

On the Mathematical Properties of the Structural Similarity Index

Dominique Brunet, Edward R. Vrscay, *Member, IEEE*, and Zhou Wang, *Member, IEEE*

Abstract—Since its introduction in 2004, the structural similarity (SSIM) index has gained widespread popularity as a tool to assess the quality of images and to evaluate the performance of image processing algorithms and systems. There has been also a growing interest of using SSIM as an objective function in optimization problems in a variety of image processing applications. One major issue that could strongly impede the progress of such efforts is the lack of understanding of the mathematical properties of the SSIM measure. For example, some highly desirable properties such as convexity and triangular inequality that are possessed by the mean squared error may not hold. In this paper, we first construct a series of normalized and generalized (vector-valued) metrics based on the important ingredients of SSIM. We then show that such modified measures are valid distance metrics and have many useful properties, among which the most significant ones include quasi-convexity, a region of convexity around the minimizer, and distance preservation under orthogonal or unitary transformations. The groundwork laid here extends the potentials of SSIM in both theoretical development and practical applications.¹

Index Terms—Cone metrics, normalized metrics, perceptually optimized algorithms and methods, quality metrics and assessment tools, quasi-convexity and convexity, structural similarity (SSIM) index.

I. INTRODUCTION

IMAGE similarity (or dissimilarity) assessment is a fundamental issue in many signal processing applications. Traditionally, mean squared error (MSE), or the L^2 metric, has been the standard measure not only for the evaluation of signal fidelity and quality but also for the design and optimization of signal processing algorithms and systems [1]. MSE is easy to

compute and understand. It also has many desirable mathematical properties [1], i.e., its square root is a valid distance metric in \mathbf{R}^N ; it is convex, symmetric, and differentiable; it is distance preserving under orthogonal and unitary transformations; and it is additive for statistically independent sources of distortions. All of these make MSE an ideal choice for optimization purposes, where analytic linear solutions can be often found. Even when closed-form solutions are difficult, it is often convenient to formulate iterative numerical optimization algorithms as the gradient and the Hessian matrix of the MSE are also easy to compute.

Unfortunately, MSE (as well as any other L^p metric) turns out to be an extremely poor measure when the purpose is to predict perceived signal fidelity or quality [1]. A research topic that has attracted a great deal of attention in the past decade is to design novel objective image similarity/dissimilarity measures that correlate well with perceptual image fidelity/distortion [2]. The progress has been exciting as a handful of objective measures such as the structural similarity (SSIM) index [3], [4] and its derivatives, multiscale SSIM (MS-SSIM) [5] and information content-weighted SSIM (IW-SSIM) [6], the visual information fidelity [7], the visual signal-to-noise ratio [8], the most apparent distortion measure [9], and the feature similarity index [10] significantly and consistently outperformed MSE and its direct derivative, the peak signal-to-noise ratio, in terms of correlations with subjective image quality evaluations in extensive tests using a number of independent large-scale subject-rated image databases [6], [10]. While such achievement is remarkable and these new perceptual measures are satisfying in image quality assessment tasks, we are often faced with major difficulties when it comes to applying them to optimization problems because they lose most (if not all) of the useful mathematical properties possessed by MSE. There is an exception to this, i.e., measures based on divisive normalization, which are originally introduced in [11], have been formulated as Riemannian metrics in [12] and [13] and shown to have good geometric properties in [14].

Among the recent objective perceptual measures, the SSIM index has the simplest formulation and has gained widespread popularity in a broad range of practical applications [1]. There have been also several attempts to incorporate SSIM in optimization frameworks in order to improve perceived image/video quality in a number of image processing problems, which include image denoising [15], [16], image restoration [17], [18], equalizer design [19], contrast enhancement [20], watermarking [20], image approximation [21], image quantization and coding [22]–[24], and rate–distortion optimization in standard video compression [25]–[27]. Some mathematical aspects of SSIM

Manuscript received March 29, 2011; revised August 22, 2011 and October 12, 2011; accepted October 12, 2011. Date of publication October 24, 2011; date of current version March 21, 2012. This work was supported by the Natural Sciences and Engineering Research Council of Canada. The work of D. Brunet and Z. Wang was supported in part by the Province of Ontario Ministry of Research and Innovation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jesus Malo.

D. Brunet and E. R. Vrscay are with the Department of Applied Mathematics, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: dbrunet@uwaterloo.ca; ervrscay@uwaterloo.ca).

Z. Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: zhouwang@ieec.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2011.2173206

¹Some preliminary results of this paper (specifically, parts of Section II) were presented at *International Conference on Image and Analysis and Recognition*, Burnaby, BC, Canada, June 2011.

have been also investigated. For example, the gradient of SSIM with respect to the image has been derived and used in iterative optimization procedures [28], [29]. Some aspects of the convexity and quadratic properties of SSIM have been also studied [19], [30]. However, in general, deep analysis of the mathematical properties of SSIM is seriously lacking. In practice, SSIM is often used as a black box in optimization tasks as merely an adhesive control unit outside the main optimization module. Consequently, the potentials of SSIM optimization cannot be fully exploited, far behind its widespread usage as purely an assessment or comparison tool in other applications.

The purpose of this paper is to study the important mathematical properties of SSIM. We are particularly interested in seeing whether the pleasing properties of MSE also hold for SSIM (or its direct variants). During our derivation, the potential application in our minds is the general optimal image reconstruction problem, which may be formulated as follows: Assume that there exists an original image X that is unknown to us, but for which we have some prior knowledge. What we observe is some partial or distorted information Y generated by an operator \mathcal{D} applied upon X , i.e., $Y = \mathcal{D}(X)$. The goal is to design a reconstruction function \mathcal{R} that, when applied to observed data Y , creates a reconstructed image $X^{(R)} = \mathcal{R}(Y)$, which is desired to be a good approximation of X . The problem to be solved is to find the optimal function \mathcal{R} such that some error energy function (e.g., expected value of squared distance) defined by the distance

$$d(X, X^{(R)})$$

is minimized. Depending on the operator \mathcal{D} , this general framework covers a wide range of practical problems. For example, when \mathcal{D} is an image compression operator, then Y is a coded bitstream of image X , and this is an image decompression problem; when \mathcal{D} represents noise contamination, then Y is an observed noisy image, and this is an image denoising problem; when \mathcal{D} denotes a blur process, it becomes a deblurring problem; when \mathcal{D} is a downsampling operator based on a predefined sampling lattice, it corresponds to an image interpolation problem; when \mathcal{D} denotes a “blotching” operator that damages some regions in an image, this is an image inpainting problem; when \mathcal{D} is a quantization operator, this is a dequantization problem; when \mathcal{D} is a compressive sensing matrix, then this becomes a problem of image reconstruction from compressed sensing data.

What we are concerned about here is the definition of distance d , which should not only reflect perceptual image distortions (since, in most applications, human eyes are the ultimate consumers of images) but should also possess good mathematical properties (so that finding optimal solutions is a manageable task). We assert that this is a problem of fundamental importance because a different distance function could lead to substantial progress in all the aforementioned practical applications.

The division of this paper is as follows: In Section II, we define the SSIM index and show the metric space properties of its modifications. Then, in Section III, we show more basic properties of the SSIM and focus particularly on its convexity properties. We conclude this paper by summarizing all the properties

in tables and citing some research avenues in Section IV. The principal contribution of this paper resides in the proofs of two important properties of SSIM, but due to their technical nature, they are left in the Appendix.

II. SSIM, NORMALIZED, AND GENERALIZED METRICS

For the remainder of this paper, we let \mathbf{R}_+ denote the nonnegative real line, i.e., $[0, \infty)$, and \mathbf{R}_+^N denote the first orthant, i.e., the set of N -dimensional vectors with nonnegative components.

Given two images $x \in \mathbf{R}_+^N$ and $y \in \mathbf{R}_+^N$ to be compared, the SSIM index [4] combines three components, namely, a luminance (mean) distortion term, a contrast (variance) distortion term, and a correlation term as follows:

$$l(x, y) = \frac{2\bar{x}\bar{y} + c_1}{\bar{x}^2 + \bar{y}^2 + c_1} \quad (1)$$

$$c(x, y) = \frac{2s_x s_y + c_2}{s_x^2 + s_y^2 + c_2} \quad (2)$$

$$s(x, y) = \frac{s_{x,y} + c_3}{s_x s_y + c_3} \quad (3)$$

where \bar{x} , \bar{y} , s_x^2 , s_y^2 , and $s_{x,y}$ represent, respectively, the local mean of x and y , the local variance of x and y , and the local covariance between x and y . Constants c_1 , c_2 , and c_3 are small constants that aim to characterize the saturation effects of the visual system at low luminance and contrast regions and that assure numerical stability when the denominators are close to zero. The universal image quality index [3], an earlier version of the SSIM index, corresponds to the case that $c_1 = c_2 = c_3 = 0$. The first two terms account for nonstructural distortion of the image, whereas the last term accounts for structural distortion (absence of correlation) of the image.

