Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Mathematical Psychology 48 (2004) 409–424

Journal of Mathematical Psychology

www.elsevier.com/locate/jmp

Dual scaling of comparison and reference stimuli in multi-dimensional psychological space

Jun Zhang*

Department of Psychology, University of Michigan, 525 East University Ave, Ann Arbor, MI 48109, USA

Received 15 December 2003; received in revised form 2 September 2004

Abstract

Dzhafarov and Colonius (Psychol. Bull. Rev. 6 (1999)239; J. Math. Psychol. 45(2001)670) proposed a theory of Fechnerian scaling of the stimulus space based on the psychometric (discrimination probability) function of a human subject in a same–different comparison task. Here, we investigate a related but different paradigm, namely, referent–probe comparison task, in which the pair of stimuli (x and y) under comparison assumes substantively different psychological status, one serving as a referent and the other as a probe. The duality between a pair of psychometric functions, arising from assigning either x or y to be the fixed reference stimulus and the other to be the varying comparison stimulus, and the 1-to-1 mapping between the two stimulus spaces X and Y under either assignment are analyzed. Following Dzhafarov and Colonius, we investigate two properties characteristic of a referent–probe comparison task, namely, (i) Regular cross-minimality—for the pair of stimulus values involved in referent–probe comparison, each minimizes a discrimination probability function where the other is treated as the fixed reference stimulus; (ii) Nonconstant self-similarity—the value of the discrimination probability function at its minima is a nonconstant function of the reference stimulus value. For the particular form of psychometric functions investigated, it is shown that imposing the condition of regular cross-minimality on the pair of psychometric functions forces a consistent (but otherwise still arbitrary) mapping between X and Y, such that it is independent of the assignment of reference/comparison status to x and to y. The resulting psychometric differentials under both assignments are equal, and take an asymmetric, dualistic form reminiscent of the so-called divergence measure that appeared in the context of differential geometry of the probability manifold with dually flat connections (Differential Geometric Methods in Statistics, Lecture Notes in Statistics, Vol. 28, Springer, New York, 1985). The pair of divergence functions on X and on Y, respectively, induce a Riemannian metric in the small, with psychometric order (defined in Dzhafarov & Colonius, 1999) equal to 2. The difference between the Finsler–Riemann geometric approach to the stimulus space (Dzhafarov & Colonius, 1999) and this dually-affine Riemannian geometric approach to the dual scaling of the comparison and the reference stimuli is discussed.

Keywords: Referent-probe comparision task; Psychometric differential; Divergence function; Convex representation; Convex conjugate; Legendre–Fenchel duality

1. Introduction

In geometric theories of the scaling of the stimulus space, the notion of "distance" plays a key role in describing the proximity between various stimuli whose features are defined in some multi-dimensional vector space R^n. Often, such distance measures are induced by the so-called *norm* of a vector space, formally defined as a real-valued function $R^n \to R_+ \equiv R^+ \cup \{0\}$ and denoted $\|\cdot\|$, that satisfies the following conditions for all $x, y \in R^n$ and $\alpha \in R$: (i) $\|x\| \geqslant 0$ with the equality holding if and only if $x = 0$; (ii) $\|\alpha x\| = |\alpha| \cdot \|x\|$; (iii) $\|x + y\| \leqslant \|x\| + \|y\|$. The distance measure or *metric* induced by such a norm takes the form

$$\Delta(x, x') = \|x - x'\|. \tag{1}$$

*Fax: +734 763 7480.
E-mail address: junz@umich.edu.

0022-2496/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2004.09.001

Obviously, such a metric satisfies, in addition to being continuous with respect to its arguments, the axioms of (i) nonnegativity: $\Delta(x, x') \geqslant 0$, with 0 attained if and only if $x = x'$; (ii) symmetry: $\Delta(x, x') = \Delta(x', x)$; and (iii) triangle inequality: $\Delta(x, x') + \Delta(x', x'') \geqslant \Delta(x, x'')$ for any triplet $x, x', x'' \in R^n$. A familiar example is the Minkowski metric, which is induced on a vector space equipped with an L_p norm: $\|x\|_p = \sum_{i=1}^{n} (|x^i|^p)^{1/p}$ (here x^i, $i = 1, 2, \ldots, n$ denote vector components). The distance measure $\Delta(x, x')$ forms the basis of many geometric models of stimulus similarity in the psychological space (Shepard, 1962a,b).

Mounting evidence has accumulated indicating that comparative judgment in humans violated almost all of the basic axioms about a norm-induced metric, therefore questioning the psychological validity of (1) as a measure of proximity. Violation of nonnegativity (in the strict sense which includes the condition of attaining zero) occurred when the point of subjective equality in a psychometric measurement does not always turn out to be identical to the fixed, reference stimulus value. Violation of symmetry was frequently demonstrated following Tversky's (1977) report that subjects rated less prominent Communist countries (e.g., North Korea) as more similar to more prominent country of the same Communist category (e.g., China) than vice versa—such asymmetry was attributed to a difference in the weighting of distinctive/unique features that the comparison and the reference stimulus each possessed (Tversky, 1977), to an emphasis of the common features that they shared (Ortony, 1979; Medin, Goldstone, & Gentner, 1993), to the prototypicality or saliency of the two stimuli (Rosch, 1975), or even to their relative frequencies of recent exposure (Polk, Behensky, Gonzalez, & Smith, 2002). Finally, violation of triangle inequality was demonstrated by Tversky and Gati (1982) when perceived dissimilarity for objects that differed on two separable dimensions (attributes) was found to be larger than predicted from their perceived unidimensional differences. While violation of nonnegativity can be easily remedied by postulating a rescaling for comparison or reference stimulus, it is difficult to salvage the axioms of symmetry and of triangle inequality under any metric-based framework. In fact, it is unclear whether symmetry is a desirable requirement in comparative judgment tasks, simply because the different roles a fixed referent and a variable probe may play.

Recently, a complete reformulation and re-characterization of the distance measure in the stimulus space has been proposed by Dzhafarov and Colonius (1999, 2001), utilizing the language of Finsler geometry. As a major departure from the standard approach of inducing the metric through a norm via (1), the framework uses the empirically observed psychometric function (specifically, the discrimination probability function in the same–different discrimination task) to construct a Finslerian metric function in the small (i.e., locally) and then, through an integration and minimization procedure, the Finslerian distance in the large. To be more specific, that theory starts with a discrimination probability function in a same/different judgment task, denoted here as $\Psi_x(y)$ between a comparison stimulus $y \in R^n$ against a fixed, reference stimulus $x \in R^n$. The psychometric differential is defined as[1]

$$\Psi_x(x + us) - \Psi_x(x),$$

where $s > 0$ is the magnitude of the transition from stimulus x to stimulus $x + us$ in the direction $u \neq 0$ (i.e., $u \in R^n - \{0\}$). Under some reasonable assumptions about how this differential (viewed as a function of x, u) would behave when $s \to 0$, Dzhafarov and Colonius (1999, 2001) showed that there exists a global transformation (global in the sense of being independent of x, u), denoted $g : R_+ \to R_+$, such that when it is applied to the psychometric differential, the following holds:

$$\lim_{s \to 0^+} \frac{g(\Psi_x(x + us) - \Psi_x(x))}{s} \equiv F(x, u).$$

This was called *The Fundamental Theorem* of Fechnerian scaling; and the resultant function $F(x, u)$ naturally satisfies Euler homogeneity with respect to its second variable: $F(x, ku) = kF(x, u)$ for $k > 0$. The function $F(x, u)$, or its convex closure if it is not already convex (denoted $\tilde{F}(x, u)$ in Dzhafarov & Colonius (2001) but we ignore such technicality), becomes the local metric function of an underlying Finsler space. When g takes the form of a power function: $g(t) = t^{1/\mu}$ or more generally $g(t) = (tl(t))^{1/\mu}$ where $l(t)$ is slowly varying, then $\mu \in R^+$ is referred to as the "psychometric order" of the discrimination probability function and, due to a fundamental assumption in their papers, that of the stimulus space.

Note that $F(x, u)$ introduced above, termed Finslerian metric function in the mathematical theory of Finsler space (c.f., Rund, 1959), is actually a norm of the tangent space (defining the u variable) associated with a point (defining the x variable) sitting in the base manifold—roughly, it is proportional to the distance between infinitesimally nearby points on the base manifold $\mathscr{B}$ along the tangent vector u. Unlike (1), to obtain the Finslerian distance in the large between any

[1]This notation for a psychometric function was used in Dzhafarov and Colonius (1999, 2001) to denote the probability of same–different judgment for a pairwise presentation of stimulus values. The subscript merely refers to fixing a reference point during post hoc analysis of the psychometric function obtained through this procedure. For this reason, in Dzhafarov (2002d), the $\Psi(x, y)$ notation is favored to stress the fact that the psychometric function in same–different comparison task is a two-variable function. Here for mathematical clarity, we retain the subscript notation to indicate the stimulus that is fixed (either procedurally or mentally by instruction).

two points x and x' on the base manifold $\mathcal{B}$, one first constructs (*step one*) a path γ connecting x to x': $[0,1] \ni t \mapsto \gamma(t) \in \mathcal{B}$ with $\gamma(0) = x, \gamma(1) = x'$, and then computes (*step two*) the length of such a path by integrating

$$L_\gamma = \int_0^1 F\left(\gamma(t), \frac{d\gamma(t)}{dt}\right) dt$$

(note that we have replaced x by $\gamma(t)$ and u by $d\gamma/dt$ in the two slots of $F(\cdot,\cdot)$, and that Euler homogeneity ensures the above integral to be invariant with respect to any re-parameterization of the path $\gamma(\cdot)$). Then finally, one takes (*step three*) the infimum of all such paths linking the two points x, x' as their end points to define the Finslerian distance $G(x, x')$ in the large:

$$G(x, x') = \inf_{\{\gamma(t):\gamma(0)=x,\gamma(1)=x'\}} L_\gamma.$$

The function $G(x, x')$ can easily been shown to satisfy nonnegativity and triangle inequality but not necessarily symmetry in general. A distance function $G(x, x')$ in the large constructed through this procedure (following steps 1–3 above) from a suitably defined metric function $F(x, u)$ in the small is said to be "internal" (or "intrinsic"). Whenever, after a global transformation h (continuous and increasing) on the distance function, such construction is possible, the distance function is called "internalizable" (Dzhafarov, 2002b). The idea of an "internal" or "internalizable" metric for a stimulus space is closed linked with, and in fact a mathematical articulation of Gustav Fechner's program of "psychophysics", i.e., linking distances measured in the large ("stimulus scaling") to just noticeable differences (jnd's) measured in the small ("thresholds" of discrimination).

