{2}\delta(\mathbf{x},\overline{\mathbf{u}},\varepsilon)\right\}.$$

Clearly,

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{|\mathbf{b}_n-\mathbf{a}_n|} - F(\mathbf{x}_n,\overline{\mathbf{u}}_n) \to 0.$$

At the same time, for all sufficiently large n,

$$\max\{|\mathbf{x}_n-\mathbf{x}|, |\overline{\mathbf{u}}_n-\overline{\mathbf{u}}|\} < \frac{1}{2}\delta(\mathbf{x},\overline{\mathbf{u}},\varepsilon),$$

implying

$$\max\left\{|\mathbf{a}_n-\mathbf{x}|, |\mathbf{b}_n-\mathbf{x}|, \left|\overline{\mathbf{b}_n-\mathbf{a}_n}-\overline{\mathbf{u}}\right|\right\} < \delta(\mathbf{x},\overline{\mathbf{u}},\varepsilon).$$

But then

$$\left|\frac{D\mathbf{a}_n\mathbf{b}_n}{|\mathbf{b}_n-\mathbf{a}_n|} - F(\mathbf{x},\overline{\mathbf{u}})\right| < \varepsilon,$$

and, as ε can be chosen arbitrarily small, we have

$$\frac{D\mathbf{a}_n\mathbf{b}_n}{|\mathbf{b}_n-\mathbf{a}_n|} - F(\mathbf{x},\overline{\mathbf{u}}) \to 0.$$

The convergence

$$F(\mathbf{x}_n,\overline{\mathbf{u}}_n) \to F(\mathbf{x},\overline{\mathbf{u}})$$

follows, establishing the continuity of $F(\mathbf{x},\overline{\mathbf{u}})$. Now, for $\mathbf{u} \neq \mathbf{0}$, denoting $\mathbf{u} = |\mathbf{u}|\,\overline{\mathbf{u}}$,

$$F(\mathbf{x},\mathbf{u}) = \lim_{s\to 0+}\frac{D\mathbf{x}[\mathbf{x}+\mathbf{u}s]}{s} = |\mathbf{u}| \lim_{|\mathbf{u}|s\to 0+}\frac{D\mathbf{x}[\mathbf{x}+\overline{\mathbf{u}}\,|\mathbf{u}|\,s]}{|\mathbf{u}|\,s} = |\mathbf{u}|\,F(\mathbf{x},\overline{\mathbf{u}}).$$

It immediately follows that $F(\mathbf{x},\mathbf{u})$ exists, that it is positive and continuous, and that

$$F(\mathbf{x},\mathbf{u}) = |\mathbf{u}|\,F(\mathbf{x},\overline{\mathbf{u}}).$$

So, for $k > 0$,

$$F(\mathbf{x}, k\mathbf{u}) = k|\mathbf{u}| F(\mathbf{x}, \overline{\mathbf{u}}) = kF(\mathbf{x}, \mathbf{u}).$$

Finally, since any convergence of $(\mathbf{x}_n, \mathbf{u}_n) \to (\mathbf{x}, \mathbf{0})$ with $\mathbf{u}_n \neq \mathbf{0}$ can be presented as $(\mathbf{x}_n, |\mathbf{u}_n| \overline{\mathbf{u}}_n) \to (\mathbf{x}, \mathbf{0})$ with $|\mathbf{u}_n| \to 0$, we have

$$F(\mathbf{x}_n, \mathbf{u}_n) = |\mathbf{u}_n| F(\mathbf{x}_n, \overline{\mathbf{u}}_n) \to 0,$$

because within a small ball around $\mathbf{x}$ and on a compact set of unit vectors the function $F(\mathbf{x}_n, \overline{\mathbf{u}}_n)$ does not exceed some finite value. Thus $F(\mathbf{x}_n, \mathbf{u}_n)$ extends to $F(\mathbf{x}, \mathbf{0}) = 0$ by continuity. $\square$

Lemma 42. *For any $(\mathbf{a}, \mathbf{u}) \in \mathbb{T}$, the maximal production of $\mathbf{u}$ in $\mathbb{I}_\mathbf{a}$ can be presented as a convex combination of n (not necessarily distinct) radius-vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \delta\mathbb{I}_\mathbf{a}$.*