An SSIM map is produced by computing the SSIM on a local window. A single index is then obtained by pooling all the local errors computed. Note that, for the purpose of optimizing the image quality measure, a single SSIM index is generally not relevant, unless we are dealing with a bit budget as in image compression.

If we choose $c_3 = c_2/2$, then the SSIM index can be then written as

$$\text{SSIM}(x, y) = S_1(x, y)S_2(x, y) \quad (4)$$

where

$$S_1(x, y) = l(x, y) = \frac{2\bar{x}\bar{y} + c_1}{\bar{x}^2 + \bar{y}^2 + c_1} \quad (5)$$

$$S_2(x, y) = c(x, y)s(x, y) = \frac{2s_x s_y + c_2}{s_x^2 + s_y^2 + c_2}. \quad (6)$$

For $x, y \in \mathbf{R}_+^N$, one can show that $-1 \leq \text{SSIM}(x, y) \leq 1$ and $\text{SSIM}(x, y) = 1 \iff x = y$. In other words, the closer that x and y are to each other, the closer $\text{SSIM}(x, y)$ is to 1.

Clearly, the SSIM index is not a metric. In what follows, we show that $\sqrt{1 - S_1}$ and $\sqrt{1 - S_2}$ are normalized metrics. We then combine them to form a vector-valued metric or find a way to “scalarize” them to form a metric.

A. Normalized Metrics

Observing that $S_1(x, y) = 1$ if and only if $\bar{x} = \bar{y}$ and that $S_2(x, y) = 1$ if and only if $x - \bar{x} = y - \bar{y}$ leads to the question whether

$$T_i(x, y) = 1 - S_i(x, y), \quad i = 1, 2 \quad (7)$$

is a metric. It is not a metric since the triangular inequality does not hold. However, we will show that metrics are obtained if the square root of T_i is used, i.e.,

$$\begin{aligned} d_1(\bar{x}, \bar{y}) &:= \sqrt{\frac{|\bar{x} - \bar{y}|^2}{\bar{x}^2 + \bar{y}^2 + c_1}} \\ &= \sqrt{1 - S_1(x, y)} \end{aligned} \quad (8)$$

$$\begin{aligned} d_2(x - \bar{x}, y - \bar{y}) &:= \sqrt{\frac{\|(x - \bar{x}) - (y - \bar{y})\|^2}{\|x - \bar{x}\|^2 + \|y - \bar{y}\|^2 + (N - 1)c_2}} \\ &= \sqrt{\frac{s_x^2 - 2s_{x,y} + s_y^2}{s_x^2 + s_y^2 + c_2}} \\ &= \sqrt{1 - S_2(x, y)}. \end{aligned} \quad (9)$$

The very similar forms of d_1 and d_2 can be written as a root mean-squared-error (RMSE) normalized by a modification of power mean $A_2(x, y) = ((x^2 + y^2)/2)^{1/2}$.

Definition 2.1: We define the *normalized root mean-squares-error* (NRMSE) with stability constant c to be the metric of the form

$$\text{NRMSE}(x, y) := \frac{\|x - y\|}{\sqrt{\|x\|^2 + \|y\|^2 + c}}. \quad (10)$$

This is an example of a *normalized metric*, which is a special kind of metric that measures a relative error.

Definition 2.2: Let $(X, \|\cdot\|)$ be a normed space. Given $x, y \in X$, a *normalized metric* or *relative distance* is a metric of the form

$$d_n(x, y) = \frac{\|x - y\|}{M(\|x\|, \|y\|)} \quad (11)$$

where $M : \mathbf{R}_+ \times \mathbf{R}_+ \rightarrow \mathbf{R}_+$ is a function chosen so that d_n is a metric. By convention, we set $0/0 := 0$.

In this paper, we are interested in case $X = \mathbf{R}^N$ with the L^2 -norm. A few authors have studied normalized metrics in this space. Klamkin and Meir [31] showed that, for

$$M(\|x\|, \|y\|) = (\|x\|^p + \|y\|^p)^{q/p} \quad (12)$$

d_n is a metric for $q = 1$ and for all $p \geq 1$. Hästö [32] generalized the proof for $0 < q \leq 1$ and for $p \geq \max(1 - q, (2 - q/3))$. He also showed isometries, quasi-convexity, and region of convexity properties for these metrics [33]. Finally, in [34], Yianilos showed that

$$d_n(x, y) := \frac{\|x - y\|}{\|x\| + \|y\| + c} \quad (13)$$

is a metric for $c \geq 0$. In addition, worthy of mention is the work of Li *et al.* [35] on normalized metrics for binary strings using

Kolmogorov complexity. However, none of these authors have covered the case of interest for our application.

It remains to be shown that the NRMSE, hence d_1 and d_2 , are metrics. We need two lemmas.

The first lemma is Ptolemy's inequality, an important geometric identity for points in Euclidean space \mathbf{R}^N . This is a generalization of the well-known geometric identity due to Ptolemy that states that, if a quadrilateral is inscribed in a circle, then the sum of the products of the lengths of its opposed edges is equal to the product of the length of its diagonals.

Lemma 2.3 (Ptolemy's Inequality for \mathbf{R}^n): Let A, B, C , and D be the four points of \mathbf{R}^N . Then

$$\overline{AB} \cdot \overline{CD} + \overline{BC} \cdot \overline{AD} \geq \overline{BD} \cdot \overline{AC}. \quad (14)$$

Various proofs of this lemma can be found in the literature, e.g., [36], but it seems that not all authors are aware of the following simple proof outlined in [37].

Proof: Let A', B' , and C' denote the inversion of A, B , and C , respectively, around a sphere centered at D and of radius r . By the triangular inequality, we have

$$\overline{A'B'} + \overline{B'C'} \geq \overline{A'C'}. \quad (15)$$

One can verify the following circle inversion property:

$$\overline{A'B'} = r^2 \frac{\overline{AB}}{\overline{AD} \cdot \overline{BD}}. \quad (16)$$

Substituting (16) into (15), we obtain

$$r^2 \frac{\overline{AB}}{\overline{AD} \cdot \overline{BD}} + r^2 \frac{\overline{BC}}{\overline{BD} \cdot \overline{CD}} \geq r^2 \frac{\overline{AC}}{\overline{AD} \cdot \overline{CD}}. \quad (17)$$

A rearrangement yields Ptolemy's inequality (14). ■

In general, a metric space (X, d) is called Ptolemaic if

$$d(x, w)d(y, z) + d(z, w)d(x, y) \geq d(y, w)d(x, z) \quad (18)$$

for all $x, y, z, w \in X$.

The second lemma asserts that, under certain conditions, it suffices to study normalized metrics on \mathbf{R}_+ . Before stating the result, we need to define moderately increasing functions.

Definition 2.4: A function $f : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ is moderately increasing (nondecreasing) if

- 1 $f(x)$ is increasing (nondecreasing);
- 2 $f(x)/x$ is decreasing (nonincreasing) for $x > 0$.

For example, one can easily verify that

$$f(x) = \sqrt{x^2 + y^2 + c} \quad (19)$$

is moderately increasing.

Lemma 2.5: Let $(X, \|\cdot\|)$ be a Ptolemaic normed space, and let $M(\|x\|, \|y\|)$ be a moderately nondecreasing function with respect to $\|x\|$, then

$$d_n(x, y) = \frac{\|x - y\|}{M(\|x\|, \|y\|)} \quad (20)$$

is a metric on X if and only if it is a metric on \mathbf{R}_+ .

Proof: See [32]. ■

We are now ready to state the main result of this section.

Theorem 2.6: The NRMSE

$$d(x, y) := \text{NRMSE}(x, y, c) = \frac{\|x - y\|_2}{\sqrt{\|x\|_2^2 + \|y\|_2^2 + c}} \quad (21)$$

is a metric on \mathbf{R}^N for all $c \geq 0$.

Proof: See the Appendix. ■

B. Metric Spaces Arising From SSIM: From a “Scalarized” Metric to a “Cone Metric Space”

From the previous section, each of the components $S_1(x, y)$ and $S_2(x, y)$ of the SSIM function $S(x, y)$ in (4) gives rise to a distance function of the form $d_i(x, y) = \sqrt{1 - S_i(x, y)}$, $i = 1, 2$. The natural question is what to do with these distance functions. The first step is to consider them as components of a vector $\mathbf{d}(x, y)$, as done in [38]. First of all, it is useful to recall the following classical theorem.