The mathematical framework of Finsler geometry, which extends Riemannian geometry in a natural way (c.f., the treatise on the topic by Rund, 1959), makes clear the conceptual difference and the connection between a metric in the large $G(x, x')$ that describes the distance of two points with finite separation *on the base manifold* $\mathcal{B}$ and a metric in the small $F(x, u)$ that describes the distance between points that are infinitesimally close—in the latter case between a point x and another point that is slightly off in the direction of u. The basic premise of a differential manifold is the assumption that the infinitesimal neighborhood of a point x on the base manifold forms a space that is diffeomorphic to a vector space (diffeomorphism meaning, roughly, a transformation of the vector space characterized by an arbitrary change of coordinates with nonvanishing Jacobian), called the tangent space and denoted by the u variable, so that nearby points on the base manifold can now be represented by different vectors *in the tangent space* that is situated on that particular point of the original manifold. To avoid confusion, we call $G(x, x')$ "distance" and called $F(x, u)$ "metric function". In the application to the scaling of the stimulus space, the former corresponds to the subjective distance between two stimuli x and x' sufficiently farther apart, while the latter describes psychometric measurements of discrimination threshold when the value of the comparison stimulus changes in different direction u around a fixed stimulus x.

This Fechnerian scaling framework advanced by Dzhafarov and Colonius (1999) is seen as a reformulation of the generalized Weber's Law (Luce & Galanter, 1963)

$$\Psi(x + w(x, s)) - \Psi(x) = \eta(s),$$

where the psychophysical function $\Psi(x)$ measures the subjective scale associated with a physical stimulus x, and the Weber fraction $w(x, s)$ is used to induce an increment $\eta(s)$ of subjective distance in the stimulus space. In fact Finsler geometry is seen as a more faithful implementation of Fechner's original proposal and a mathematically correct procedure for deriving distance in the stimulus space based on experimental measurements of psychometric thresholds. For one thing, it provides a natural and precise expression of the Probability-Distance Hypothesis, namely, the representability of the discrimination probability $\Psi_x(y)$ as a function $h : R_+ \to R_+$ of some distance function $\Delta(x, y)$ of the underlying stimulus space: $\Psi_x(y) = h(\Delta(x, y))$. It was shown (Dzhafarov, 2002b) that if such $\Delta(x, y)$ exists and is either internal or internalizable, then $\Delta(x, y)$ is either $G(x, y)$ itself (as induced by the $\Psi_x(y) \to F(x, u) \to G(x, x')$ procedure) or a transformation thereof; this is due to the fundamental relation between the distance $G(x, x')$ and the metric function $F(x, u)$ in a Finsler space:

$$\lim_{s\to 0^+} \frac{G(x, x + us)}{s} = F(x, u),$$

$$\inf_{\text{all } \gamma(t)|_{\gamma(0)=x}^{\gamma(1)=x'}} \int_0^1 F\left(\gamma(t), \frac{d\gamma(t)}{dt}\right) dt = G(x, x')$$

(here $F(x, u)$ is assumed to be convex in u, and $G(x, x')$ satisfies metric axioms).

This well-formulated theory of psychological scaling using Finsler geometry techniques invoked certain assumptions about the asymptotic shape of a psychometric function as the two stimuli approach one another. Dzhafarov (2002d) recently investigated two conditions on psychometric functions that appear theoretically motivated and empirically robust (see, e.g., Zimmer & Colonius (2000), and many other studies cited in Dzhafarov (2002d)): (i) the condition of regular minimality—for the pair of stimulus values involved, each minimizes the discrimination probability function when the other is *viewed* i.e., mathematically (but not procedurally taken) as fixed; and (ii) the condition of nonconstant self-similarity—the values of the discrimination probability function at its minima when either stimulus is mathematically fixed is a nonconstant function of the fixed stimulus value. While their exact meaning will be elaborated in the next subsection, these two conditions have been shown to preclude a large class of well-behaved Thurstonian scaling functions (Dzhafarov, 2003a,b). In fact, they impose severe constraints on the psychometric order of any discrimination probability function and on the stimulus space. Nonconstant self-similarity alone would also preclude the Probability-Distance Hypothesis, as the latter necessarily implies constant self-similarity (Dzhafarov, 2002b).

It should be noted that this Fechnerian scaling framework applied to the task of *same–different comparison*, namely, subjects are asked to make a same–different judgment (discrimination) on a pair of stimuli presented in distinct observation intervals that are separated in space and/or in time; the two stimuli involved are nonetheless *not* treated differentially in any other regards. Hence this procedure will yield *one* psychometric function in its canonical form $\Psi(x, y)$, though not necessarily symmetric in its arguments $\Psi(x, y) \neq \Psi(y, x)$. The placeholders for the two arguments in $\Psi(\cdot, \cdot)$ are referred as the two "observation areas" (Dzhafarov, 2002d) without assuming that one of them is to hold a stimulus as the reference and the other as the comparison; it is only in the post hoc mathematical analysis of $\Psi(x, y)$ that one talks about treating or "viewing" x as the reference stimulus $\Psi_{\hat{x}}(y) = \Psi(x, y)|_{x=\hat{x}}$ or y as the reference stimulus $\Psi_{\hat{y}}(x) = \Psi(x, y)|_{y=\hat{y}}$.

In this paper, a related but different task is investigated, which we call the *referent–probe comparison* task. Here, one stimulus in a pair is held fixed while the other is varied, randomly or in ascending/descending ("staircase") order, either controlled by the experimenter or by the subject, while the subject make the same–different judgment. The stimulus that is procedurally (or, by instruction, mentally.) fixed is called the *referent* (*reference stimulus*), while the stimulus that is procedurally (or mentally by instruction) varied is called the *probe* (*comparison stimulus*). The two stimuli, aside from their being presented in distinct temporal or spatial intervals, have substantively different psychological status that subjects must maintain in their mental representations. As an example, suppose a subject is to make a same–different judgment on the "value" of two gambles, one involving a guaranteed payoff of x-utility units and the other involving a probabilistic payoff of y-units (in which the subject will receive either y or 0 as a random event with a known, given probability). In this situation, the experimenter can either hold x fixed and have y values change in a series of trials, called the *forward procedure*, or conversely hold y fixed and have x values change, called the *reverse procedure*. Of course, this terminology can itself be reversed; we use it only to fix the notation.

In such a referent–probe comparison task, the psychometric function $\Psi_x(y)$ obtained from treating x as the reference stimulus may be an entirely different function from the dual psychometric function $\Phi_y(x)$ obtained from treating y as the reference stimulus. For the previous example, it is natural to assume that the comparative process where a fixed amount of guaranteed payoff ("certainty equivalence" or CE) is used as a mental reference for the evaluation of a series of gambles with variable payoffs as their probabilistic outcome is different from the process where a probabilistic outcome with a fixed payoff is used as a reference for the evaluation of varies amounts of certainty equivalence. Indeed, not only the "points of subjective equality" from the two procedures may be different, the entire psychometric functions obtained from these two procedures may be different; the asymmetry in this scenario may reveal some fundamental difference in a subject's mental representation between a risky and a risk-free outcome. Hershey and Schoemaker (1985) employed such mutually dual procedures to separately measure the certainty equivalence of gambles and probability equivalence (PE) of a certain gain/loss (where the probability of a fixed outcome, not the value of a fixed probabilistic outcome, is varied), and showed that subjects often displayed serious inconsistencies between the CE and PE responses in a way that strongly depended on a subject's initial risk attitude and the specific domain of gain versus loss.

Despite the reported violation of this so-called "procedural invariance" in the literature on utility measurement, which has been attributed to cognitive factors such as insufficient adjustment due to anchoring and reframing effects (Hershey & Schoemaker, 1985; Schoemaker & Hershey, 1992), the author is unaware of any study that has already tested this type of procedural invariance in the context of psychophysical measurement. We call this kind of procedural invariance *regular cross-minimality*, or regular minimality across the (mutually dual) procedures. The term is adapted from Dzhafarov's (2002d) notion of "regular minimality" in the same–different comparison tasks. In the current context, the minimality requirement is applied across *two* psychometric functions, whereas in Dzhafarov's case, it is applied to a single psychometric function. Though much stronger than the constraint of regular cross-minimality (see below), the condition of regular minimality was seen to be largely satisfied in same–-different comparison task (Dzhafarov, 2002d).

This paper deals with purely theoretical issues about the conditions of regular (cross)-minimality and nonconstant similarity. These two conditions were shown to restrict possible shapes of a psychometric function for the same–different comparison procedure (Dzhafarov, 2003a,b). One might ask whether they would place any restrictions on the pair of psychometric functions in the referent–probe comparison procedure. A priori, one should not expect any significant constraints—this is because we now have *two* psychometric functions at our disposal. The condition of nonconstant self-similarity merely calls for an opportunity to add an arbitrary single-variable function (of the reference stimulus value only) onto the psychometric function from either the forward or the reverse procedure (but not simultaneously onto both), thus resulting in no substantial constraints on their forms. Applying the regular cross-minimality requirement across two different psychometric functions would, at most, restrict the relationship between one function (for the forward procedure) and the other (for the reverse procedure). However, it will be shown that for the particular form of psychometric functions studied in this paper, this restriction amounts precisely to the requirement of consistent mappings between the two stimulus spaces, regardless of which space contains the reference stimulus and which the comparison stimulus. Regular cross-minimality requires the two psychometric differentials (but not psychometric functions) to be identical. Differential status of the referent and of the probe and asymmetry in the psychological divergence function can manifest in a dualistic fashion—the same psychometric differential can take two alternative yet dually symmetric forms expressing precisely the duality in the stimulus status/psychometric procedure.