Proof. With no loss of generality, let $\mathbf{u} \in \delta\mathbb{I}_\mathbf{a}$, and let κ stand for $\kappa(\mathbf{a}, \mathbf{u})$. By Corollary 36, for some $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1} \in \mathbb{I}_\mathbf{a}$, the system of $n+1$ linear equations

$$\begin{cases} \kappa\mathbf{u} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_{n+1}\mathbf{v}_{n+1} \\ \lambda_1 + \ldots + \lambda_{n+1} = 1 \end{cases}$$

has a solution $\lambda_1, \ldots, \lambda_{n+1} \in [0, 1]$. Assume that $\lambda_1, \ldots, \lambda_{n+1}$ are all positive (if some of them are zero, the theorem's statement holds). If the determinant of the matrix of coefficients for this system were nonzero, then, for any ε, the modified system

$$\begin{cases} [\kappa + \varepsilon]\,\mathbf{u} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_{n+1}\mathbf{v}_{n+1} \\ \lambda_1 + \ldots + \lambda_{n+1} = 1 \end{cases}$$

would also have a solution $\lambda'_1, \ldots, \lambda'_{n+1}$, and choosing ε positive and sufficiently small, this solution (by continuity) would also satisfy $\lambda'_1 > 0, \ldots, \lambda'_{n+1} > 0$. But this would mean that $[\kappa + \varepsilon]\,\mathbf{u}$ belongs to the convex hull of $\mathbb{I}_\mathbf{a}$, which is impossible since $\kappa\mathbf{u}$ is the maximal production of $\mathbf{u}$. Hence

$$\det \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_{n+1} \\ 1 & \cdots & 1 \end{bmatrix} = 0,$$

where we treat $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ as n-element columns. But this means that, for some $\gamma_1, \ldots, \gamma_{n+1}$, not all zero,

$$\gamma_1 \begin{bmatrix} \mathbf{v}_1 \\ 1 \end{bmatrix} + \ldots + \gamma_{n+1} \begin{bmatrix} \mathbf{v}_{n+1} \\ 1 \end{bmatrix} = \mathbf{0},$$

which indicates the affine dependence of $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$. It follows from Lemma 35 that $\mathbf{u}$ can be presented as a convex combination of some $m < n+1$ (not necessarily distinct) nonzero vectors in $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1} \in \mathbb{I}_\mathbf{a}$. Let them be the first m vectors in the list, $\mathbf{v}_1, \ldots, \mathbf{v}_m$. We have now the system

$$\begin{cases} \kappa\mathbf{u} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_m\mathbf{v}_m \\ \lambda_1 + \ldots + \lambda_m = 1 \end{cases}$$

with a solution $\lambda_1 > 0, \ldots, \lambda_m > 0$ (zero values here would simply decrease m). Rewriting it as

$$\begin{cases} \kappa\mathbf{u} = \lambda_1 c_1 \widetilde{\mathbf{v}}_1 + \ldots + \lambda_m c_m \widetilde{\mathbf{v}}_m \\ \lambda_1 + \ldots + \lambda_m = 1 \end{cases},$$

where $\widetilde{\mathbf{v}}_i \in \delta\mathbb{I}_{\mathbf{a}}$ is codirectional with $\mathbf{v}_i$ $(i = 1, \ldots, m)$, it is clear by Lemma 34 that for κ to have a maximal possible value, all c_i should have maximal possible values. In $\mathbb{I}_{\mathbf{a}}$ these values are $c_1 = \ldots = c_m = 1$, that is, all vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are radius-vectors. This completes the proof. □

Theorem 46. *The minimal submetric function $\widehat{F}(\mathbf{a}, \mathbf{u})$ has all the properties of a submetric function: it is positive for $\mathbf{u} \neq \mathbf{0}$, Euler homogeneous, and continuous.*

Proof. We only prove the continuity, as the other properties follow trivially from the definition of $\widehat{F}$ and the analogous properties of F. Consider a sequence of line elements

$$(\mathbf{a}_k, \mathbf{u}_k) \to (\mathbf{a}, \mathbf{u}).$$

Let $(\mathbf{v}_1, ..., \mathbf{v}_n)$ be a minimizing chain for $(\mathbf{a}, \mathbf{u})$ (or a sequence of n zero vectors if $\mathbf{u} = \mathbf{0}$). For every k, consider the sequence $\mathbf{v}_1 + (\mathbf{u}_k - \mathbf{u}), \mathbf{v}_2, ..., \mathbf{v}_n$, which differs from the minimizing chain in the first element only. Its elements sum to $\mathbf{u}_k$, because of which

$$F(\mathbf{a}_k, \mathbf{v}_1 + (\mathbf{u}_k - \mathbf{u})) + F(\mathbf{a}_k, \mathbf{v}_2) + ... + F(\mathbf{a}_k, \mathbf{v}_n) \geq \widehat{F}(\mathbf{a}_k, \mathbf{u}_k).$$