Lemma 2.7 (Orthogonal Decomposition Theorem): Let $\{X_m\}_{m=1}^M$ be an orthogonal decomposition of \mathbf{R}^N , i.e.,

$$\bigcup_{m=1}^M X_m = \mathbf{R}^N \\ X_i \perp X_j \quad \text{for all } i \neq j.$$

Then, each element x of \mathbf{R}^N can be uniquely written as

$$x = \sum_{m=1}^M x_m \quad (22)$$

where $x_m = P_m(x)$ are the projections of x on X_m .

From (8) and (9), we are now motivated to decompose \mathbf{R}_+^N into the following direct sum:

$$\mathbf{R}_+^N = \mathbf{R}_1 \oplus \mathbf{R}_2 \quad (23)$$

where $\mathbf{R}_1 = \text{span}(\mathbf{1}) = \text{span}(1, 1, \dots, 1)$ and $\mathbf{R}_2 = \mathbf{R}_+^\perp$. A given signal $x \in \mathbf{R}_+^N$ will be decomposed into a 1-D mean component $x = \bar{x}\mathbf{1} \in \mathbf{R}_1$ and an $(N - 1)$ -dimensional zero-mean component $x - \bar{x}\mathbf{1} \in \mathbf{R}_2$. Our resulting vector of metrics becomes

$$\mathbf{d}(x, y) = (d_1(\bar{x}, \bar{y}), d_2(x - \bar{x}, y - \bar{y})) \quad (24)$$

where d_1 and d_2 are defined in (8) and (9), respectively.

We now form a metric from a combination of the two components d_1 and d_2 of the vector of metrics. The goal is to find a suitable approximation of SSIM that will belong to a metric space. For this, we will need the following property.

Definition 2.8: A norm in \mathbf{R}^M is said to satisfy the *increasing property* if for any $a \in \mathbf{R}_+^M$ and $b \in \mathbf{R}_+^M$

$$\|a + b\| \geq \|a\|. \quad (25)$$

The increasing property in (25) holds for suitably weighted l^p norms, e.g.,

$$\|x\|_{w,p} = \left[\sum_{m=1}^M w_m |x_m|^p \right]^{1/p} \quad (26)$$

where $w_{k,p} > 0$ and $1 \leq p < \infty$ for $1 \leq m \leq M$. It also holds for the L^∞ -norm, e.g.,

$$\|x\|_\infty = \max_{1 \leq m \leq M} \{ |x_m| \}. \quad (27)$$

However, one can show that (25) does not hold for **all** norms.

The following theorem shows that, in general, if $\|\cdot\|$ possesses the increasing property, then the norm of the vector of metrics such as (24) is also a metric.

Theorem 2.9: Let $\{X_m\}_{m=1}^M$ be an orthogonal decomposition of \mathbf{R}^N and $P_m : \mathbf{R}^N \rightarrow X_m$ for $1 \leq m \leq M$ be the projection of $x \in \mathbf{R}^N$ to X_m with $x_m = P_m(x)$. Let $\{d_m\}_{m=1}^M$ be M metrics on, respectively, $\{X_m\}_{m=1}^M$. Define

$$\mathbf{d}(x, y) : \mathbf{R}^N \times \mathbf{R}^N \rightarrow \mathbf{R}_+^M \\ (x, y) \mapsto (d_1(x_1, y_1), \dots, d_M(x_M, y_M)). \quad (28)$$

Let $\|\cdot\|$ be a norm in \mathbf{R}^M that satisfies the increasing property (25) in \mathbf{R}_+^M . Then, for \mathbf{d} defined in (28)

$$d(x, y) := \|\mathbf{d}\| \quad (29)$$

is a metric in \mathbf{R}^N .

Proof: It is clear that $x = y$ implies that $\mathbf{d} = \mathbf{0}$. Now, if $\mathbf{d}(x, y) = \mathbf{0}$, then $d_m(x_m, y_m) = 0$ for all $1 \leq m \leq M$; thus, $x_m = y_m$ for all $1 \leq m \leq M$ since d_m is a metric. By the Orthogonal Decomposition Theorem (see Lemma 2.7), this implies that $x = y$. Therefore, by the property of norms, we have $\|\mathbf{d}\| = \mathbf{0} \iff x = y$. The symmetry property is straightforward. It remains to prove that $\|\mathbf{d}\|$ satisfies the triangle inequality. Since d_1, \dots, d_M are metrics, every component of \mathbf{d} satisfies the triangular inequality. From the assumption that the $\|\cdot\|$ norm satisfies the increasing property, we conclude that

$$\|\mathbf{d}(x, z)\| \leq \|\mathbf{d}(x, y) + \mathbf{d}(y, z)\| \\ \leq \|\mathbf{d}(x, y)\| + \|\mathbf{d}(y, z)\|. \quad (30)$$

The above procedure represents a “scalarization,” producing a single real number, e.g., the distance $\|\mathbf{d}(x, y)\|_2$ in (33), from the two-component metric. This is analogous to scalarization of (4) in which the two components S_1 and S_2 are multiplied to produce a single real number. Indeed, one of the attractive features of using real numbers for quantities such as distances is that they are *totally ordered*, i.e., given two real numbers $a, b \in \mathbf{R}$, one of the following properties must hold: 1) $a < b$; 2) $a = b$; or 3) $a > b$.

However, it may be desirable, and this is still an open question in image processing, to keep the metric \mathbf{d} in vector form, i.e., as an element in \mathbf{R}^2 , or \mathbf{R}^M , $M > 1$, in general. In this case, we lose the ordering property of the real numbers, i.e., the spaces \mathbf{R}^M for $M > 1$ are *partially ordered sets*. The most convenient example of a partial ordering on \mathbf{R}^M is the following. Given $x, y \in \mathbf{R}^M$, $M > 1$, we write that “ $x \succeq y$ ” if the following conditions hold:

$$x_k \geq y_k, \quad 1 \leq k \leq M. \quad (31)$$

If one or more of the above conditions do not hold, then partial ordering $\mathbf{x} \succeq \mathbf{y}$ is not applicable and we say that \mathbf{x} and \mathbf{y} cannot be ordered.

It is not difficult to show that the above partial ordering defines a *vector-valued metric* on a space \mathbf{X}

$$\mathbf{d}(x, y) : X \times X \rightarrow \mathbf{R}_+^M \quad (32)$$

satisfying the following conditions for all $x, y, z \in X$:

- 1) $\mathbf{d}(x, y) = \mathbf{0} \iff x = y$.
- 2) $\mathbf{d}(x, y) = \mathbf{d}(y, x)$.
- 3) $\mathbf{d}(x, y) + \mathbf{d}(y, z) \succeq \mathbf{d}(x, z)$.

The above is a particular example of a partial ordering in \mathbf{R}_+^M . It can be replaced by a more general condition involving *cones*.

Definition 2.10: Let $X = \mathbf{R}^M$. A cone $P \subset X$ obeys the following set of conditions.

- 1) P is closed, nonempty, and $P \neq \{0\}$.
- 2) If $x, y \in P$, then $ax + by \in P$ for all $a, b \in \mathbf{R}_+$ (i.e., $a, b \geq 0$),
- 3) $x \in P$ and $-x \in P$ imply that $x = 0$.

A partial ordering may be then defined with respect to a given cone $P \subset X$ as follows: For $x, y \in X$, we write $x \succeq y$ if and only if $x - y \in P$. This is also symbolically written as $y \in x + P$.

Example: Positive orthant $\mathbf{R}_+^M \subset \mathbf{R}^M$ is an example of a cone. In addition, (31) is the partial ordering associated with this cone.

In general, a cone $P \subset \mathbf{R}_+^M$ may be then used to define a *cone metric* that will satisfy the three properties of a vector-valued metric above, particularly the triangle inequality with respect to the partial ordering defined by P . Here, we mention that the mathematical definition of a cone metric is actually more general in form: A cone metric d is a mapping from $X \times X$ to a Banach space E [39].

C. Perceptual Validation of the Modification of SSIM

For practical applications, the metric obtained from the modification of SSIM is essentially equivalent to the original SSIM index. This assertion is shown in two ways, i.e., theoretically and with a perceptual experiment.