1.1. Regular cross-minimality and nonconstant self-similarity

Consider the standard two-alternative choice paradigm, in which two stimuli, assigned the role of comparison stimulus and reference stimulus, respectively, are presented in spatially (or temporally) separated observation intervals. The value of the reference stimulus is fixed while that of the comparison stimulus is variable (adjustable) during a block of trials. Let the psychometric function (called the discrimination probability function in a discrimination paradigm) $\Psi_x(y)$ denote the probability that a comparison stimulus y (written as the functional argument) is judged as being "different" in magnitude, when compared against a reference stimulus x (written as the subscript). The condition of regular cross-minimality states that *if*, corresponding to a particular value of the reference stimulus $\hat{x}$, there exists a unique value of the comparison stimulus $y = \hat{y}$ such that

$$\hat{y} = \operatorname{argmin}_y \Psi_{\hat{x}}(y),$$

then when the entire procedure is reversed, i.e., when $\hat{y}$ is held fixed and x becomes adjustable, the psychometric function $\Phi_{\hat{y}}(x)$ so obtained would yield a minimum value at $x = \hat{x}$:

$$\hat{x} = \operatorname{argmin}_x \Phi_{\hat{y}}(x).$$

The condition of nonconstant self-similarity simply states that the psychometric functions evaluated at their respective minima

$$\Psi_{\hat{x}}(y)|_{y=\hat{y}} \neq const,$$

$$\Phi_{\hat{y}}(x)|_{x=\hat{x}} \neq const,$$

that is, each minimum, in general, is a nonconstant function of the reference stimulus value $\hat{x}$ (in the forward procedure) or $\hat{y}$ (in the reverse procedure), respectively.[2]

To gain insights to what these conditions entail, let us consider the case of unidimensional stimuli. Let $\mathscr{X} \subset R$ and $\mathscr{Y} \subset R$ denote, respectively, the domain where stimulus $x \in \mathscr{X}$ and stimulus $y \in \mathscr{Y}$ are defined. In the forward procedure, elements of $\mathscr{Y}$ are chosen to be the varying comparison stimuli whereas an element of $\mathscr{X}$ is selected as the fixed reference stimulus; the two domains are hence denoted as $\mathscr{Y}_c$ and $\mathscr{X}_r$, respectively. Because the stimuli under consideration can at least be ordered (i.e., defined on an ordinal scale), there must exist some strictly increasing (and hence invertible) transformation $\psi : \mathscr{Y}_c \to \mathscr{X}_r$, so that x and y are comparable. For convenience, we assume differentiability of ψ (as well as ψ^{-1}). For any fixed reference stimulus $\hat{x}$, ψ is chosen so that the psychometric function $\Psi_{\hat{x}}(y)$ achieves its minimum at $y = \psi^{-1}(\hat{x})$, so

$$\left.\frac{d\Psi_{\hat{x}}(y)}{dy}\right|_{y=\psi^{-1}(\hat{x})} = 0$$

and

$$\left.\frac{d^2\Psi_{\hat{x}}(y)}{dy^2}\right|_{y=\psi^{-1}(\hat{x})} > 0.$$

Obviously, if we require

$$\frac{d\Psi_{\hat{x}}(y)}{dy} = \psi(y) - \hat{x},$$

then both of the above would be satisfied (since the first derivative of a strictly increasing function $\psi(\cdot)$ is always positive). Integrating the above yields

$$\begin{aligned}\Psi_{\hat{x}}(y) - \Psi_{\hat{x}}(\psi^{-1}(\hat{x})) \\ = \Psi(y) - \Psi(\psi^{-1}(\hat{x})) - \hat{x}\,(y - \psi^{-1}(\hat{x})),\end{aligned} \tag{2}$$

[2] The reader is reminded that $x, \hat{x}$ always denote the stimulus that is used as a fixed reference in the forward procedure whereas $y, \hat{y}$ always denote the stimulus that is used as a fixed reference in the reverse procedure. The former procedure contains a series of y values used as comparison stimuli, whereas the latter procedure contains a series of x values used as comparison stimuli.

where

$$\Psi(y) = \int_c^y \psi(y)\, dy$$

is strictly convex[3] (the constant associated with the indefinite integral can be any arbitrary number c). Denote

$$\Psi^*(\hat{x}) = \hat{x}\psi^{-1}(\hat{x}) - \Psi(\psi^{-1}(\hat{x})), \tag{3}$$

which is known as the convex conjugate of Ψ (the $*$ on the convex function indicates the conjugacy operation), and introduce what will be called the *psychometric differential*

$$\mathscr{A}^{(\Psi)}(\hat{x}, y) = \Psi(y) - y\hat{x} + \Psi^*(\hat{x}), \tag{4}$$

which attains 0 when $y = \psi^{-1}(\hat{x})$, due to (3). The psychometric function $\Psi_{\hat{x}}(y)$ then takes the form

$$\Psi_{\hat{x}}(y) = \Psi_{\hat{x}}(\psi^{-1}(\hat{x})) + \mathscr{A}^{(\Psi)}(\hat{x}, y). \tag{5}$$

Keeping in mind that $\Psi_{\hat{x}}(y)$ refers to the psychometric/discrimination function arising from the procedure (of referent–probe comparison) whereby stimulus $x = \hat{x}$ is being fixed (a referent); it is viewed as a function of y, the varying comparison stimulus. The minimum value of the discrimination function $\Psi_{\hat{x}}(\psi^{-1}(\hat{x}))$, now viewed as a function of the reference stimulus $\hat{x}$, is usually a nonconstant function—this is what the condition of "nonconstant self-similarity" requires.

Let us study the psychometric function $\Phi_y(x)$ of the reverse procedure, whereby y is being treated as a fixed referent and x as a varying probe. Assume that in the reverse procedure, the two stimulus domains, now denoted as $\mathscr{X}_c$ and $\mathscr{Y}_r$, are related to each other via another smooth and strictly increasing function $\phi : \mathscr{X}_c \to \mathscr{Y}_r$. An analogous expression for the psychometric function $\Phi_{\hat{y}}(x)$ of the reverse procedure can be derived using a strictly convex function $\Phi(\cdot)$

$$\Phi_{\hat{y}}(x) = \Phi_{\hat{y}}(\phi^{-1}(\hat{y})) + \mathscr{A}^{(\Phi)}(\hat{y}, x),$$

where $\mathscr{A}$ is defined earlier, and $\Phi_{\hat{y}}(\phi^{-1}(\hat{y}))$ is the minimal value of the psychometric function, as a function of the referent $\hat{y}$ now; it is in general a nonconstant function.

Now, the two convex functions Ψ and Φ, which are used to construct a pair of psychometric functions $\Psi_x(y)$ and $\Phi_y(x)$, are not necessarily related to each other in obvious ways. However, if one imposes the condition of regular cross-minimality, then it will be shown in this paper that the convex functions Ψ and Φ must be *conjugated*: $\Phi = \Psi^* \leftrightarrow \Phi^* = \Psi$, with $(\Psi^*)^* = \Psi, (\Phi^*)^* = \Phi$. In convex analysis (see Section A.1), a pair of conjugated convex functions Ψ and Ψ^* are related through the Legendre–Fenchel transformation. Furthermore, since their derivatives are inverse of one another, we obtain $\phi = \psi^{-1}$. This means that the regular cross-minimality condition, when applied to a referent–probe comparison task, forces a *consistent* diffeomorphic mapping between the two stimulus spaces $\mathscr{X}$ and $\mathscr{Y}$ regardless of which space contains reference and which comparison stimuli. The mapping $\phi = \psi^{-1} : \mathscr{X} \to \mathscr{Y} \longleftrightarrow \psi = \phi^{-1} : \mathscr{Y} \to \mathscr{X}$ can be any strictly increasing transformation, so the correspondence between the two stimulus spaces is allowed to be in a quite arbitrary form. Nevertheless, it must be the one and the same for both the forward and the reverse procedures; hence the subscripts indicating the reference/comparison role each space serves (r versus c) can be dropped. As a consequence, the psychometric differentials for the forward and reverse procedures are equal and have the following dualistic relation:

$$\mathscr{A}^{(\Psi)}(x, y) = \mathscr{A}^{(\Psi^*)}(y, x).$$

In the following, we extend these considerations from the unidimensional stimulus space to the multi-dimensional one, investigate the properties of regular cross-minimality and nonconstant self-similarity, and characterize the relationship between the pair of psychometric differentials arising from the forward and reverse procedures and the dual scaling of the two stimulus spaces. In Section 2.1, we show how regular cross-minimality amounts to requiring the one and same 1-to-1 mapping between the two stimulus spaces regardless of the assignment of the reference/comparison role to each. In Section 2.2, we investigate the dualistic expression of the psychometric differentials for the two procedures and show they become equal under regular cross-minimality. This allows the introduction of an asymmetrically defined *divergence function* between a referent and a probe, both expressed in the same stimulus space. Mathematically the dualistic expression of the nonnegative psychometric differential arises from the fundamental inequality (Fenchel inequality) associated with the conjugacy operation (Legendre transformation) in convex analysis. In Section 2.3, we investigate the global properties of the psychometric differential and the divergence function, and prove their nonnegativity, conjugacy, triangle relation, quadrilateral relation, and dualistic representability. In Section 2.4, the local property of the divergence function is studied; it is shown that this proposed measure of psychological dissimilarity, when considered in the small, induces a Riemannian metric indeed, a special case of the

[3]Since $\psi(\cdot)$ is strictly monotone increasing, the function $\Psi(y)$ must be strictly convex (see Section A.1 for a brief review of convex functions). Now, any strictly monotonic (possibly $\hat{x}$-dependent) transformation of $\Psi_{\hat{x}}(y)$ itself would not change the location of its minimum, $\Psi_{\hat{x}}(y)$ can thus be taken to represent either the psychometric/discrimination probability function or the "discrimination function"—so named to indicate that its range may lie outside [0, 1]—so long as the two are monotonically related to one another. Therefore, no further distinction will be made between a psychometric (discrimination probability) function and a discrimination function, and the terms will be used interchangeably.

Finslerian one. Finally, since our analysis hinges on a particular form of the psychometric function, we discuss its characterization from the perspective of biorthogonal mappings associated with a Riemannian manifold. The paper closes with a discussion about of the significance of the current framework of dual scaling of the comparison stimulus and the reference stimulus, and its difference from the Finslerian geometric framework of the stimulus space proposed by Dzhafarov and Colonius (1999, 2001). The main conclusion of the present analysis is to demonstrate that, while regular cross-minimality requires the correspondence between the two stimulus spaces be independent of the reference/comparison status each space is assigned to serve, one may still construct a pair of psychometric differentials, one for either assignment, which turn out to be identical yet respect the differential status of a reference stimulus and that of a comparison stimulus—the two stimulus spaces involved in the referent–probe comparison tasks are hence "dual" with respect to one another in this precise sense.

2. Stimulus dissimilarity and dual scaling of the reference and the comparison stimuli

In this section, a systematic framework is advanced to deal with the basic asymmetry in the two stimuli under referent–probe comparison, i.e., between the one serving as the comparison stimulus ("probe") and the one serving as the reference stimulus ("referent"). Following Dzhafarov and Colonius' proposal for the same–different comparison task, we advance the notions of (i) regular cross-minimality and (ii) nonconstant self-similarity in a pair of psychometric/discrimination function. The discrimination function under investigation can be expressed using a strictly convex (but otherwise arbitrary) function that satisfies nonconstant self-similarity in a straightforward way. Imposing the regular cross-minimality requirement across the two psychometric functions, each associated with a convex function, reveals that the two convex functions must be conjugated to one another. This in turn means that the scaling of either stimulus space, as measured by the (generally) asymmetric divergence functions, is dualistic with respect to the psychometric procedure (whether forward or reverse). Mathematically, regular cross-minimality and nonconstant self-similarity are properties that naturally arise from the fundamental duality (Legendre–Fenchel transformation) associated with the conjugacy operation on any convex functions.