At the same time, by continuity of F,

$$\begin{aligned} &F(\mathbf{a}_k, \mathbf{v}_1 + (\mathbf{u}_k - \mathbf{u})) + F(\mathbf{a}_k, \mathbf{v}_2) + ... + F(\mathbf{a}_k, \mathbf{v}_n) \\ &\to F(\mathbf{a}_k, \mathbf{v}_1) + F(\mathbf{a}_k, \mathbf{v}_2) + ... + F(\mathbf{a}_k, \mathbf{v}_n) = \widehat{F}(\mathbf{a}, \mathbf{u}), \end{aligned}$$

whence it follows that

$$\limsup_{k\to\infty} \widehat{F}(\mathbf{a}_k, \mathbf{u}_k) \leq \widehat{F}(\mathbf{a}, \mathbf{u}).$$

To prove that at the same time

$$\liminf_{k\to\infty} \widehat{F}(\mathbf{a}_k, \mathbf{u}_k) \geq \widehat{F}(\mathbf{a}, \mathbf{u}),$$

let $(\mathbf{v}_{1k}, ..., \mathbf{v}_{nk})$ be a minimizing chain for $(\mathbf{a}_k, \mathbf{u}_k)$, for every k, and consider the sequence $\mathbf{v}_{1k} + (\mathbf{u} - \mathbf{u}_k), \mathbf{v}_{2k}, ..., \mathbf{v}_{nk}$, which differs from the minimizing chain in the first element only. Its elements sum to $\mathbf{u}$, because of which

$$F(\mathbf{a}, \mathbf{v}_{1k} + (\mathbf{u} - \mathbf{u}_k)) + F(\mathbf{a}, \mathbf{v}_{2k}) + ... + F(\mathbf{a}, \mathbf{v}_{nk}) \geq \widehat{F}(\mathbf{a}, \mathbf{u}).$$

We will arrive at the desired inequality for lim inf if we show that

$$[F(\mathbf{a}, \mathbf{v}_{1k} + (\mathbf{u} - \mathbf{u}_k)) + F(\mathbf{a}, \mathbf{v}_{2k}) + ... + F(\mathbf{a}, \mathbf{v}_{nk})] - \widehat{F}(\mathbf{a}_k, \mathbf{u}_k) \to 0.$$

The left-hand side difference here is

$$\begin{aligned}
&[F(\mathbf{a},\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k))+F(\mathbf{a},\mathbf{v}_{2k})+...+F(\mathbf{a},\mathbf{v}_{nk})] \\
&\qquad -[F(\mathbf{a}_k,\mathbf{v}_{1k})+F(\mathbf{a}_k,\mathbf{v}_{2k})+...+F(\mathbf{a}_k,\mathbf{v}_{nk})] \\
&= [F(\mathbf{a},\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k))-F(\mathbf{a}_k,\mathbf{v}_{1k})]+[F(\mathbf{a},\mathbf{v}_{2k})-F(\mathbf{a}_k,\mathbf{v}_{2k})] \\
&\qquad +\ldots+[F(\mathbf{a},\mathbf{v}_{nk})-F(\mathbf{a}_k,\mathbf{v}_{nk})],
\end{aligned}$$

where

$$\begin{aligned}
&F(\mathbf{a},\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k))-F(\mathbf{a}_k,\mathbf{v}_{1k}) \\
&\qquad = (|\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)|-|\mathbf{v}_{1k}|)\,F\left(\mathbf{a},\overline{\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)}\right) \\
&\qquad\qquad + |\mathbf{v}_{1k}|\left[F\left(\mathbf{a},\overline{\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)}\right)-F(\mathbf{a}_k,\overline{\mathbf{v}}_{1k})\right],
\end{aligned}$$

and

$$F(\mathbf{a},\mathbf{v}_{ik})-F(\mathbf{a}_k,\mathbf{v}_{ik})=|\mathbf{v}_{ik}|\,[F(\mathbf{a},\overline{\mathbf{v}}_{ik})-F(\mathbf{a}_k,\overline{\mathbf{v}}_{ik})]\,, i=2,\ldots,n.$$