We can apply Theorem 2.9 to the particular case of the vector of metrics given in (24). Taking a weighted l^p norm, as shown in (26), with $w = (1, 1)$ and $p = 2$ leads to

$$\begin{aligned} D_2 &:= \|\mathbf{d}(x, y)\|_2 = \sqrt{d_1^2(\bar{x}, \bar{y}) + d_2^2(x - \bar{x}, y - \bar{y})} \\ &= \sqrt{2 - S_1(x, y) - S_2(x, y)}. \end{aligned} \quad (33)$$

Metric $\|\mathbf{d}(x, y)\|_2$ can be seen as a low-order estimation of $\sqrt{1 - \text{SSIM}}$. Indeed, we have

$$\begin{aligned} \sqrt{1 - \text{SSIM}} &= \sqrt{1 - S_1 S_2} \\ &= \sqrt{1 - (1 - d_1^2)(1 - d_2^2)} \\ &= \sqrt{d_1^2 + d_2^2 - d_1^2 d_2^2}. \end{aligned} \quad (34)$$

If either $\bar{x} = \bar{y}$ or $x - \bar{x} = y - \bar{y}$, then $S_1(x, y) = 1$ or $S_2(x, y) = 1$ and we have

$$\|\mathbf{d}(x, y)\|_2 = \sqrt{1 - \text{SSIM}(x, y)}. \quad (35)$$

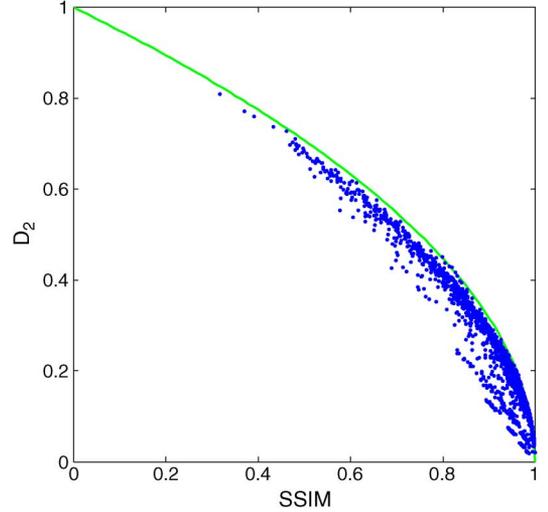


Fig. 1. Comparison of the SSIM index and metric $D_2 = \sqrt{2 - S_1 - S_2}$ for the distorted images of the TID2008 database. It is observed that $\sqrt{1 - \text{SSIM}}$ follows the data very closely.

In fact, most typical image distortions $Y = \mathcal{D}(X)$, e.g., additive noise, blur, and image compression, roughly preserve the local mean of images so that $\bar{x} \simeq \bar{y}$. This explains the observation by Rouse and Hemami [40] on the Laboratory for Image and Video Engineering (LIVE) database [41] that the second (2) and third (3) components of SSIM are the most important to predict the human perception of image distortion and justify the use of $\|\mathbf{d}(x, y)\|_2$ as an approximation of SSIM. However, in the case that there exist significant changes in the mean image intensity, as shown in the TID2008 database [42], [43], such an approximation is no longer valid. Indeed, the full SSIM measure with all three components included performs the best for the full data set in the TID2008 database [9].

To illustrate this fact, in Fig. 1, the SSIM index is compared with the score given by $D_2 := \|\mathbf{d}(x, y)\|_2$ on the 1700 distorted images of the TID2008 database. As shown in [4], the quality maps were computed on 11×11 local patches with circular Gaussian weighting (standard deviation of 1.5 pixels) of a down-sampled image and then pooled into a single index by averaging. The resulting Pearson's correlation coefficient between D_2 and $\sqrt{1 - \text{SSIM}}$ is 0.967, showing a very strong relationship. If only the images distorted with the most commonly encountered distortions are considered (see [42, Table IV], for details), then a nearly perfect correlation of 0.994 is obtained. As expected, the same experiment with the LIVE database also gives a very high correlation of 0.997.

The mean opinion scores (MOSs) of the image quality of both TID2008 and LIVE databases, release 2, are compared with D_2 and SSIM in Table I. Two kinds of rank correlation were computed, namely, Spearman's and Kendall's correlations. It is observed that the correlation coefficients match very well for LIVE and TID2008 databases restricted to realistic distortions (including only noise distortions, denoising, blurring, and compression). It is when the full database is used that the comparison between D_2 and SSIM breaks down, metric D_2 performing slightly worse than SSIM. This problem may be rectified by adjusting the weights assigned to d_1 and d_2 in (24).

TABLE I
CORRELATION BETWEEN MOS AND IMAGE QUALITY ASSESSMENT MEASURES

Correlation Coefficient	Measure	LIVE	TID2008 Full Set	TID2008 Realistic Distortions
Kendall's tau	SSIM	0.815	0.577	0.663
	D_2	0.811	0.501	0.642
Spearman's rho	SSIM	0.950	0.775	0.857
	D_2	0.948	0.667	0.839

III. CONVEXITY PROPERTIES

Here, we are concerned with the mathematical properties of the normalized MSE (NMSE), because of its connection to the SSIM function, cf. (8) and (9).

Definition 3.1: The NMSE is defined as the square of the NRMSE

$$\text{NMSE}(x, y, c) := \frac{\|x - y\|^2}{\|x\|^2 + \|y\|^2 + c} \quad (36)$$

where $c \geq 0$.

As it is well known, most image processing analysis and algorithms are based on L^2 -based methods, namely, RMSE and MSE. The $\text{RMSE}(x, y)$ function is a metric (in the strict mathematical sense). Nevertheless, many optimization routines are performed with respect to the $\text{MSE}(x, y)$ function, taking advantage of its convexity. This also applies for the NRMSE and the NMSE.

One problem of particular interest is to find the region of convexity for the NMSE. For the benefit of the reader, we outline here the definitions of convexity of a set and of a function.

Definition 3.2: A subset X of a vector space is convex if for all $x, y \in X$ and for all $\lambda \in (0, 1)$

$$\lambda x + (1 - \lambda)y \in X. \quad (37)$$

Definition 3.3: A function $f : X \rightarrow \mathbf{R}$ is convex if for all $x, y \in X$ and for all $\lambda \in (0, 1)$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (38)$$

A. Preliminary Properties

We start with some basic invariance properties of the NMSE, namely, invariance under: 1) multiplication by a scalar; and 2) isometries (e.g., rotations and reflections). The properties, along with others that follow, will also hold for each of the two components, i.e., S_1 and S_2 , of SSIM since

$$S_1(\bar{x}, \bar{y}) = 1 - \text{NMSE}(\bar{x}, \bar{y}, c_1) \quad (39)$$

$$S_2(x - \bar{x}, y - \bar{y}) = 1 - \text{NMSE}(x - \bar{x}, y - \bar{y}, c_2). \quad (40)$$

The first property is analog to Parseval's identity for the L^2 -norm.

Proposition 3.4: $\text{NMSE}(x, y, c)$ is invariant under isometries.

Proof: If U is an orthogonal or unitary matrix, then

$$\begin{aligned} \text{NMSE}(Ux, Uy, c) &= \frac{\|Ux - Uy\|^2}{\|Ux\|^2 + \|Uy\|^2 + c} \\ &= \frac{\|x - y\|^2}{\|x\|^2 + \|y\|^2 + c} \\ &= \text{NMSE}(x, y, c). \end{aligned} \quad (41)$$

Proposition 3.5: $\text{NMSE}(x, y, 0)$ is invariant under scaling. ■

Proof: If $\alpha > 0$, then

$$\begin{aligned} \text{NMSE}(\alpha x, \alpha y, 0) &= \frac{\|\alpha x - \alpha y\|^2}{\|\alpha x\|^2 + \|\alpha y\|^2} \\ &= \frac{\alpha^2 \|x - y\|^2}{\alpha^2 (\|x\|^2 + \|y\|^2)} \\ &= \text{NMSE}(x, y, 0). \end{aligned} \quad (42)$$

This explains the preserved SSIM measure in the discrete cosine transform domain [23]. ■

The scale invariance property is valid only when $c = 0$. For $c > 0$

$$\text{NMSE}(\alpha x, \alpha y, c) = \text{NMSE}(x, y, c/\alpha^2) \quad (43)$$

which is very close to scaling invariance if c and c/α^2 are negligible in comparison to $\|x\|_2^2 + \|y\|_2^2$.

The next property will be very useful for analyzing $\text{NMSE}(x, y, c)$ when y is fixed, reducing the dimensions of interest to only two.

Proposition 3.6: For $x, y \in \mathbf{R}^N$ ($N > 1$) with y fixed, $g(x) := \text{NMSE}(x, y, c)$ exhibits a rotational symmetry around the direction of y .