2.1. Convex representation of psychometric functions

Within the context of a referent–probe comparison experiment, a trial consists of the presentation of a comparison stimulus and of a reference stimulus in two alternative observational intervals that are spatially and/or temporally separated (e.g., on two sides of the screen, in two successive frames). A *reference stimulus* or *referent* is one that is held constant during a block of experimental trials (in the method of staircase or the method of constant stimuli) or during a trial (in the method of adjustment). A *comparison stimulus* or *probe* is one that is varied in magnitude during its multiple presentations, in either ascending or descending order, or with randomization. The referent–probe distinction refers to the status of the actual physical stimuli being presented to the subjects. This status is understood by the subject as a result of either the delivery procedure and/or the mental representation of the stimulus. Therefore, any asymmetry that arises out of a switch of their status necessarily indicates the presence of some intrinsic difference in information processing by the subject, rather than merely a difference of observation intervals, whether spatial or temporal, within which the stimuli are presented.

To account for the fundamental difference in the referent and the probe, we would need to assume two distinct psychometric (discrimination probability) functions: $\Psi_x(y)$ for the forward procedure where x is presented in one interval as the referent and y in another interval as the probe, and $\Phi_y(x)$ for the reverse procedure where x is now allowed to vary and assumes the role of the probe while y is treated as a fixed referent. The two functions $\Psi_x(y)$ and $\Phi_y(x)$ may not be obviously related to each other. But we require them to satisfy regular cross-minimality and nonconstant self-similarity as described in Section 1.1.

The psychometric function $\Psi_x(y)$ is a bivariate function where $x = [x^1, \ldots, x^n]$ and $y = [y^1, \ldots, y^n]$ represent, respectively, the two stimuli used in comparison judgment, where each is defined in some subsets of R^n, denoted here and below as $\mathscr{X}$ and $\mathscr{Y}$. To denote that x is treated as the reference stimulus (referent) and y as the comparison stimulus (probe), one variable is written as the subscript and the other as the functional argument of the two-variable function $\Psi_x(y)$. In the framework of Dzhafarov and Colonius (1999), $\Psi_x(y)$ is taken to be the probability that y is judged as different from x. Here, a less restrictive interpretation is adopted: $\Psi_x(y)$ is taken to represent some "degree" of discrimination between a variable probe y and a fixed referent x, and need not be bounded between $[0, 1]$, as long as it is strictly monotonically related to the relative frequency that y is judged different from x in a discrimination task (the transformation itself can be x-dependent, see footnote 3).

Following the arguments developed in Section 1.1, we write the psychometric function in the form (c.f. Eq. (5))

$$\Psi_x(y) = C_1(x) + \mathscr{A}^{(\Psi)}(x, y), \tag{6}$$

where the multi-dimensional version of the psychometric differential (of the forward procedure) has the form

$$\mathscr{A}^{(\Psi)}(x,y) = \Psi(y) - x \cdot y + \Psi^*(x). \tag{7}$$

Here, $x \cdot y \equiv \sum_i^n x^i y^i$ denotes the dot product of two vectors x and y, $\Psi(\cdot) : R^n \to R$ is a smooth and strictly convex function with convex conjugate

$$\Psi^*(x) = (\nabla\Psi)^{-1}(x) \cdot x - \Psi((\nabla\Psi)^{-1}(x)) \tag{8}$$

(∇ denotes the gradient operation, see below). In (6), the function $C_1(\cdot) : R^n \to R$ is smooth but otherwise arbitrary. The rationale for writing the psychometric differential in the form of (7) will be further investigated in Section 2.5, where characterization results are provided.

With x as the fixed reference stimulus, $\Psi_x(y)$ is viewed as a function of the comparison stimulus y. The first derivative or *gradient* of $\Psi_x(y)$ with respect to y is

$$\nabla_y(\Psi_x(y)) = \nabla\Psi(y) - x,$$

where the gradient operator is $\nabla_y = [\frac{\partial}{\partial y^1}, \dots, \frac{\partial}{\partial y^n}]$ (for clarity, we omit the subscript in ∇ when the function being operated on has single variable). The first derivative is zero if and only if

$$x = \nabla\Psi(y).$$

The second derivative or *Hessian* of $\Psi_x(y)$, with respect to y again, is

$$\nabla^T(\nabla\Psi(y)) \equiv \mathrm{H}\Psi(y) \tag{9}$$

(∇^T is treated as a column vector, and ∇ as a row vector). Since $\mathrm{H}\Psi$ is positive semi-definite by the strict convexity property of the function $\Psi(\cdot)$, the mapping $y \mapsto x = (\nabla\Psi)(y)$ is 1-to-1 and hence invertible. Furthermore, for any fixed value of $x = \hat{x}$, since the Hessian of $\Psi_x(y)$ with respect to y is positive semi-definite due to positive semi-definiteness of $\mathrm{H}\Psi$, the point

$$\hat{y} = (\nabla\Psi)^{-1}(x)|_{x=\hat{x}} \tag{10}$$

is a unique, global minimum of $\Psi_{\hat{x}}(y)$ when the latter is viewed as a function of y:

$$\hat{y} = \mathrm{argmin}_y\, \Psi_{\hat{x}}(y).$$

The minimum value of such a discrimination function, for the fixed referent x, is

$$\min_y \Psi_x(y) = C_1(x).$$

So $C_1(x)$ is precisely the function of "self-similarity" in the notion of Dzhafarov (2002d), which varies according to the value of the reference stimulus x. Clearly, it need not be a constant.

Now turning to the reverse procedure in which x is treated as a comparison stimulus and hence variable while y is treated as the reference stimulus and hence fixed. Analogous to (5), the psychometric function $\Phi_y(x)$ is assumed to take the form:

$$\Phi_y(x) = C_2(y) + \mathscr{A}^{(\Phi)}(y,x), \tag{11}$$

where $\Phi(\cdot) : R^n \to R$ is another smooth and strictly convex function, and $\mathscr{A}^{(\Phi)}$ is simply

$$\mathscr{A}^{(\Phi)}(y,x) = \Phi(x) - y \cdot x + \Phi^*(y). \tag{12}$$

The function $C_2(\cdot) : R^n \to R$ is smooth but otherwise arbitrary; it is the minimal value function of the reverse discrimination function $\Phi_y(x)$ when minimizing over x, for any fixed referent y.

We can, for the fixed referent $y = \hat{y}$, find the value of $x = \check{x}$ that achieves the global minimum in this reverse discrimination task

$$\check{x} = \mathrm{argmin}_x\, \Phi_{\hat{y}}(x).$$

Differentiating the right side of (12) with respect to x, and setting $x = \check{x}, y = \hat{y}$ yields

$$\hat{y} = \nabla\Phi(\check{x}).$$

This is the equation that $\check{x}$ would have to satisfy.

Note that the condition of regular cross-minimality requires that

$$\check{x} \equiv \hat{x}.$$

Because of (10), we can see that

$$\nabla\Phi(\hat{x}) = (\nabla\Psi)^{-1}(\hat{x}).$$

Since $(\nabla\Psi)^{-1} = \nabla\Psi^*$ by the property of convex conjugacy (see Section A.2), we have

$$\Phi = \Psi^*$$

apart from a constant. Therefore, regular cross-minimality dictates that the convex function Φ in expression (11) for the psychometric function of the reverse procedure to be conjugated to the convex function Ψ in expression (6) for the psychometric function of the forward procedure; consequently the two psychometric differentials are equal

$$\mathscr{A}^{(\Psi)}(x,y) = \mathscr{A}^{(\Psi^*)}(y,x).$$

To summarize, using a pair of conjugated convex functions $\Psi(\cdot)$ and $\Psi^*(\cdot)$, we may express the two discrimination functions $\Psi_x(y)$ and $\Phi_y(x)$ (that characterize the degree of discrimination performance when x or y is fixed, respectively) by

$$\Psi_x(y) = \mathscr{A}^{(\Psi)}(x,y) + C_1(x),$$

$$\Phi_y(x) = \mathscr{A}^{(\Psi^*)}(y,x) + C_2(y),$$

such that (i) their global minima are both nonconstant functions $C_1(\cdot), C_2(\cdot)$ of the value of the fixed referent (subscript of the function)

$$\min_y \Psi_x(y) = C_1(x),$$

$$\min_x \Phi_y(x) = C_2(y)$$

and that (ii) the value of the probe (argument of the function) that attains their respective global minimum

$$\hat{y} = \mathrm{argmin}_y \, \Psi_{\hat{x}}(y),$$

$$\hat{x} = \mathrm{argmin}_x \, \Phi_{\hat{y}}(x),$$

satisfies

$$\hat{y} = \nabla \Psi^*(x)|_{x=\hat{x}},$$

$$\hat{x} = \nabla \Psi(y)|_{y=\hat{y}},$$

where $\nabla \Psi^* = (\nabla \Psi)^{-1}$. The use of a pair of conjugated functions Ψ and Ψ^* in representing the respective discrimination function $\Psi_x(y)$ and $\Phi_y(x)$ is a direct consequence of imposing the regular cross-minimality condition on the two psychometric functions.

2.2. Divergence functions and dual scaling of stimulus spaces

The psychometric differential $\mathscr{A}^{(\Psi)}(x, y)$, as given by (7), measures the difference in the large between a reference stimulus x and a comparison stimulus y (the term "differential" is really a misnomer since its value need not be infinitesimal). Under regular cross-minimality the two psychometric differentials for the forward and the reverse procedure the are equal. The basic convex duality $(\Psi^*)^* = \Psi$ results in the dualistic form of the psychometric differential

$$\mathscr{A}^{(\Psi^*)}(y, x) = \mathscr{A}^{((\Psi^*)^*)}(x, y) = \mathscr{A}^{(\Psi)}(x, y).$$

Psychometric differentials equal zero if and only if the functional arguements x and y satisfy

$$x = (\nabla \Psi)(y) \leftrightarrow \quad y = (\nabla \Psi^*)(x). \tag{13}$$

In such case, $\mathscr{A}^{(\Psi)}(x, y) = \mathscr{A}^{(\Psi^*)}(y, x) = 0$ simply expresses the Legendre–Fenchel duality (see Eq. (22) in Section A.1).