Since $\mathbf{u}_k\to\mathbf{u}$, $\mathbf{a}_k\to\mathbf{a}$, and F is uniformly continuous and bounded on the compact set of unit vectors, we have

$$\begin{gathered}
|\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)|-|\mathbf{v}_{1k}|\to 0, \\
F\left(\mathbf{a},\overline{\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)}\right)-F(\mathbf{a}_k,\overline{\mathbf{v}}_{1k})\to 0, \\
(|\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)|-|\mathbf{v}_{1k}|)\,F\left(\mathbf{a},\overline{\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k)}\right)\to 0, \\
F(\mathbf{a},\overline{\mathbf{v}}_{ik})-F(\mathbf{a}_k,\overline{\mathbf{v}}_{ik})\to 0.
\end{gathered}$$

To see that

$$\begin{gathered}
F(\mathbf{a},\mathbf{v}_{1k}+(\mathbf{u}-\mathbf{u}_k))-F(\mathbf{a}_k,\mathbf{v}_{1k})\to 0, \\
F(\mathbf{a},\mathbf{v}_{ik})-F(\mathbf{a}_k,\mathbf{v}_{ik})\to 0, \qquad i=2,\ldots,n,
\end{gathered}$$

it remains to show that $|\mathbf{v}_{ik}|$ is bounded for $i=2,\ldots,n$. But this follows from the fact that

$$F(\mathbf{a}_k,\mathbf{v}_{1k})+\ldots+F(\mathbf{a}_k,\mathbf{v}_{nk})\le F(\mathbf{a}_k,\mathbf{u}_k)\to F(\mathbf{a},\mathbf{u})\,,$$

because of which

$$F(\mathbf{a}_k,\mathbf{v}_{ik})=|\mathbf{v}_{ik}|\,F(\mathbf{a}_k,\overline{\mathbf{v}}_{ik})\le F(\mathbf{a},\mathbf{u})+C,$$

where C is some positive constant. □

Theorem 52. *The distance $G(\mathbf{x},\mathbf{x}+\mathbf{u}s)$ is differentiable at $s=0+$ for any $(\mathbf{x},\mathbf{u})\in\mathbb{T}$, and*

$$\left.\frac{\mathrm{d}G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{\mathrm{d}s+}\right|_{s=0}=\lim_{s\to 0+}\frac{G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{s}=\widehat{F}(\mathbf{x},\mathbf{u})\,.$$

Proof. We prove first that

$$\limsup_{s\to 0+}\frac{G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{s\widehat{F}(\mathbf{x},\mathbf{u})}\leq 1.$$

Let $(\mathbf{u}_1,\ldots,\mathbf{u}_n)$ be a minimizing vector chain for $(\mathbf{x},\mathbf{u})$, so that

$$\widehat{F}(\mathbf{x},\mathbf{u})=F(\mathbf{x},\mathbf{u}_1)+\ldots+F(\mathbf{x},\mathbf{u}_n).$$

Consider the chain of points

$$\mathbf{x}\ [\mathbf{x}+\mathbf{u}_1 s]\ [\mathbf{x}+(\mathbf{u}_1+\mathbf{u}_2)s]\ \ldots\ [\mathbf{x}+(\mathbf{u}_1+\ldots+\mathbf{u}_n)s],$$

in which the last point coincides with $\mathbf{x}+\mathbf{u}s$. We will generically refer to a point in this chain as

$$\mathbf{x}+(\mathbf{u}_1+\ldots+\mathbf{u}_i)s,\qquad i=0,1,\ldots,n,$$

with the obvious convention for $i=0$. For all sufficiently small s, all these points belong to a compact ball in $\mathfrak{S}$ centered at $\mathbf{x}$. Then, by Theorem 31 and the continuity of F, we have, as $s\to 0+$,

$$\begin{aligned}&\frac{D[\mathbf{x}+(\mathbf{u}_1+\ldots\mathbf{u}_i)s][\mathbf{x}+(\mathbf{u}_1+\ldots+\mathbf{u}_{i+1})s]}{sF(\mathbf{x},\mathbf{u}_{i+1})}\\&\quad=\frac{D[\mathbf{x}+(\mathbf{u}_1+\ldots\mathbf{u}_i)s][\mathbf{x}+(\mathbf{u}_1+\ldots+\mathbf{u}_{i+1})s]}{F(\mathbf{x}+(\mathbf{u}_1+\ldots\mathbf{u}_i)s,\mathbf{u}_{i+1}s)}\\&\qquad\times\frac{sF(\mathbf{x}+(\mathbf{u}_1+\ldots\mathbf{u}_i)s,\mathbf{u}_{i+1})}{sF(\mathbf{x},\mathbf{u}_{i+1})}\to 1,\end{aligned}$$