Proof: Consider any $x_{(1)} = ay + bu$ and $x_{(2)} = ay + bv$ where $u, v \in \mathbf{R}^N$ with $\langle y, u \rangle = \langle y, v \rangle = 0$ and $\|u\| = \|v\| = 1$. In this case

$$g(x_{(1)}) = 1 - \frac{(a-1)^2 \|y\|^2 + b^2 + c}{(a^2 + 1)\|y\|^2 + b^2 + c} = g(x_{(2)}) \quad (44)$$

for all $a \in \mathbf{R}$, $b \in \mathbf{R}$, and for all $y \in \mathbf{R}^N$. ■

The last preliminary property of $\text{NMSE}(x, y, c)$ will be used in the region of convexity proof.

Proposition 3.7: For $x, y \in \mathbf{R}^N$ with $N > 1$, $g(x) := \text{NMSE}(x, y, c)$ is monotonically increasing perpendicularly to the direction of y .

Proof: Consider any $x_{(1)} = ay + bu$ and $x_{(2)} = ay + b'u$ where $u \in \mathbf{R}^N$ with $\langle y, u \rangle = 0$ and $\|u\| = 1$. In this case

$$\begin{aligned} \text{NMSE}(x_{(1)}, y, c) &= 1 - \frac{(a-1)^2 \|y\|^2 + b^2 + c}{(a^2 + 1)\|y\|^2 + b^2 + c} \\ &\leq 1 - \frac{(a-1)^2 \|y\|^2 + b'^2 + c}{(a^2 + 1)\|y\|^2 + b'^2 + c} \\ &= \text{NMSE}(x_{(2)}, y, c) \end{aligned} \quad (45)$$

for all $a \in \mathbf{R}$, $b \leq b' \in \mathbf{R}$, and for all $y \in \mathbf{R}^N$. ■

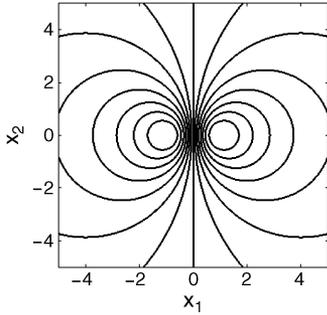


Fig. 2. Level sets of $\text{NMSE}(x, y, c)$ with $c = 0$ for $x = (x_1, x_2)$ and $y = (1, 0)$.

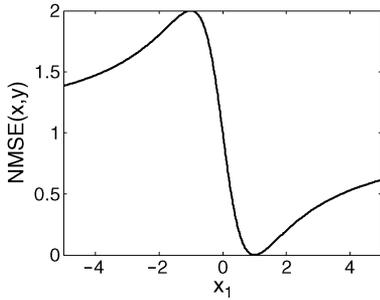


Fig. 3. Cross section of $\text{NMSE}(x, y, c)$ with $c = 0$ for $x = (x_1, x_2)$ and $y = (1, 0)$ along the vertical plane $x_2 = 0$.

We now draw the level sets and one cross section of the graph of the NMSE for a fixed image $y \in \mathbf{R}^N$. By rotational symmetry of x around y (see Prop. 3.6), we only need to consider two directions in order to have full knowledge of the shape of $\text{NMSE}(x, y, c)$, namely, y and any direction orthogonal to y . Notice also that, by invariance under isometries (see Prop. 3.4), we can consider, without loss of generality, that $y = (\|y\|, 0)$. Finally, notice that for x and y sufficiently large, constant c is negligible. Therefore, by scale invariance (see Prop. 3.5), we can, without loss of generality, consider $y = (1, 0)$. The level sets of $\text{NMSE}(x, y, c)$ are shown in Fig. 2. Its cross section along the second component of x , i.e., in the direction of $y = (1, 0)$, is shown in Fig. 3.

B. Quasi-Convexity

As it is clear by looking at Fig. 3, $g(x) = \text{NMSE}(x, y, c)$ is not convex. However, a weaker form of convexity called quasi-convexity can be shown to hold. Quasi-convexity is a useful property for nonlinear optimization: it assures the existence of a global minimum on any convex subset of the function domain (see, e.g., [44]).

Definition 3.8: Given a convex set X , a function $f : X \rightarrow \mathbf{R}$ is said to be quasi-convex if its h -sublevel set, defined as

$$X_h = \{x \in X | f(x) \leq h\} \quad (46)$$

is a convex set for all $h \in \text{Range}(f)$.

Theorem 3.9: For y fixed, $\text{NMSE}(x, y, c)$ is quasi-convex on $H = \{x \in \mathbf{R}^N : \langle x, y \rangle \geq (-c/2)\}$.

Proof: We easily see that

$$x \in H \iff \text{NMSE}(x, y, c) \leq 1. \quad (47)$$

Therefore, it suffices to show that X_h is convex for all $h \leq 1$. To show this, we first examine the inequality

$$\text{NMSE}(x, y, c) = \frac{\|x - y\|_2^2}{\|x\|^2 + \|y\|^2 + c} \leq h. \quad (48)$$

For $h < 1$, this inequality becomes

$$\|x\|^2 - \frac{2}{1 - h^2} \langle y, x \rangle \leq -\|y\|^2 + c \frac{h^2}{1 - h^2}. \quad (49)$$

Completing the square yields

$$\left\| x - \frac{y}{1 - h^2} \right\|^2 \leq \|y\|_2^2 \left(\frac{1}{(1 - h^2)^2} - 1 \right) + c \frac{h^2}{1 - h^2}. \quad (50)$$

The set of points x for which this inequality holds is a hypersphere, which is a convex set.

In the particular case when $h = 1$, we have the inequality

$$-2 \langle y, x \rangle \leq c \quad (51)$$

which is satisfied for all $x \in H$, as previously defined. Set H is bounded by a hyperplane and, therefore, convex. ■

The quasi-convexity region H_1 of $S_1(\bar{x}, \bar{y})$ is thus

$$H_1 = \left\{ \bar{x} : \bar{x} \bar{y} \geq -\frac{c_1}{2} \right\} \quad (52)$$

and the quasi-convexity region H_2 of $S_2(x - \bar{x}, y - \bar{y})$ is

$$H_2 = \left\{ x - \bar{x} : \langle x - \bar{x}, y - \bar{y} \rangle \geq -\frac{c_2}{2} \right\}. \quad (53)$$

C. Region of Convexity

For y fixed, we wish to find the set of all points where $g(x) = \text{NMSE}(x, y, c)$ is convex. In particular, we shall show that $\text{NMSE}(x, y, c)$ is locally convex, i.e., there exists a sphere around the minimum $x = y$ for which $\text{NMSE}(x, y, c)$ is convex. For this, the following technical lemma is needed.

Lemma 3.10: Let X be a 2-D convex subset of \mathbf{R}^N , and let $y \in X$. If f is convex on X , has a rotational symmetry around the direction of y , and is monotonically increasing in a direction perpendicular to y , then f is convex on the rotation of X around y .

Proof: See the Appendix. ■

Theorem 3.11: For $x, y \in \mathbf{R}$, $g(x) = \text{NMSE}(x, y, 0)$ is convex on

$$0 \leq x \leq \sqrt{3}|y|. \quad (54)$$

For $x, y \in \mathbf{R}^N$ with $N > 1$, $g(x) = \text{NMSE}(x, y, 0)$ is convex on the region bounded by the rotation of

$$v(t) = \sqrt{\frac{-2t^2 - \|y\|^2 + \sqrt{16t^2\|y\|^2 + \|y\|^4}}{2}} \quad (55)$$

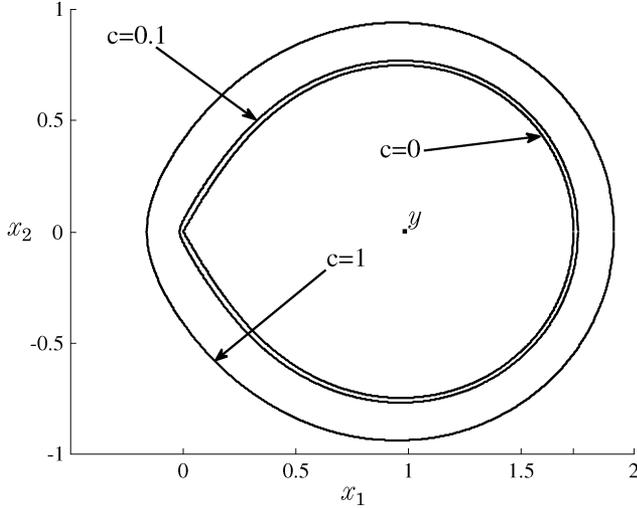


Fig. 4. Boundary of the region where $g(x) = \text{NMSE}(x, y, c)$ is convex with $y = (1, 0)$ and $x = (x_1, x_2)$ for $c = 0$, $c = 0.1$, and $c = 1$.