Denote $\psi(y) = (\nabla \Psi)(y)$ and $\phi(x) = \psi^{-1}(x) = (\nabla \Psi)^{-1}(x)$. We observe a diffeomorphic correspondence between the space $\mathscr{X}$ and the space $\mathscr{Y}$, namely $\phi : \mathscr{X} \to \mathscr{Y} \longleftrightarrow \psi : \mathscr{Y} \to \mathscr{X}$, regardless of which is being used to contain the comparison stimulus and which the reference stimulus (i.e., whether the forward or the reverse procedure):

$$\mathscr{X} \ni x \mapsto \phi(x) = y \in \mathscr{Y} \longleftrightarrow \mathscr{Y} \ni y \mapsto \psi(y) = x \in \mathscr{X}.$$

Because the mapping between the two spaces is continuous and 1-to-1, we can now express the psychometric differential *either* using values defined in $\mathscr{X}$ alone *or* in $\mathscr{Y}$ alone (i.e., both the comparison and the reference stimuli are now expressed in the same space). For stimulus values $y, y' \in \mathscr{Y}$ where y is the referent and y' is the probe, we define

$$\begin{aligned}\mathscr{D}^{(\Psi)}(y, y') &= \mathscr{A}^{(\Psi)}(\psi(y), y') \\ &= \Psi(y') - \psi(y) \cdot y' + \Psi^*(\psi(y)).\end{aligned}$$

Substituting the expression of $\Psi^*(\cdot)$ from (8) and rewriting $\psi(y) = (\nabla \Psi)(y)$, we have

$$\mathscr{D}^{(\Psi)}(y, y') = \Psi(y') - \Psi(y) - (\nabla \Psi)(y) \cdot (y' - y). \tag{14}$$

This defines a distance-like measure between a probe y' and a referent y both expressed in the same space. We call this measure $\mathscr{D}^{(\Psi)}(y, y')$ the (*psychological*) *divergence function* between two stimuli in the multidimensional stimulus space; it provides a scaling of the stimulus space $\mathscr{Y}$. Similarly, for two stimuli $x, x' \in \mathscr{X}$ serving as referent and probe, respectively, their dissimilarity is

$$\mathscr{D}^{(\Psi^*)}(x, x') = \mathscr{A}^{(\Psi^*)}(\phi(x), x'),$$

or writing out explicitly (recall $\nabla \Psi^*$ is simply $(\nabla \Psi)^{-1} = \psi^{-1} = \phi$)

$$\mathscr{D}^{(\Psi^*)}(x, x') = \Psi^*(x') - \Psi^*(x) - (x' - x) \cdot (\nabla \Psi^*)(x), \tag{15}$$

providing a scaling of the stimulus space $\mathscr{X}$. The duality between the two dissimilarity measures presents itself as the identity

$$\mathscr{D}^{(\Psi^*)}(x, x') = \mathscr{D}^{(\Psi)}(\phi(x'), \phi(x)),$$

$$\mathscr{D}^{(\Psi)}(y, y') = \mathscr{D}^{(\Psi^*)}(\psi(y'), \psi(y)).$$

Expressions (14) and (15) are formally identical since $(\Psi^*)^* = \Psi$. Henceforth, we will not distinguish, between Ψ and Ψ^*, the original versus the conjugated status of a convex function; we only need to note that the two stimulus spaces are in correspondence through the use of the gradients of a pair of mutually conjugated convex functions. In particular, we will freely write $\mathscr{D}^{(\Psi)}(x, x')$ to denote the divergence function for the stimulus space $\mathscr{X}$ as "scaled" by the convex function Ψ (c.f. Eq. (15))

$$\mathscr{D}^{(\Psi)}(x, x') = \Psi(x') - \Psi(x) - (\nabla \Psi)(x) \cdot (x' - x).$$

It is noteworthy that the psychometric differential defined above coincides in form with the so-called "canonical divergence" encountered in the analysis of the differential manifold structure of parametric probability distributions where the technique of Legendre transform is also used (Amari, 1985; Amari & Nagaoka, 2000). The psychological divergence function $\mathscr{D}^{(\Psi)}(x, x')$ has appeared in an entirely different context—in the field of optimization and machine learning—and is known as the Bregman (1967) divergence. Clearly, from the above analysis, it is the canonical divergence function $\mathscr{A}\Psi$ in disguise. It is called "divergence" as opposed to "distance", because $\mathscr{D}^{(\Psi)}(x, x')$ in general does not satisfy the symmetry axiom $\mathscr{D}^{(\Psi)}(x, x') \neq \mathscr{D}^{(\Psi)}(x', x)$ nor the axiom of triangle inequality. In fact, it will be shown elsewhere (Zhang, 2004) that a parametric family of dissimilarity (divergence) functions based on a single convex function can be constructed which include the current form as a special case.

Note that our present construction, like that of Dzhafarov and Colonius, is inherently multi-dimensional. When the stimulus dimensions are noninteracting (i.e., they are perceptually separable), then

$$\Psi(x) = \sum_{i=1}^{n} f(x^i)$$

for smooth and strictly convex $f : R \to R$, and

$$\nabla\Psi = [f'(x^1), f'(x^2), \ldots, f'(x^n)],$$

where f' denotes the ordinary derivative. This allows one to define a dissimilarity measure on each dimension separately

$$D^{(f)}(x^i, x'^i) = f(x'^i) - f(x^i) - (x'^i - x^i)f'(x^i),$$

such that the total dissimilarity is

$$\mathscr{D}^{(\Psi)}(x, x') = \sum_i D^{(f)}(x^i, x'^i).$$

This is the case of "perceptual separability" (Dzhafarov, 2002c) in the representation of psychological divergence functions.

2.3. Properties of psychometric differentials and divergence functions

The psychometric differential $\mathscr{A}^{(\Psi)}(x, y)$ defines a measure of dissimilarity between the stimulus x as a referent and a comparison stimulus y as a probe when the former is defined in one stimulus space $\mathscr{X}$ and the latter in another stimulus space $\mathscr{Y}$; its conjugate $\mathscr{A}^{(\Psi^*)}(y, x)$ measures the dissimilarity in the two stimuli when their respective roles are switched. It can be shown that the psychometric differential satisfies:

(A1) *Nonnegativity*: For all $x \in \mathscr{X}, y \in \mathscr{Y}$,

$$\mathscr{A}^{(\Psi)}(x, y) \geqslant 0,$$

with the equality holding if and only if $x = (\nabla\Psi)(y)$.

(A2) *Conjugacy*: For all $x \in \mathscr{X}, y \in \mathscr{Y}$,

$$\mathscr{A}^{(\Psi)}(x, y) = \mathscr{A}^{(\Psi^*)}(y, x).$$

(A3) *Triangle relation* (*generalized cosine*): For any three points $x \in \mathscr{X}$, $y, y' \in \mathscr{Y}$,

$$\mathscr{A}^{(\Psi)}(x, y) = \mathscr{A}^{(\Psi)}(x, y') + \mathscr{A}^{(\Psi)}(\nabla\Psi(y'), y) - (x - \nabla\Psi(y')) \cdot (y - y').$$

(A4) *Quadrilateral relation*: For any four points $x, x' \in \mathscr{X}$, $y, y' \in \mathscr{Y}$,

$$\mathscr{A}^{(\Psi)}(x, y') + \mathscr{A}^{(\Psi)}(x', y) - \mathscr{A}^{(\Psi)}(x, y) - \mathscr{A}^{(\Psi)}(x', y') = (x - x') \cdot (y - y').$$

As a special case, if x' and y' are in correspondence, i.e., $x' = (\nabla\Psi)(y')$, then $\mathscr{A}^{(\Psi)}(x', y') = 0$, and (A4) becomes the triangle relation (A3).

In parallel to properties enjoyed by $\mathscr{A}^{\Psi}$, the psychological divergence function $\mathscr{D}^{(\Psi)}(x, x')$, and similarly $\mathscr{D}^{(\Psi^*)}(y, y')$, satisfies the following properties:

(D1) *Nonnegativity*: For all $x, x' \in \mathscr{X}$,

$$\mathscr{D}^{(\Psi)}(x, x') \geqslant 0,$$

with the equality holding if and only if $x = x'$.

(D2) *Conjugacy*: For all $x, x' \in \mathscr{X}$,

$$\mathscr{D}^{(\Psi)}(x, x') = \mathscr{D}^{(\Psi^*)}(\nabla\Psi(x'), \nabla\Psi(x)).$$

(D3) *Triangle relation* (*generalized cosine*): For any three points $x, x', x'' \in \mathscr{X}$,

$$\mathscr{D}^{(\Psi)}(x, x') + \mathscr{D}^{(\Psi)}(x', x'') - \mathscr{D}^{(\Psi)}(x, x'') = (x'' - x') \cdot (\nabla\Psi(x) - \nabla\Psi(x')).$$

(D4) *Quadrilateral relation*: For any four points $x, x', x'', x''' \in \mathscr{X}$,

$$\mathscr{D}^{(\Psi)}(x, x') + \mathscr{D}^{(\Psi)}(x''', x'') - \mathscr{D}^{(\Psi)}(x, x'') - \mathscr{D}^{(\Psi)}(x''', x') = (x'' - x') \cdot (\nabla\Psi(x) - \nabla\Psi(x''')).$$

As a special case, when $x''' = x'$, $\mathscr{D}^{(\Psi)}(x''', x') = 0$, (D4) reduces to the triangle relation (D3).

Finally, we have the connection between the psychometric differential $\mathscr{A}^{\Psi}$ and the psychological divergence function $\mathscr{D}^{\Psi}$, namely, the *dualistic representability*:

$$\mathscr{D}^{(\Psi)}(x, x') = \mathscr{A}^{(\Psi)}((\nabla\Psi)(x), x') = \mathscr{A}^{(\Psi^*)}(x', (\nabla\Psi^*)^{-1}(x)).$$

The proof for (D1)–(D4) and for the dualistic representability is given in Section A.2. Since (A1)–(A4) are counterparts to (D1)–(D4), their proof is omitted.

2.4. Riemannian metric induced by the divergence functions

Recall that in the Fechnerian scaling of the stimulus space (Dzhafarov & Colonius, 1999), the Finslerian metric function $F(x, u)$ plays a pivotal role in linking a distance measure in the small and that in the large. Here $x = [x^1, \ldots, x^n]$ is a point on the base manifold and $u = [u^1, \ldots, u^n]$ is a nonzero vector on the tangent space (attached to the point x). For a given x, the set of u-vectors satisfying $F(x, u) = 1$ is called the indicatrix. The *metric tensor* g associated with a Finslerian metric function has components (see Rund, 1959)

$$g_{ij}(x, u) \equiv \frac{1}{2} \frac{\partial^2 (F(x, u))^2}{\partial u^i \partial u^j},$$

which can be shown to be positive semi-definite. Because of the Euler homogeneity

$$F(x, ku) = kF(x, u) \quad (k > 0),$$

the following identity can be derived:

$$(F(x, u))^2 = \sum_{i,j} g_{ij}(x, u) u^i u^j.$$

The right-hand side of the above equation resembles the expression of the line element under the Riemannian geometry, only that the metric tensor $g(x, u)$, called Finsler–Riemannn metric, now depends on u in addition to depending on x. Therefore, Finsler geometry can be seen as an extension of Riemannian geometry properly defined by the quadratic form of the line-element. On the other hand, when the Finslerian metric function $F(x, u) = F(u)$ does not depend on x, the Minkowski distance measure results. For these reasons, Finslerian metric function extends both the Riemannian metric and the Minkowski metric.