whence

$$\begin{aligned}&\frac{D\mathbf{x}[\mathbf{x}+\mathbf{u}_1 s]\ldots[\mathbf{x}+\mathbf{u}s]}{s\widehat{F}(\mathbf{x},\mathbf{u})}\\&\quad=\frac{\sum_{i=0}^{n-1}D[\mathbf{x}+(\mathbf{u}_1+\ldots\mathbf{u}_i)s][\mathbf{x}+(\mathbf{u}_1+\ldots+\mathbf{u}_{i+1})s]}{s\sum_{i=1}^{n}F(\mathbf{x},\mathbf{u}_i)}\to 1.\end{aligned}$$

But then

$$\limsup_{s\to 0+}\frac{G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{s\widehat{F}(\mathbf{x},\mathbf{u})}=\limsup_{s\to 0+}\frac{G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{D\mathbf{x}[\mathbf{x}+\mathbf{u}_1 s]\ldots[\mathbf{x}+\mathbf{u}s]}\leq 1,$$

by the definition of G. We prove next that

$$\liminf_{s\to 0+}\frac{G(\mathbf{x},\mathbf{x}+\mathbf{u}s)}{s\widehat{F}(\mathbf{x},\mathbf{u})}\geq 1.$$

Consider a sequence of chains

$$\mathbf{x}[\mathbf{x}+\mathbf{v}_{1k}s_k][\mathbf{x}+(\mathbf{v}_{1k}+\mathbf{v}_{2k})s_k]\ldots[\mathbf{x}+(\mathbf{v}_{1k}+\ldots+\mathbf{v}_{m_k k})s_k],\qquad k=1,2,\ldots,$$

such that

$$s_k \to 0+,$$

$$\mathbf{v}_{1k} + \ldots + \mathbf{v}_{m_k k} = \mathbf{u}, \qquad k = 1, 2, \ldots,$$

and

$$\frac{D\mathbf{x}\left[\mathbf{x} + \mathbf{v}_{1k} s_k\right] \ldots \left[\mathbf{x} + \mathbf{u} s_k\right]}{G\left(\mathbf{x}, \mathbf{x} + \mathbf{u} s\right)} \to 1.$$

Again, it is easy to see that for all all k sufficiently large (i.e., s_k sufficiently small) all these chains fall within a compact ball in $\mathfrak{S}$ centered at $\mathbf{x}$. Then, for $i = 0, 1, \ldots, m_k - 1$, by Theorem 31 and the continuity of F, as $k \to \infty$,

$$\begin{aligned} &\frac{D\left[\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots \mathbf{v}_{ik}\right) s_k\right]\left[\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots + \mathbf{v}_{i+1,k}\right) s_k\right]}{s_k F\left(\mathbf{x}, \mathbf{v}_{i+1,k}\right)} \\ &\qquad = \frac{D\left[\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots \mathbf{v}_{ik}\right) s_k\right]\left[\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots + \mathbf{v}_{i+1,k}\right) s_k\right]}{F\left(\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots \mathbf{v}_{ik}\right) s_k, \mathbf{v}_{i+1,k} s_k\right)} \\ &\qquad\qquad \times \frac{s_k F\left(\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots \mathbf{v}_{ik}\right) s_k, \mathbf{v}_{i+1,k}\right)}{s_k F\left(\mathbf{x}, \mathbf{v}_{i+1,k}\right)} \to 1 \end{aligned}$$

uniformly across all choices of $\left(\mathbf{v}_{1k} + \ldots \mathbf{v}_{m_k k}\right)$. It follows that

$$\begin{aligned} &\frac{D\mathbf{x}\left[\mathbf{x} + \mathbf{v}_{1k} s_k\right] \ldots \left[\mathbf{x} + \mathbf{u} s_k\right]}{s_k \sum_{i=1}^{m_k} F\left(\mathbf{x}, \mathbf{v}_{ik}\right)} \\ &\qquad = \frac{\sum_{i=0}^{m_k - 1} D\left[\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots \mathbf{v}_{ik}\right) s_k\right]\left[\mathbf{x} + \left(\mathbf{v}_{1k} + \ldots + \mathbf{v}_{i+1,k}\right) s_k\right]}{s_k \sum_{i=1}^{m_k} F\left(\mathbf{x}, \mathbf{v}_{ik}\right)} \to 1. \end{aligned}$$