$0 \leq t \leq \sqrt{3}\|y\|$, around the direction of y , i.e., g is convex on the region bounded by $ty/\|y\| + v(t)u/\|u\|$ for any u perpendicular to y (see Fig. 4).

Proof: The 1-D case can be easily verified by finding the region where the second derivative of $g(x)$ is positive. For the general case, we proceed in three steps.

1) First, consider the 2-D case and denote $x = (x_1, x_2)$. For $y = (1, 0)$, we have

$$g(x) = \text{NMSE}(x, y, 0) = \frac{(x_1 - 1)^2 + x_2^2}{x_1^2 + x_2^2 + 1}. \quad (56)$$

To determine the region of convexity of g , it suffices to compute the eigenvalues of its Hessian matrix and to determine when they are nonnegative. One can verify that this happens when

$$|x_2| \leq \sqrt{\frac{-2x_1^2 - 1 + \sqrt{16x_1^2 + 1}}{2}} := v(x_1) \quad (57)$$

$$0 \leq x_1 \leq \sqrt{3}. \quad (58)$$

We shall call this region X . It is convex since $v''(x_1) \leq 0$.

2) For any $y \in \mathbf{R}^2$, we can write $y = (\|y\| \cos(\theta), \|y\| \sin(\theta))$, where θ is the angle between y and $(1, 0)$. Vector y can be thought as a transformation of $e_1 = (1, 0)$

$$y = \|y\| R_\theta e_1. \quad (59)$$

By the invariance properties of NMSE (see Prop. 3.4 and 3.5), we have

$$\text{NMSE}(x, y, 0) = \text{NMSE}(R_{-\theta}x/\|y\|, e_1, 0). \quad (60)$$

Therefore, $\text{NMSE}(x, y, 0)$ is convex on $\|y\| R_\theta X$, i.e., on a rotation of X by angle θ and a scaling by factor $\|y\|$.

3) Since $g(x) = \text{NMSE}(x, y, 0)$ is convex on convex region X , by rotational symmetry and monotonicity of g (see

Prop. 3.6 and 3.7), it will be also convex in the region consisting of the rotation of X around the direction of y in virtue of Lemma 3.10. ■

- **Note 1:** Here, we considered case $c = 0$. The general case is similar, but since we cannot use the scaling property, we need to consider $y = (\|y\|, 0)$ in Step 1), which leads to a complex equation for the region of convexity. The region with $y = (1, 0)$ and $c \neq 0$ will look very similar than the one for $c = 0$, but the region will be slightly larger: It adds approximately a sausage of diameter increasing with c (see Fig. 4). For practical purpose, c is usually very small ($c \leq 0.0001$) and the region of convexity of $\text{NMSE}(x, y, c)$ is almost identical to the region of convexity of $\text{NMSE}(x, y, 0)$.
- **Note 2:** There is another region of convexity for some negative x , but it is not of interest since it is not around y , i.e., the minimum of $\text{NMSE}(x, y, c)$.
- **Note 3:** An easy estimate of the region of convexity is the (hyper)sphere centered at y and of radius $(\sqrt{3} - 1)\|y\|$. Indeed, one can verify that this sphere is inscribed in the region of convexity.

From Theorem 3.11, we can directly deduce the convexity (in fact, concavity) region for $S_1(\bar{x}, \bar{y})$ and $S_2(x - \bar{x}, y - \bar{y})$.

D. Convexity of Generalized (Cone) Metrics

We show that generalized (cone) metrics carry a generalized kind of convexity. In particular, it will apply to the metric \mathbf{d} in (28).

Theorem 3.12: If d_m is convex on R_m for $1 \leq m \leq M$, then cone metric $\mathbf{d} = (d_1, \dots, d_m)$ is convex on $R_1 \times \dots \times R_M$ in the sense that, for all $x, y, z \in \mathbf{R}^N$ and for all $0 \leq \lambda \leq 1$

$$\mathbf{d}(\lambda x + (1 - \lambda)y, z) \preceq \lambda \mathbf{d}(x, z) + (1 - \lambda) \mathbf{d}(y, z) \quad (61)$$

where \preceq is the product order (31).

Proof: Since projection operator $P_m(x) = x_m$ is linear and d_m is convex on R_m for $1 \leq m \leq M$, we have

$$\begin{aligned} d_m(P_m(\lambda x + (1 - \lambda)y), P_m(z)) \\ &= d_m(\lambda x_m + (1 - \lambda)y_m, z_m) \\ &\leq \lambda d_m(x_m, z_m) + (1 - \lambda) d_m(y_m, z_m). \end{aligned} \quad (62)$$

Since the inequality holds for all $1 \leq m \leq M$, we conclude that \mathbf{d} is convex on $R_1 \times \dots \times R_M$ in the sense of (61). ■

We can also show the convexity of $\|\mathbf{d}\|$.

Theorem 3.13: If d_m is convex on R_m for $1 \leq m \leq M$ and if $\|\cdot\|$ is a norm with the increasing property [see (25)], then $\|\mathbf{d}\|$ is convex on $R_1 \times \dots \times R_M$.

Proof: From Theorem 3.12, \mathbf{d} is convex on $R_1 \times \dots \times R_M$. Since $\|\cdot\|$ is a norm, it is convex as well. Therefore, by the increasing property, we have that on $R_1 \times \dots \times R_M$

$$\begin{aligned} \|\mathbf{d}(\lambda x + (1 - \lambda)y, z)\| &\leq \|\lambda \mathbf{d}(x, z) + (1 - \lambda) \mathbf{d}(y, z)\| \\ &\leq \lambda \|\mathbf{d}(x, z)\| + (1 - \lambda) \|\mathbf{d}(y, z)\|. \end{aligned} \quad (63)$$

■

TABLE II
REGION OF VALIDITY FOR THE PROPERTIES OF THE COMPONENTS OF SSIM

Properties	S_1	S_2	$\ (S_1, S_2)\ $	(S_1, S_2)
Distance Preserving	\mathbf{R}_+	\mathbf{R}^{N-1}	\mathbf{R}^N	$\mathbf{R} \times \mathbf{R}^{N-1}$
Quasiconcavity	\mathbf{R}_+	H_2	-	$\mathbf{R}_+ \times H_2$
Concavity	R_1	R_2	$R_1 \times R_2$	$R_1 \times R_2$

where $H_2 = \{x - \bar{x} \in \mathbf{R}^{N-1} : \langle x - \bar{x}, y - \bar{y} \rangle \geq (-c_2/2)\}$, $R_1 = [0, \sqrt{3}\bar{y}]$, R_2 : A teardrop shape pointed at 0 containing the ball centered at $y - \bar{y}$ and of radius $(\sqrt{3} - 1)\|y - \bar{y}\|$ (see Theorem 3.11).

For our case of interest, we deduce that the region of convexity of $\mathbf{S} = (S_1(x, y), S_2(x, y))$ and of $\|\mathbf{S}\|_2$ is the (hyper)cylinder with $0 \leq \bar{x} \leq \sqrt{3}\bar{y}$ and with $x - \bar{x}$ following a teardrop shape around $y - \bar{y}$ like it is described in Theorem 3.11.

Contrary to convex functions, the sum of quasi-convex functions is not necessarily quasi-convex. Nevertheless, we can find a similar theorem about the quasi-convexity of \mathbf{d} , but we cannot establish the same result for $\|\mathbf{d}\|$.

Theorem 3.14: If d_m is quasi-convex on H_m for $1 \leq m \leq M$, then \mathbf{d} is quasi-convex on $H = H_1 \times \dots \times H_M$.

Proof: Since d_m is quasi-convex on H_m for $1 \leq m \leq M$, we have that the Cartesian product

$$\prod_{m=1}^M \{P_m(x) \in H_m | d_m(P_m(x), P_m(y)) \leq h_m\} \quad (64)$$

is a convex set for all $h = (h_1, h_2, \dots, h_M)$. By the Orthogonal Decomposition Theorem 2.7, (64) is equivalent to

$$\left\{ x = \sum_{m=1}^M P_m(x) \in H | \mathbf{d}(x, y) \leq h \right\} \quad (65)$$

which we know to be convex; hence, that \mathbf{d} is quasi-convex. ■

Hence, $\mathbf{S} = (S_1(x, y), S_2(x, y))$ is quasi-convex in region $H_1 \times H_2$, where H_1 and H_2 are the regions of quasi-convexity described in (52) and (53), respectively.