We next investigate local properties of the psychological divergence function introduced above. According to *The Fundamental Theorem* of Fechnerian scaling (Dzhafarov & Colonius, 1999, and more elaborated in Dzhafarov, 2002a), under a certain co-measurability condition, there exists a global psychometric transformation $g(\cdot)$ such that, when globally applied,

$$\lim_{s \to 0^+} \frac{g(\mathscr{D}(x, x + su))}{s},$$

exists and is to be identified with $F(x, u)$, the metric function of Finsler space. The "order" of such psychometric transformation is determined by the Taylor expansion of $\mathscr{D}(x, x + su)$ to the lowest nonvanishing term in s:

$$\begin{aligned}\mathscr{D}^{(\Psi)}(x, x + su) &= \Psi(x + su) - \Psi(x) - \nabla\Psi(x) \cdot (su) \\ &\approx \tfrac{1}{2} su^T \cdot (\mathrm{H}\Psi)(x) \cdot su,\end{aligned}$$

where H was defined in (9) as the Hessian matrix (second derivative) operator. With the square-root operation $g(t) = t^{1/2}$,

$$\lim_{s \to 0^+} \frac{g(\mathscr{D}^{(\Psi)}(x, x + su))}{s} = \left(\frac{1}{2}\, u^T (\mathrm{H}\Psi)(x) u\right)^{1/2}$$

Analogously, if one expands $\mathscr{D}(x + su, x)$, the same $F(x, u)$ results—the Finslerian metric function obtained is "balanced" using the terminology of Dzhafarov (2002d). The psychometric order is 2. Furthermore,

$$F(x, u) = \sqrt{\frac{1}{2} \sum_{ij} \frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j} u^i u^j}$$

is actually a Riemannian metric function. So, the psychological dissimilarity measure, though asymmetric in the large, actually induces (i.e., is compatible with) a symmetric Riemannian metric in the small.

For reader's information, we may also calculate the "level function"—the ω-function as defined in Dzhafarov's (2002d),

$$\Psi_{x+su}(x + su) - \Psi_x(x) \simeq s \sum_{i,j} \frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j}\, x^i u^j,$$

$$\Phi_{x+su}(x + su) - \Phi_x(x) \simeq s \sum_{i,j} \frac{\partial^2 \Psi^*(x)}{\partial x^i \partial x^j}\, x^i u^j.$$

They are both in the first order of s, though the psychometric differentials $\mathscr{D}^{(\Psi)}(x, x + su)$, $\mathscr{D}^{(\Psi)}(x + su, x)$, $\mathscr{D}^{(\Psi^*)}(x, x + su)$, $\mathscr{D}^{(\Psi^*)}(x + su, x)$ all approach zero in the second order of s (and hence with psychometric order of 2). This is not inconsistent with the results of Dzhafarov (2002d), see further discussions in Section 3.

2.5. Biorthogonality and characterization of dualistic psychometric differentials

The divergence function defined on the stimulus space is different from the distance function defined on it. Recall the three axioms of the distance function (see Section 1) and contrast them with the properties of the divergence function $\mathscr{D}^{(\Psi)}$ or its equivalence $\mathscr{A}^{(\Psi)}$. Notably, the nonnegativity axiom is replaced by a modified requirement of $\mathscr{A}^{(\Psi)}(x, y) \geqslant 0$, with 0 achieved if and only if $y = \psi(x)$ for some diffeomorphic (1-to-1 and continuous) transformation $\psi(\cdot)$ in multi-dimensional case or strictly increasing transformation in one-dimensional case. The symmetry axiom is replaced by a dual symmetry: $\mathscr{A}^{(\Psi)}(x, y) = \mathscr{A}^{(\Psi^*)}(y, x)$ with Ψ^* satisfying $(\Psi^*)^* = \Psi$. Lastly, triangle inequality is replaced by the triangle relation (generalized cosine law). These properties of a divergence function differ sharply from a distance function $\Delta(x, y)$ which enjoys nonnegativity, symmetry, and triangle inequality.

This paper investigated a particular form of the discrimination function (6), which results in the specific form of the psychometric differential (7) or equivalently the divergence function in the form (14). As mentioned earlier, in the unidimensional case, such form (5) arises out of the presumed existence of a smooth, strictly increasing (i.e., order-preserving) transformation between the two stimulus spaces (Section 1.1). In the multi-dimensional case, monotonicity of the mapping no longer makes sense. Therefore, the assumption of smooth, order-preserving transformations is to be replaced by some other restrictions on the class of permissible diffeomorphic transformations $\psi : \mathscr{Y} \to \mathscr{X}$ which, by definition, must have nonvanishing Jacobian. In fact, if we require the Jacobian to be symmetric (here ψ^i and ψ^j denote the ith and jth component of the vector-valued map ψ)

$$\frac{\partial \psi^i}{\partial y^j} = \frac{\partial \psi^j}{\partial y^i}, \tag{16}$$

as well as positive-definite, then convex analysis tells us that there must exist a strictly convex function Ψ such that $\psi = \nabla\Psi$. This is to say, the assumption (in Section 2.1) of the existence of

$$x = \psi(y) = \phi^{-1}(y) = \nabla\Psi(y)$$
$$\longleftrightarrow y = \psi^{-1}(x) = \phi(x) = \nabla\Psi^*(x)$$

amounts to a requirement (16) that the mapping from $\mathscr{Y}$ to $\mathscr{X}$ be "curl"-less.

When points in the two spaces $\mathscr{X}$ and $\mathscr{Y}$ are 1-to-1 identified in such a way, i.e., when they satisfy (16), x and y may be viewed as two coordinate representations of some underlying abstract manifold (remember a manifold is locally differeomorphic to a subset of R^n, in this case $\mathscr{X}$ or $\mathscr{Y}$). The metric tensor g of the underlying manifold is then given by (in matrix form)

$$ij\text{th element of } \mathrm{g} = \frac{\partial \psi^i}{\partial y^j} = \frac{\partial \psi^j}{\partial y^i},$$

whose inverse is

$$ij\text{th element of } \mathrm{g}^{-1} = \frac{\partial \phi^i}{\partial x^j} = \frac{\partial \phi^j}{\partial x^i}.$$

The two coordinate systems x and y are said to be "biorthogonal" with respect to the metric. The manifold possesses, in addition to a Riemannian metric, a pair of dually flat affine connections (related to Ψ and Ψ^*) that can be used for parallel transport of vectors. Such pair of connections, though not satisfying the Levi–Civita condition and hence incompatible with the Riemannian metric, are dually flat in the sense that a unique, canonical divergence function exists between any two points on the manifold. This is precisely the (asymmetric) psychometric differential/divergence functions discussed in Section 2.2. These important concepts (biorthogonal coordinates, dual connections, canonical divergence) have been advanced in the field of "information geometry"(Amari, 1985; Amari & Nagaoka, 2000), which studies the manifold of all probability functions where the Fisher information defines the Riemannian metric.

So far, we have characterized the dual correspondence between $\mathscr{X}$ and $\mathscr{Y}$ through the gradients of a pair of conjugated convex functions Ψ and Ψ^*. To characterize psychometric functions $\Psi_x(y)$ which take the exact form of (6), we examine, equivalently, the characterization of divergence functions $\mathscr{D}^{(\Psi)}(x, x')$ which take the exact form (14).

Kaas and Vos (1997, p. 240) showed that (14) arises uniquely if the following three conditions are satisfied (i) $\mathscr{D}^{(\Psi)}(x, x') \geqslant 0$, the equality holding if and only if $x = x'$; (ii) the first derivatives $\nabla_x \mathscr{D}^{(\Psi)}(x, x')|_{x'=x} = \nabla_{x'} \mathscr{D}^{(\Psi)}(x, x')|_{x=x'} = 0$; and (iii) the second derivative (Hessian) $\mathrm{H}_{x'} \mathscr{D}^{(\Psi)}(x, x')$ is positive semi-definite in x' and independent of x. (Here $\nabla_{x'}, \mathrm{H}_{x'}$ denote the first and the second derivative operators with respect to the the x'-variable). Eguchi (1983), in his theory on divergence (called also contrast) functions, did not use (iii) above, but stipulated instead that (iii)' the mixed second derivative $(\nabla_x)^T(\nabla_{x'})\mathscr{D}^{(\Psi)}(x, x')$, which is necessarily symmetric, be negative semi-definite. The reader may verify the above claims by simply noting the following relations:

$$\nabla_{x'} \mathscr{D}^{(\Psi)}(x, x') = \nabla\Psi(x') - \nabla\Psi(x),$$

$$\nabla_x \mathscr{D}^{(\Psi)}(x, x') = (x - x') \cdot (\mathrm{H}\Psi)(x),$$

$$\mathrm{H}_{x'} \mathscr{D}^{(\Psi)}(x, x') = (\mathrm{H}\Psi)(x').$$

3. Discussion

Geometrization of the multi-dimensional stimulus space has always captured the interest of mathematical psychologists. Differential geometric descriptions of the perceptual space, excluding the large body of work on binocular depth perception where geometry is substantively involved (Luneburg, 1947; Smith, 1959; Indow, 1982, 1991), have recently been applied to the discrimination and comparison of stimuli in multi-dimensional setting invoking the notion of affine connection for vector comparison (Yamazaki, 1987; Levine, 2000), the emergence of perceptual oneness in segregated objects invoking the notion of intrinsic parallelism (Zhang & Wu, 1990; Zhang, 1995), the perception of complex visual stimuli through infinite-dimensional analysis (Townsend, Solomon, & Smith, 2001), and the characterization of perceptual distance in the large via local measurement of stimulus discrimination using the Finsler geometry approach (Dzhafarov & Colonius, 1999, 2001). Here, in the same spirit, the foundation of comparative judgment between a probe (comparison stimulus) and a referent (reference stimulus) is investigated, with a special interest in the issue of asymmetry in comparison.