But then

$$\begin{aligned} \liminf_{s \to 0+} \frac{G\left(\mathbf{x}, \mathbf{x} + \mathbf{u} s\right)}{s \widehat{F}\left(\mathbf{x}, \mathbf{u}\right)} &= \liminf_{k \to \infty} \frac{D\mathbf{x}\left[\mathbf{x} + \mathbf{v}_{1k} s_k\right] \ldots \left[\mathbf{x} + \mathbf{u} s_k\right]}{s_k \widehat{F}\left(\mathbf{x}, \mathbf{u}\right)} \\ &= \liminf_{k \to \infty} \frac{\sum_{i=1}^{m_k} F\left(\mathbf{x}, \mathbf{v}_{ik}\right)}{\widehat{F}\left(\mathbf{x}, \mathbf{u}\right)} \geq 1, \end{aligned}$$

by the definition of $\widehat{F}$ in terms of minimizing chains. This establishes

$$\lim_{s \to 0+} \frac{G\left(\mathbf{x}, \mathbf{x} + \mathbf{u} s\right)}{s \widehat{F}\left(\mathbf{x}, \mathbf{u}\right)} = 1,$$

and the theorem is proved. □

Theorem 56. *For every path* $\mathbf{h}|\left[a, b\right]$ *connecting* $\mathbf{a}$ *to* $\mathbf{b}$ *one can find a piecewise linear path from* $\mathbf{a}$ *to* $\mathbf{b}$ *which is arbitrarily close to* $\mathbf{h}|\left[a, b\right]$ *pointwise and in its length.*

Proof. Let

$$\mu_n = \left\{a = t_{n0}, ..., t_{ni}, t_{n,i+1}, ..., t_{n,k_n+1} = b\right\}$$

be a sequence of nets with $\delta\mu_n \to 0$. Since the set $\mathbf{h}([a,b])$ is compact, n can be chosen sufficiently large so that any two successive $\mathbf{h}(\alpha = t_{ni})$ and $\mathbf{h}(\beta = t_{n,i+1})$ can be connected by a straight line segment

$$\mathbf{s}_{ni}(t) = \mathbf{h}(\alpha) + \frac{\mathbf{h}(\beta) - \mathbf{h}(\alpha)}{\beta - \alpha}(t - \alpha).$$

Then n can further be increased to ensure

$$1 - \varepsilon < \frac{G\mathbf{h}(\alpha)\,\mathbf{h}(\beta)}{\widehat{F}(\mathbf{h}(\alpha), \mathbf{h}(\beta) - \mathbf{h}(\alpha))} < 1 + \varepsilon$$

and

$$1 - \varepsilon < \frac{D\mathbf{s}_{ni}|\,[\alpha,\beta]}{\widehat{F}(\mathbf{h}(\alpha), \mathbf{h}(\beta) - \mathbf{h}(\alpha))} < 1 + \varepsilon.$$

The latter follows from

$$D\mathbf{s}_{ni}|\,[\alpha,\beta] = \int_\alpha^\beta \widehat{F}\left(\mathbf{h}(x), \dot{\mathbf{h}}(x)\right) \mathrm{d}x = \widehat{F}\left(\mathbf{h}(\xi), \frac{\mathbf{h}(\beta) - \mathbf{h}(\alpha)}{\beta - \alpha}\right)(\beta - \alpha),$$

for some $\alpha \le \xi \le \beta$. Combining the two double-inequalities, for any $\delta > 0$ and all sufficiently large n,

$$1 - \delta < \frac{G\mathbf{h}(t_{ni})\,\mathbf{h}(t_{n,i+1})}{D\mathbf{s}_{ni}|\,[t_{ni}, t_{n,i+1}]} < 1 + \delta,$$

whence

$$1 - \delta < \frac{\sum_{i=0}^{k_n} G\mathbf{h}(t_{ni})\,\mathbf{h}(t_{n,i+1})}{D\mathbf{s}_n|\,[a,b]} < 1 + \delta,$$

where $\mathbf{s}_n|\,[a,b]$ is the *piecewise linear* path concatenating together $\mathbf{s}_{ni}|\,[t_{ni}, t_{n,i+1}]$, $i = 0, \ldots, k_n$. By the definition of $D\mathbf{h}|\,[a,b]$, we have then

$$\lim_{n\to\infty} D\mathbf{s}_n|\,[a,b] = D\mathbf{h}|\,[a,b].$$

Since it is obvious that, as $n \to \infty$, $\mathbf{s}_n|\,[a,b]$ tends to $\mathbf{h}|\,[a,b]$ pointwise, the theorem is proved. □

References

[1] Blumenthal, L.M. (1953). *Theory and Applications of Distance Geometry.* London: Oxford University.