IV. CONCLUSION

Several important mathematical properties of the SSIM index have been investigated in this paper. First, it was shown that SSIM can be partitioned into two components S_1 and S_2 , each of which may be transformed into a valid distance metric via $d_1 = \sqrt{1 - S_1}$ and $d_2 = \sqrt{1 - S_2}$, respectively. The two metrics can be either combined to a single scalar-valued distance metric applying a norm with the increasing property (see Theorem 2.9) or be used to form a vector-valued generalized (cone) metric. We then proved that, like the MSE, the SSIM index and SSIM-based metrics are preserved under orthogonal or unitary transformations. Convexity, quasi-convexity, and generalized convexity have been also shown to locally hold for the metrics derived from SSIM. It is summarized in Table II.

The properties proved here may find broad applications in many optimization problems in image processing where objective functions correlated with perceptual image quality are desirable, as described in Section 1. For example, the distance preservation property could be useful in transform-domain algorithm design and error analysis. The convexity and quasi-convexity properties are important in the design of efficient optimization

algorithms. The property of being a valid distance metric may help analyze and prove the convergence properties of many iterative algorithms. In particular, it could be useful for fractal imaging (see [45]). In the general cone metric space setting, fixed-point theorems were also found [39]. The convexity of the generalized metric opens the door for multicriteria optimization and goal programming (see [44] and [46]). Furthermore, the results of this paper may also expedite the explorations of other new research directions in image and signal processing, such as sparse representation and collective sensing [47].

APPENDIX

Proof of Theorem 2.6: From Lemma 2.3, we know that $(\mathbf{R}^N, \|\cdot\|_2)$ is Ptolemaic. As we have already noticed, $M(\|x\|_2, \|y\|_2) = \sqrt{\|x\|_2^2 + \|y\|_2^2 + c}$ is moderately increasing with respect to $\|x\|_2$. Therefore, by Lemma 2.5, $d(x, y) := \text{NMRSE}(x, y, c)$ will be a metric on \mathbf{R}^N if it is a metric on \mathbf{R}_+ . It is easy to verify that

$$d(x, y) = 0 \iff x = y \quad (66)$$

$$d(x, y) = d(y, x) \quad \forall x, y \in \mathbf{R}. \quad (67)$$

It remains to prove the triangular inequality. We denote

$$(*) = \frac{(|z|^2 + |x|^2 + c)^{1/2}}{(|x|^2 + |y|^2 + c)^{1/2}} |x - y| + \frac{(|z|^2 + |x|^2 + c)^{1/2}}{(|z|^2 + |y|^2 + c)^{1/2}} |y - z|.$$

It must be shown that

$$(*) \geq |x - z|. \quad (68)$$

Since (*) is symmetric, we can assume, without loss of generality, that $x \leq z$. There are three cases to consider.

Case 1) $y \leq x \leq z$. We have

$$\frac{(|z|^2 + |x|^2 + c)^{1/2}}{(|x|^2 + |y|^2 + c)^{1/2}} \geq 1 \quad (69)$$

$$\frac{(|z|^2 + |x|^2 + c)^{1/2}}{(|z|^2 + |y|^2 + c)^{1/2}} \geq 1. \quad (70)$$

These imply that

$$(*) \geq |x - y| + |y - z|. \quad (71)$$

Equation (68) follows from the triangular inequality for the L^2 -norm.

Case 2) $x < y < z$. For any stability constant $c \geq 0$ and for any constant $a \geq 0$, define

$$f(w) := \sqrt{w^4 + 2cw^2 + a}. \quad (72)$$

By looking at the sign of the second derivative of the function, one can verify that it is a convex function. Therefore, by definition of convexity, we have for all $\lambda \in (0, 1)$, with $y = \lambda x + (1 - \lambda)z$

$$f(y) \leq \lambda f(x) + (1 - \lambda)f(z). \quad (73)$$

Choosing the constant

$$a := x^2 y^2 + y^2 z^2 + x^2 z^2 + c(x^2 + y^2 + z^2) + c^2$$

and rearranging the terms, we find

$$\frac{1}{\sqrt{x^2 + z^2 + c}} \leq \frac{\lambda}{\sqrt{z^2 + y^2 + c}} + \frac{1 - \lambda}{\sqrt{x^2 + y^2 + c}}. \quad (74)$$

Finally, by taking

$$\lambda = \frac{|z - y|}{|x - z|}$$

we obtain

$$\frac{|x - z|}{\sqrt{x^2 + z^2 + c}} \leq \frac{|y - z|}{\sqrt{z^2 + y^2 + c}} + \frac{|x - y|}{\sqrt{x^2 + y^2 + c}}. \quad (75)$$

Case 3) $x \leq z \leq y$. We then have

$$\frac{(|z|^2 + |x|^2 + c)^{1/2}}{(|x|^2 + |y|^2 + c)^{1/2}} \geq \frac{|z|}{|y|} \quad (76)$$

$$\frac{(|z|^2 + |x|^2 + c)^{1/2}}{(|z|^2 + |y|^2 + c)^{1/2}} \geq \frac{|x|}{|y|}. \quad (77)$$

These imply that

$$(*) \geq \frac{|z|}{|y|}|x - y| + \frac{|x|}{|y|}|y - z|. \quad (78)$$

From Ptolemy's inequality for \mathbf{R}^N (see Lemma 2.3), the proof is complete. ■

Note that this proof could be generalized for any normalized metric of the form

$$d_n(x, y) := \frac{\|x - y\|}{(\|x\|^p + \|y\|^p + c)^{q/p}} \quad (79)$$

with $p \geq 1$ and $q \geq 1/2$. The idea is to instead use function $f(w) = (w^{2p} + 2cw^p + a)^{q/p}$ in the argument using the convexity of f .

Proof of Lemma 3.10: Consider any two points $x_{(1)}$ and $x_{(2)}$ in the region consisting of the rotation of X around the direction of y and consider a third point

$$z = \lambda x_{(1)} + (1 - \lambda)x_{(2)}, \quad \text{where } 0 \leq \lambda \leq 1 \quad (80)$$

somewhere in the line segment relying these two points. For the convexity of f , we need to show that

$$f(z) \leq \lambda f(x_{(1)}) + (1 - \lambda)f(x_{(2)}). \quad (81)$$

We decompose $x_{(1)}$ and $x_{(2)}$ as $x_{(1)} = a_1y + b_1u$ and $x_{(2)} = a_2y + b_2v$, respectively, where u and v are unit vectors perpendicular to y , i.e., $\langle y, u \rangle = \langle y, v \rangle = 0$ and $\|u\| = \|v\| = 1$. Now, consider two points $x_{(3)} = a_1y + b_1v$ and $x_{(4)} = a_2y + b_2u$ and construct two other points $x_{(5)} = \lambda x_{(1)} + (1 - \lambda)x_{(4)}$ and $x_{(6)} = \lambda x_{(2)} + (1 - \lambda)x_{(3)}$ (see Fig. 5).

Considering

$$z' = \mu x_{(5)} + (1 - \mu)x_{(6)} \quad (82)$$

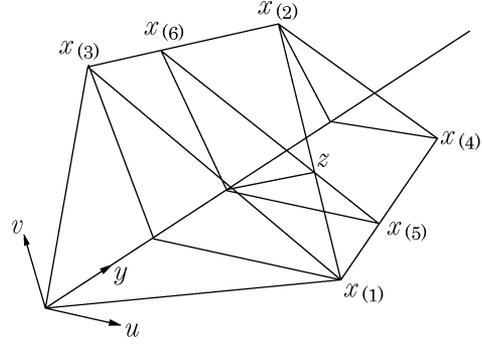


Fig. 5. Points x_1, x_2, \dots, x_6 and z in the proof of Lemma 3.10.

where

$$\mu = \frac{\lambda b_1}{\lambda b_1 + (1 - \lambda)b_2} \quad (83)$$

one can check that, in fact, $z = z'$.