Scalings for the comparison stimulus space and for the reference stimulus space that are dual to each other are formally constructed. The duality between the comparison and the reference stimuli, which are nevertheless in diffeomorphic correspondence, gives rise to a difference structure characterized by an asymmetric (but dually symmetric) measure of dissimilarity, in contrast to the symmetric distance measure of similarity. The nonnegativity axiom is relaxed to state that $\mathscr{A}^{(\Psi)}(x, y) \geqslant 0$ with 0 attained iff $y = \psi(x)$ for some diffeomorphic (1-to-1 and continuous) transformation

$\psi = \nabla\Psi$. In the unidimensional case, the requirement of diffeomorphic transformation is reduced to the assumption of a monotonic (order-preserving) mapping between the two stimulus spaces $\mathscr{X}$ and $\mathscr{Y}$. In lieu of the symmetry axiom, a dualistic relation is in place $\mathscr{A}^{(\Psi)}(x,y) = \mathscr{A}^{(\Psi^*)}(y,x)$ with Ψ^* satisfying $(\Psi^*)^* = \Psi$. In lieu of the triangle inequality, the triangle relation (generalized cosine) is satisfied by $\mathscr{A}^{\Psi}$. These properties of the psychometric differential (or its alternative form, the psychological divergence function) have been shown to be consistent with the property of regular cross-minimality without violating the property of nonconstant self-similarity in the psychometric functions themselves; in fact regular cross-minimality dictates that the differomorphic mapping between the two spaces $\mathscr{X}$ and $\mathscr{Y}$ to be the one and the same regardless of which is taken to contain the reference stimulus and which the comparison stimulus, and the psychometric differentials for either assignment to be identical.

Asymmetric measures of dissimilarity are perhaps, we claim, more important for a stimulus space than perhaps the symmetric, distance measure of similarity[4] because of the pervasiveness of asymmetry in comparative judgment $\Psi_x(y) \neq \Phi_y(x)$. The differential status of a fixed referent (status quo) and a variable probe is of fundamental importance to any theory of the underlying psychological process of comparative judgment, for example between one stimulus in perception and another in memory during categorization, between the current state and the goal state during planning, between the frame of reference (anchoring point) and the potential targets to be mentally searched during problem solving, between the known status quo and the uncertain gains or losses during decision-making, etc. In the framework of Dzhafarov and Colonius (1999, 2001), asymmetry in same–different comparison is possible because, in general, the Fechnerian metric function in the small $F(x,u)$ can be made to be asymmetric: $F(x,-u) \neq F(x,u)$. Put in another way, the Finsler geometry generalizes the conventional Riemannian geometry by allowing directional dependence in the metric tensor (in component form) g_{ij} that define the quadratic line element $ds^2 = \sum_{i,j} g_{ij}(x,u)u^i u^j$. Since such a mechanism works for unidimensional scaling as well, asymmetry is seen to arise fundamentally as the "building-up" of local, asymmetric psychometric judgments in the small, as opposed to arising from the multidimensional nature of the stimuli, as in the contrast model of similarity (Tversky, 1977). In the present framework, asymmetry in the discrimination function in the large does not arise as a consequence of local asymmetry—in fact the local, Riemannian metric induced is a symmetric one. Asymmetry in comparative judgments in the dualistic psychometric procedures arises from an additive, nonconstant self-similarity term reflecting the properties of the reference stimulus; the psychometric differentials themselves are dually symmetric. The divergence function on the one hand respects the differential status of a variable probe and a fixed referent; on the other hand it is expressible in one of the two dually equivalent forms, one for each stimulus space. To the extent that self-similarity is related to the typicality of a stimulus within a category (as discussed in Smith, 1995), our account is different from the contrast model of Tversky (1977) which attributes the asymmetry in performance to distinct features a comparison stimulus and a reference stimulus each possesses and the differential weighting among the unique features.

It is worth pointing out that, from a purely mathematical standpoint, the framework of Dzhafarov and Colonius and the current framework do not subsume one another. Dzhafarov and Colonius (1999, 2001) models discrimination probability by a single two-variable function $\Psi(x,y)$ that is quite arbitrary to begin with (apart from a certain co-measurability-in-the-small condition), whereas the current framework allows two two-variable functions while restricting them to a particular representation involving a convex function. Applying the conditions of regular minimality and nonconstant self-similarity to Dzhafarov and Colonius' model would rule out certain functional forms of $\Psi(x,y)$ (Dzhafarov, 2003a,b) and also restrict the psychometric order associated with $\Psi(x,y)$ to be less than or equal to 1 (Dzhafarov, 2002d). Applying the condition of regular cross-minimality to the two psychometric functions $\Psi_x(y)$ and $\Phi_y(x)$ here results in one and the same psychometric differential that can be cast in dualistic forms, while the condition of nonconstant self-similarity is satisfied by construction. So mathematically the two frameworks should be viewed as special cases of the most general situation where $\Psi_x(y)$ and $\Phi_y(x)$ are arbitrary and nonidentical, with Dzhafarov and Colonius' framework assuming an identical *psychometric function*

$$\Psi_x(y) = \Phi_y(x) \tag{17}$$

and the current framework resulting in an identical *psychometric differential*

$$\Psi_x(y) - \min_y \Psi_x(y) = \Phi_y(x) - \min_x \Phi_y(x). \tag{18}$$

That these two frameworks are nonintersecting can be appreciated when one tries to force the pair of psychometric functions studied here (i.e. satisfying

[4] Our form of divergence function includes the Euclidean metric as a special case, but is distinct from all other Minkowski metric functions. In fact, when in the large, Euclidean metric is the *only* common element that is both a (necessarily symmetric) distance function and a (generally asymmetric) divergence function.

(18)) to be identical (i.e. satisfying (17))—it turns out that requiring both (17) and (18) leaves no other choice but $\min_y \Psi_x(y) = \min_x \Phi_y(x) = const$. In other words, nonconstant self-similarity must be violated in this situation (this fact was first pointed out to the author by Dzhafarov in his review of an earlier draft of this paper). From the perspective of Dzhafarov and Colonius' framework, imposing (17) on the pair of psychometric functions requires the psychometric procedure to be such that stimulus $x \in \mathscr{X}$ is always presented in the first observation area and stimulus $y \in \mathscr{Y}$ in the second observation area. In this case, the regular cross-minimality condition across the two functions $\Psi_x(y)$ and $\Phi_y(x)$ is enforced as the regular minimality condition across the two variables of a single function $\Psi(x, y) = \Psi_x(y) = \Phi_y(x)$. The forms of the psychometric functions studied in this paper, which allow for a dualistic form of psychometric differential, are simply too specially chosen and would have constant self-similarity. (Note: the reader should not be confused that nonconstant self-similarity is always satisfied and in fact comes for free so long as $\Psi_x(y)$ and $\Phi_y(x)$ are not required to be the one and the same.) This analysis shows that the Dzhafarov and Colonius' framework and the current framework *cannot* reduce to one another; hence there is no mathematical inconsistency in the respective conclusions regarding the psychometric order.[5]

While the two mathematical approaches differ, both have extended the Riemannian geometrical models of the stimulus space, a tradition that dates back to as early as Schrödinger(1920a,b) and later Stiles (1946) on color perception. The Finsler geometric approach (Dzhafarov & Colonius, 1999, 2001) allowed for locally non-Riemannian metric function, and defined distance through integration-and-minimization ("path-integral") procedure; this is based on the basic mathematical fact that there exists a correspondence between the Finslerian distance in the large and the Finslerian metric function in the small provided that the distance function is internal or internalizable. The present approach, on the other hand, is rooted in dually affine Riemannian geometry—the latter still uses a locally (symmetric) Riemannian metric function, but allows for nonmetrical affine connections for parallel transport of vectors. This approach is based on the basic mathematical fact that there exists a pair of global, biorthogonal coordinates, connected through the Legendre transform, that serve as geodesics of the Riemannian manifold with a pair of conjugate connections. So long as the pair of conjugated affine connections are dually flat, two dually symmetric divergence functions can be constructed in the large. Both approaches extended classical Riemannian geometry in different directions. Further research is needed to understand the relationship between these two mathematical approaches in order to derive a unified and coherent geometric theory of the psychological space within which a pair of stimuli are comparable for similarity and for difference.

[5] Dzhafarov's conclusion of psychometric order being not more than 1 is based on the psychometric function satisfying both regular minimality and nonconstant self-similarity (Dzhafarov, 2002d). Psychometric order is no longer constrained if nonconstant self-similarity condition is removed.

Appendix A

A.1. Convex sets, convex functions, and convex conjugacy

A point set $\mathscr{S} \subset R^n$ is called convex if for any two points $x, x' \in \mathscr{S}$ and any real number $\lambda \in (0, 1)$,

$$\lambda x + (1 - \lambda)x' \in \mathscr{S},$$

this is to say, the line segment connecting any two points x and x' belongs to the set $\mathscr{S}$. In general, a multivariate function $\Psi : \mathscr{S} \mapsto R$ is *strictly convex* (or simply *convex*) if for all $x, x' \in \mathscr{S}$ and $\lambda \in (0, 1)$

$$\Psi(\lambda x + (1 - \lambda)x') \leqslant \lambda \Psi(x) + (1 - \lambda)\Psi(x'),$$

where the equality holds *only when* $x = x'$. The Hessian matrix of a smooth, strictly convex function $\Psi(x)$ is positive semi-definite, i.e., for all $u \in R^n$,

$$u^T \cdot (\mathrm{H}\Psi)(x) \cdot u \geqslant 0.$$

For a general introduction to convex analysis, see Rockafellar (1970) or Hiriart-Urruty and Lemaréchal (1993). In particular, for a convex function $\Psi(\cdot)$, the gradient map $x \mapsto (\nabla\Psi)(x) = y$ establishes a diffeomorphism between $\mathscr{S}$ and its dual gradient space $\mathscr{S}^* \ni y$.

The function $\Psi^*(\cdot)$ constructed according to

$$\Psi^*(x) = x \cdot (\nabla\Psi)^{-1}(x) - \Psi((\nabla\Psi)^{-1}(x)), \tag{19}$$

is known as the conjugate function of $\Psi(\cdot)$. It enjoys many interesting properties. In particular, Ψ^* is also a convex function. Calculating its first derivative, and introducing the vector-valued function

$$\phi(x) \equiv (\nabla\Psi)^{-1}(x),$$

we have

$$\begin{aligned}\nabla\Psi^*(x) &= \nabla\{x \cdot \phi(x) - \Psi(\phi(x))\} \\ &= \phi(x) + x \cdot \nabla\phi(x) - (\nabla\Psi)(\phi(x)) \cdot \nabla\phi(x) \\ &= \phi(x) = (\nabla\Psi)^{-1}(x),\end{aligned}$$

where the identity

$$\nabla\Psi(\phi(x)) = (\nabla\Psi)((\nabla\Psi)^{-1}(x)) \equiv x$$

has been used. This shows that the mappings $y \mapsto x = \nabla\Psi(y)$ and $x \mapsto y = \nabla\Psi^*(x)$ are inverse functions of each other:

$$\nabla\Psi^* = (\nabla\Psi)^{-1} \longleftrightarrow \nabla\Psi = (\nabla\Psi^*)^{-1}. \tag{20}$$

So the convex function $\Psi(\cdot)$ and its conjugate $\Psi^*(\cdot)$ as defined by (19) form a pair. It can be verified that the convex conjugate of $\Psi^*(\cdot)$ is in turn $\Psi(\cdot)$:

$$(\Psi^*)^* = \Psi. \tag{21}$$

As examples, the functions $\frac{1}{p}(\sum_i (|x|^i)^p)^{\frac{1}{p}}$, $p>1$ and the functions $\frac{1}{q}(\sum_i (|x|^i)^q)^{\frac{1}{q}}$, $q>1$ are conjugate convex functions when $\frac{1}{p}+\frac{1}{q}=1$. Quadratic functions ($p=q=2$) are self-conjugated.