[2] Blumenthal, L.M., & Menger, K. (1970). *Studies in Geometry.* San Francisco, CA: W.H. Freeman.

[3] Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling.* New York: Springer-Verlag.

[4] Busemann, H. (2005). *The Geometry of Geodesics*. Mineola, NY: Dover.

[5] Colonius, H. & Dzhafarov, E. N. (2012). Ultrametric Fechnerian Scaling of discrete object sets. In M. Deza, M. Petitjean, K. Markov (Eds.) *The Mathematics of Distances and Applications* (pp. 129-132). Sofia, Bulgaria: ITHEA® Publisher.

[6] Creelman, C. D. (1967). Empirical detectability scales without the jnd. *Perceptual & Motor Skills*, 24, 1079-1084.

[7] Dzhafarov, D.D., & Dzhafarov, E. N. (2011). The equivalence of two ways of computing distances from dissimilarities for arbitrary sets of stimuli. *Journal of Mathematical Psychology*, 55, 469-472.

[8] Dzhafarov, E.N. (2002). Multidimensional Fechnerian scaling: Pairwise comparisons, regular minimality, and nonconstant self-similarity. *Journal of Mathematical Psychology*, 46, 583-608.

[9] Dzhafarov, E.N. (2003a). Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and independent images. *Journal of Mathematical Psychology*, 47, 208-228.

[10] Dzhafarov, E.N. (2003b). Thurstonian-type representations for "same-different" discriminations: Probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, 47, 229-243. [see Dzhafarov, E.N. (2006). Corrigendum to "Thurstonian-type representations for 'same–different' discriminations: Probabilistic decisions and interdependent images." *Journal of Mathematical Psychology*, 50, 511.]

[11] Dzhafarov, E.N. (2006). On the law of Regular Minimality: Reply to Ennis. *Journal of Mathematical Psychology*, 50, 74-93.

[12] Dzhafarov, E.N. (2008a). Dissimilarity cumulation theory in arc-connected spaces. *Journal of Mathematical Psychology*, 52, 73–92.[see Dzhafarov, E.N. (2009). Corrigendum to: "Dissimilarity cumulation theory in arc-connected spaces." *Journal of Mathematical Psychology*, 53, 300.]

[13] Dzhafarov, E.N. (2008b). Dissimilarity cumulation theory in smoothly connected spaces. *Journal of Mathematical Psychology*, 52, 93–115.

[14] Dzhafarov, E.N. (2010a). Dissimilarity Cumulation as a procedure correcting for violations of triangle inequality. *Journal of Mathematical Psychology*, 54, 284-287.

[15] Dzhafarov E.N. (2010b). FSCAMDS—Fechnerian Scaling Followed by Clustering and MDS. MATLAB program version 20100713, URL http://www.psych.purdue.edu/ ~ehtibar/links.

[16] Dzhafarov, E.N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychonomic Bulletin and Review*, 6, 239-268.

[17] Dzhafarov, E.N., & Colonius, H. (2001). Multidimensional Fechnerian scaling: Basics. *Journal of Mathematical Psychology*, 45, 670-719.

[18] Dzhafarov, E.N., & Colonius, H. (2006a). Reconstructing distances among objects from their discriminability. *Psychometrika*, 71, 365 - 386.

[19] Dzhafarov, E.N., & Colonius, H. (2006b). Regular Minimality: A fundamental law of discrimination. In H. Colonius & E.N. Dzhafarov (Eds.), *Measurement and Representation of Sensations* (pp. 1-46). Mahwah, NJ: Erlbaum.

[20] Dzhafarov, E.N., & Colonius, H. (2006c). Generalized Fechnerian Scaling. In H. Colonius & E.N. Dzhafarov (Eds.), *Measurement and Representation of Sensations* (pp. 47-88). Mahwah, NJ: Erlbaum.