A rearrangement of (19) or a replacement of x in (19) by $\nabla\Psi(y)$ yields, respectively, the following identity:

$$\Psi^*(x) + \Psi(\nabla\Psi^*(x)) = x \cdot \nabla\Psi^*(x), \tag{22a}$$

$$\Psi(y) + \Psi^*(\nabla\Psi(y)) = y \cdot \nabla\Psi(y). \tag{22b}$$

These equalities between a convex function $\Psi(\cdot)$ and its convex conjugate $\Psi^*(\cdot)$ are called *Legendre–Fenchel duality* in convex analysis (see, e.g., Rockafellar, 1970). They represent the fundamental duality between the vector space where x is defined, and the conjugate gradient space where $y = (\nabla\Psi^*)(x)$ is defined.

A.2. Proof of properties of the psychological divergence function (Section 2.3)

Proof of conjugacy:

$$\begin{aligned}\mathscr{D}^{(\Psi)}(x,x') &= \Psi(x') - \Psi(x) - (x'-x)\cdot\nabla\Psi(x)\\ &= (x'\cdot\nabla\Psi(x') - \Psi^*(\nabla\Psi(x')))\\ &\quad - (x\cdot\nabla\Psi(x) - \Psi^*(\nabla\Psi(x))) - (x'-x)\cdot\nabla\Psi(x)\\ &= \Psi^*(\nabla\Psi(x)) - \Psi^*(\nabla\Psi(x'))\\ &\quad - x'\cdot(\nabla\Psi(x) - \nabla\Psi(x'))\\ &= \Psi^*(\nabla\Psi(x)) - \Psi^*(\nabla\Psi(x'))\\ &\quad - \nabla\Psi^*(\nabla\Psi(x'))\cdot(\nabla\Psi(x) - \nabla\Psi(x'))\\ &= \mathscr{D}^{(\Psi^*)}(\nabla\Psi(x'),\nabla\Psi(x)).\end{aligned}$$

Proof of triangle and quadrilateral relations:
From

$$\mathscr{D}^{(\Psi)}(x,x') = \Psi(x') - \Psi(x) - (x'-x)\cdot\nabla\Psi(x),$$

$$\mathscr{D}^{(\Psi)}(x',x'') = \Psi(x'') - \Psi(x') - (x''-x')\cdot\nabla\Psi(x'),$$

$$\mathscr{D}^{(\Psi)}(x,x'') = \Psi(x') - \Psi(x) - (x''-x)\cdot\nabla\Psi(x),$$

we easily derive the triangular (generalized cosine) relation by adding the first two equations and subtracting the third. To prove the quadrilateral relation, we write out

$$\begin{aligned}\mathscr{D}^{(\Psi)}(x,x'') &= \mathscr{D}^{(\Psi)}(x,x') + \mathscr{D}^{(\Psi)}(x',x'')\\ &\quad - (x''-x')\cdot(\nabla\Psi(x) - \nabla\Psi(x')),\end{aligned}$$

which holds for all x' (therefore the right-hand side is independent of x'). Replacing x with another variable x''':

$$\begin{aligned}\mathscr{D}^{(\Psi)}(x''',x'') &= \mathscr{D}^{(\Psi)}(x''',x') + \mathscr{D}^{(\Psi)}(x',x'')\\ &\quad - (x''-x')\cdot(\nabla\Psi(x''') - \nabla\Psi(x')).\end{aligned}$$

Subtracting these two equations and rearranging yields the quadrilateral relation.

Proof of dualistic representability:

$$\begin{aligned}\mathscr{D}^{(\Psi)}(x,x') &= \Psi(x') - \Psi(x) - (x'-x)\cdot\nabla\Psi(x)\\ &= \Psi(x') - x'\cdot\nabla\Psi(x) + (x\cdot\nabla\Psi(x) - \Psi(x))\\ &= \Psi(x') - x'\cdot\nabla\Psi(x) + \Psi^*(\nabla\Psi(x))\\ &= \mathscr{A}^{(\Psi)}(\nabla\Psi(x),x')\\ &= \mathscr{A}^{(\Psi^*)}(x',\nabla\Psi(x))\\ &= \mathscr{A}^{(\Psi^*)}(x',(\nabla\Psi^*)^{-1}(x)).\end{aligned}$$

References

Amari, S., 1985. *Differential geometric methods in statistics*, Lecture notes in statistics, Vol. 28, New York: Springer (reprinted in 1990).

Amari, S., & Nagaoka, H., 2000. *Method of information geometry*. AMS monograph, Oxford: Oxford University Press.

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Physics*, *7*, 200–217.

Dzhafarov, E. N. (2002a). Multidimensional Fechnerian scaling: regular variation version. *Journal of Mathematical Psychology*, *46*, 226–244.

Dzhafarov, E. N. (2002b). Multidimensional Fechnerian scaling: probability-distance hypothesis. *Journal of Mathematical Psychology*, *46*, 352–374.

Dzhafarov, E. N. (2002c). Multidimensional Fechnerian scaling: perceptual separability. *Journal of Mathematical Psychology*, *46*, 564–582.

Dzhafarov, E. N. (2002d). Multidimensional Fechnerian scaling: pairwise comparisons, regular minimality, and nonconstant self-similarity. *Journal of Mathematical Psychology*, *46*, 583–608.

Dzhafarov, E. N. (2003a). Thurstonian-type representations for "same–different" discriminations: deterministic decisions and independent images. *Journal of Mathematical Psychology*, *47*, 208–228.

Dzhafarov, E. N. (2003b). Thurstonian-type representations for "same–different" discriminations: probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, *47*, 229–243.

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychological Bulletin and Review*, *6*, 239–268.

Dzhafarov, E. N., & Colonius, H. (2001). Multidimensional Fechnerian scaling: basics. *Journal of Mathematical Psychology*, *45*, 670–719.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Annals of Statistics*, *11*, 793–803.

Hershey, J. C., & Schoemaker, P. J. H. (1985). Probability versus certainty equivalence methods in utility measurement: are they equivalent? *Management Science*, *31*, 1213–1231.

Hiriart-Urruty, J.-B., & Lemaréchal, C., 1993. *Convex analysis and minimization algorithms*. Vols. I and II. Berlin: Springer.

Indow, T. (1982). An approach to geometry of visual space with no a priori mapping functions: multidimensional mapping according to Riemannian metrics. *Journal of Mathematical Psychology*, *26*, 204–236.

Indow, T. (1991). A critical review of Luneburg's model with regard to global structure of visual space. *Psychological Review*, *98*, 430–453.

Kaas, R. E., & Vos, P. W. (1997). *Geometric foundation of asymptotic inference*. New York: Wiley.

Levine, D. N. (2000). A differential geometric description of the relationships among perceptions. *Journal of Mathematical Psychology*, *44*, 241–284.

Luce, R. D., & Galanter, E., 1963. Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1 (pp. 191–243). New York: Wiley.

Luneburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton, NJ: Princeton University Press.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254–278.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, *86*, 161–180.

Polk, T. A., Behensky, C., Gonzalez, R., & Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: asymmetries induced by manipulating exposure frequency. *Cognition*, *82*, B75–B88.

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

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532–547.

Rund, H. (1959). *The differential geometry of Finsler spaces*. Berlin: Springer.

Schoemaker, P. J. H., & Hershey, J. C. (1992). Utility measurement: signal, noise, and bias. *Organization Behavior and Human Decision Processes*, *52*, 397–424.

Schrödinger, E., 1920a. Outline of a theory of color measurement for daylight vision. *Annelen der Physik* 63, 397–447. (English translation in D. L. MacAdam (Ed.). 1970. *Sources of color science* (pp. 134–182). Cambridge, MA: MIT Press.)

Schrödinger, E., 1920b. Thresholds of color differences. *Annelen der Physik*, *63*, 481–520 (English translation in D.L. MacAdam (Ed.), 1970. *Sources of color science* (pp. 183–193). Cambridge, MA: MIT Press.)

Shepard, R. N. (1962a). The analysis of proximities: multidimensional scaling with an unknown distance function I. *Psychometrika*, *27*, 125–140.

Shepard, R. N. (1962b). The analysis of proximities: multidimensional scaling with an unknown distance function II. *Psychometrika*, *27*, 219–246.

Smith, A. A. (1959). The geometry of visual space. *Psychological Review*, *66*, 334–337.

Smith, E. E., 1995. Concepts and categorization. In E. Smith, & D. Osherson (Eds.), *An invitation to cognitive science*: *thinking*. Vol. 3 (pp. 3–34). Cambridge, MA: MIT Press.

Stiles, W. S. (1946). A modified Helmholtz line-element in brightness-colour space. *Proceedings of the Physical Society of London*, *58*, 41–65.

Townsend, J. T., Solomon, B., & Smith, J. S., 2001. The perfect Gestalt: infinite dimensional Riemannian face spaces and other aspects of face perception. In M. J. Wenger, & J. T. Townsend (Eds.), *Computational*, *geometric*, *and process perspectives of facial Cognition*: *contexts and challenges* (pp. 39–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, *89*, 123–154.

Yamazaki, T. (1987). Non-Riemannian approach to geometry of visual space: an application of affinely connected geometry to visual alleys and horopter. *Journal of Mathematical Psychology*, *31*, 270–298.

Zhang, J. (1995). Motion detectors and motion segregation. *Spatial Vision*, *9*, 261–273.

Zhang, J. (2004). Divergence function, duality, and convex analysis. *Neural Computation*, *16*, 159–195.

Zhang, J., & Wu, S. (1990). Structure of visual perception. *Proceedings of National Academy of Sciences, USA*, *87*, 7819–7823.

Zimmer, K., & Colonius, H. (2000). Testing a new theory of Fechnerian scaling: the case of auditory intensity discrimination. *Paper presented at the 140th meeting of the acoustical society of America*.