[21] Dzhafarov, E.N., & Colonius, H. (2007). Dissimilarity Cumulation theory and subjective metrics. *Journal of Mathematical Psychology*, 51, 290-304.

[22] Dzhafarov, E.N., & Dzhafarov, D.D. (2012). The sorites paradox: A behavioral approach. In L. Rudolph (Ed.) *Qualitative Mathematics for the Social Sciences: Mathematical Models for Research on Cultural Dynamics* (pp. 105-136). London: Routledge.

[23] Dzhafarov, E.N., & Paramei, G.V. (2010). Space of facial expressions: Cumulated versus transformed dissimilarities. In A. Bastianelli & G. Vidotto (Eds.), *Fechner Day 2010* (pp. 605–610). Padua, Italy: The International Society for Psychophysics.

[24] Dzhafarov, E.N., & Perry, L. (2014). Perceptual matching and sorites: Experimental study of an ancient Greek paradox. *Attention, Perception, and Psychophysics* 76, 2441-2464.

[25] Everitt, B. S., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster Analysis, 5th Edition.* New York: John Wiley and Sons (Wiley Series in Probability and Statistics).

[26] Falmagne, J. C. (1971). The generalized Fechner problem and discrimination. *Journal of Mathematical Psychology*, 8, 22-43.

[27] Fechner, G. T. (1860). *Elemente der Psychophysik* [*Elements of Psychophysics*]. Leipzig: Breitkopf & Härtel.

[28] Fechner, G. T. (1877). *In Sachen der Psychophysik* [*In the matter of psychophysics*]. Leipzig: Breitkopf & Härtel.

[29] Fechner, G. T. (1887). Über die psychischen Massprinzipien und das Webersche Gesetz [On the principles of mental measurement and Weber's Law]. *Philosophische Studien*, 4, 161–230.

[30] Helmholtz, H. von. (1891). Versuch einer erweiterten Anwendung des Fechnerschen Gesetzes im Farbensystem [An attempt at a generalized application of Fechner's Law to the color system]. *Zeitschrift für die Psychologie und die Physiologie der Sinnesorgane*, 2, 1–30.

[31] Hocking, J.H. &Young, G.S. (1961). *Topology.* Reading, MA: Addison-Wesley.

[32] Kelly, J.L. (1955). *General Topology.* Toronto: Van Nostrand.

[33] Krantz, D. (1971). Integration of just-noticeable differences. *Journal of Mathematical Psychology*, 8, 591-599.

[34] Kujala, J.V., & Dzhafarov, E.N. (2008). On minima of discrimination functions. *Journal of Mathematical Psychology*, 52, 116–127.

[35] Kujala, J.V., & Dzhafarov, E.N. (2009a). Regular Minimality and Thurstonian-type modeling. *Journal of Mathematical Psychology*, 53, 486–501

[36] Kujala, J.V., & Dzhafarov, E.N. (2009b). A new definition of well-behaved discrimination functions. *Journal of Mathematical Psychology*, 53, 593–599.

[37] Luce, R.D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24, 178-191.

[38] Luce, R.D., & Edwards, W. (1958). The derivation of subjective scales from just noticeable differences. *Psychological Review* 65, 222-237.

[39] Luce, R.D., & Galanter, E. (1963). Discrimination. In Luce R.D., Bush R.R., Galanter E. (Eds.), *Handbook of Mathematical Psychology*, vol. 1, 191-244. Wiley, New York.

[40] Pfanzagl, J. (1962). Über die stochastische Fundierung des psychophysischen Gesetzes [On stochastic foundations of the psychophysical law]. *Biometrische Zeitschrift*, 4, 1-14.

[41] Schrödinger, E. von (1920/1970). Outline of a theory of color measurement for daylight vision. In D.L. MacAdam (Ed.), *Sources of Color Science* (pp. 397–447, 481–520). Cambridge, MA: MIT Press.

[42] Schrödinger, E. von (1926/1970). Thresholds of color differences. In D.L. MacAdam (Ed.), *Sources of Color Science* (pp. 183–193). Cambridge, MA: MIT Press.

[43] Ünlü, A., Kiefer, T., & Dzhafarov, E.N. (2009). Fechnerian Scaling in R: The package fechner. Journal of Statistical Software, 31, Issue 6, 1-24. (URL for the paper and software: http://www.jstatsoft.org/v31/i06.)