By the convexity of f in 2-D region X and by rotational symmetry, we have

$$f(x_{(5)}) \leq \lambda f(x_{(1)}) + (1 - \lambda)f(x_{(4)}) \quad (84)$$

$$f(x_{(6)}) \leq \lambda f(x_{(2)}) + (1 - \lambda)f(x_{(3)}). \quad (85)$$

Now, we show that

$$f(z) \leq \mu f(x_{(5)}) + (1 - \mu)f(x_{(6)}). \quad (86)$$

Since $x_{(5)}$ and $x_{(6)}$ have the same distance to the line generated by y , $f(x_{(5)}) = f(x_{(6)})$ by rotational symmetry. It remains to show that $f(z) \leq f(x_{(5)})$. We write

$$z = (\lambda a_1 + (1 - \lambda)a_2)y + \lambda b_1u + (1 - \lambda)b_2v \quad (87)$$

$$= ay + bw \quad (88)$$

with $\langle w, y \rangle = 0$ and $\|w\| = 1$ and

$$x_{(5)} = (\lambda a_1 + (1 - \lambda)a_2)y + (\lambda b_1 + (1 - \lambda)b_2)u \quad (89)$$

$$= ay + b'u \quad (90)$$

with $\langle u, y \rangle = 0$ and $\|u\| = 1$. From the triangular inequality for norms, $b \leq b'$. We then use the fact that f is monotonically increasing in a direction perpendicular to y to conclude that $f(z) \leq f(x_{(5)})$. From rotational symmetry, $f(x_{(4)}) = f(x_{(2)})$. From all of the above

$$f(z) \leq f(x_{(5)}) \leq \lambda f(x_{(1)}) + (1 - \lambda)f(x_{(4)}) \quad (91)$$

$$= \lambda f(x_{(1)}) + (1 - \lambda)f(x_{(2)}). \quad (92)$$

■

REFERENCES

- [1] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

- [2] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA: Morgan & Claypool, 2006.
- [3] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [5] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, 2003, pp. 1398–1402.
- [6] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
- [7] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [8] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise-ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
- [9] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011 006-1–011 006-21, Jan. 2010.
- [10] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [11] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Process.*, 1994, vol. 2, pp. 982–986.
- [12] I. Epifanio, J. Gutierrez, and J. Malo, "Linear transform for simultaneous diagonalization of co-variance and perceptual metric matrix in image coding," *Pattern Recognit.*, vol. 36, no. 8, pp. 1799–1811, 2003.
- [13] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 68–80, Jan. 2006.
- [14] V. Laparra, J. Muñoz Mariacutec, and J. Malo, "Divisive normalization image quality metric revisited," *J. Opt. Soc. Amer. A*, vol. 27, no. 4, pp. 852–864, Apr. 2010.
- [15] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Design of a linear image estimator optimized for the structural similarity index and its application to image denoising," in *Proc. IEEE Int. Conf. Image Proc.*, Atlanta, GA, 2006, pp. 2637–2640.
- [16] A. Rehman, Z. Wang, D. Brunet, and E. R. Vrscay, "SSIM-inspired image denoising using sparse representations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, 2011, pp. 1121–1124.
- [17] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "SSIM-optimal linear image restoration," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, 2008, pp. 765–768.
- [18] M. Temerinac-Ott and M. Burkhardt, "Multichannel image restoration based on optimization of the structural similarity index," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 2009, pp. 812–816.
- [19] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 857–872, Jun. 2008.
- [20] A. N. Avnani, "Exact global histogram specification optimized for structural similarity," *Opt. Rev.*, vol. 16, no. 6, pp. 613–621, Nov. 2009.
- [21] D. Brunet, E. R. Vrscay, and Z. Wang, M. Kamel and A. Campilho, Eds., "Structural similarity-based approximation of signals and images using orthogonal bases," in *Proc. Int. Conf. Image Anal. Recognit.*, 2010, vol. 6111, LNCS, pp. 11–22, Springer: Heidelberg, Germany.
- [22] Z. Wang, Q. Li, and X. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," in *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, 2007, pp. II-121–II-124.
- [23] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on SSIM index of quantized images," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1624–1639, Sep. 2008.
- [24] T. Richter and K. J. Kim, "A MS-SSIM optimal JPEG 2000 encoder," in *Proc. Data Compression Conf.*, Snowbird, UT, 2009, pp. 401–410.
- [25] C.-L. Yang, R.-K. Leung, L.-M. Po, and Z.-Y. Mai, "An SSIM-optimal H.264/AVC inter frame encoder," in *Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst.*, Shanghai, China, 2009, pp. 291–295.
- [26] Y. H. Huang, T. S. Ou, P. Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.
- [27] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Rate-SSIM optimization for video coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, 2011, pp. 833–836.
- [28] Z. Wang and E. P. Simoncelli, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," in *Proc. Human Vis. Electron. Imag. IX-SPIE*, 2004, vol. 5292, pp. 99–108.
- [29] Z. Wang and E. P. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *J. Vis.*, vol. 8, no. 12, pp. 1–13, 2008.
- [30] J. Ostergaard, M. S. Derpich, and S. S. Channappayya, "The high resolution rate-distortion function under the structural similarity index," *EURASIP J. Appl. Signal Process.*, vol. 2011, pp. 1–7, 2011.
- [31] M. S. Klamkin and A. Meir, "Ptolemy's inequality, chordal metric, multiplicative metric," *Pacific J. Math.*, vol. 101, no. 2, pp. 389–392, 1982.
- [32] P. A. Hästö, "A new weighted metric: The relative metric I," *J. Math. Anal. Appl.*, vol. 274, no. 1, pp. 38–58, 2002.
- [33] P. A. Hästö, "A new weighted metric: The relative metric II," *J. Math. Anal. Appl.*, vol. 301, no. 2, pp. 336–353, 2005.
- [34] P. N. Yianilos, "Normalized forms for two common metrics," NEC Res. Inst., Rep. 91-082-9027-1, 1991 [Online]. Available: <http://www.pnylab.com/pny>, Cambridge Univ. Press, Rev. 7/7/2002.
- [35] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitanyi, "The similarity metric," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3250–3264, Dec. 2004.
- [36] T. Apostol, "Ptolemy's inequality and the chordal metric," *Math. Mag.*, vol. 40, no. 5, pp. 233–235, Nov. 1967.
- [37] I. J. Schoenberg, "A remark on M.M. Day's characterization of inner product spaces and a conjecture of L.M. Blumenthal," *Proc. Amer. Soc.*, vol. 3, no. 6, pp. 961–964, Dec. 1952.
- [38] D. Brunet, E. R. Vrscay, and Z. Wang, "A class of image metrics based on the structural similarity quality index," in *Proc. Int. Conf. Image Anal. Recognit.*, M. Kamel and A. Campilho, Eds., 2011, vol. 6754, LNCS, pp. 264–275, Springer: Heidelberg, Germany.
- [39] L.-G. Huang and X. Zhang, "Cone metric spaces and fixed point theorems of contractive mappings," *J. Math. Anal. Appl.*, vol. 332, no. 2, pp. 1468–1476, 2007.
- [40] D. M. Rouse and S. S. Hemami, "Understanding and simplifying the structural similarity metric," in *Proc. IEEE Int. Conf. Image Process.*, 2008, pp. 1188–1191.
- [41] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack, "Image and Video Quality Assessment Research at LIVE [Online]. Available: <http://live.ece.utexas.edu/research/quality/2010>
- [42] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, "Metrics performance comparison for color image database," in *Proc. 4th Int. Workshop Video Process. Quality Metrics Consumer Electron.*, Scottsdale, AZ, 2009, pp. 1–6.
- [43] N. Ponomarenko and K. Egiazarian, Tampere Image Database 2008 TID2008 [Online]. Available: <http://www.ponomarenko.info/tid2008.htm> 2008
- [44] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
- [45] N. Lu, *Fractal Imaging*. New York: Academic, 1997.
- [46] *Handbook of Generalized Convexity and Generalized Monotonicity*, N. Hadjisavvas, S. Komlósi, and S. Schaible, Eds. New York: Springer-Verlag, 2005.
- [47] X. Li, "Collective sensing: A fixed-point approach in the metric space," in *Proc. SPIE Visual Comm. Image Process.*, P. Frossard, H. Li, F. Wu, B. Girod, S. Li, and G. Wei, Eds., 2010, vol. 7744, pp. J1–J10.



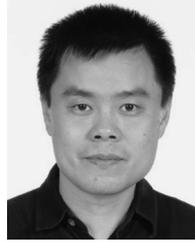
Dominique Brunet received the B.Sc. and M.Sc. degrees in mathematics from the Université Laval, Québec City, QC, Canada, in 2005 and 2007, respectively. He is currently working toward the Ph.D. degree in the Department of Applied Mathematics, University of Waterloo, Waterloo, ON, Canada.

His research interests include image processing, image-quality assessment, and mathematical and statistical image modeling.



Edward R. Vrscay (M'97) received the B.Sc. degree in chemistry and the M.Math. and Ph.D. degrees in applied mathematics from the University of Waterloo, Waterloo, ON, Canada, in 1975, 1977, and 1984, respectively.

From 1984 to 1986, he was a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellow (NSERC) and a Visiting Assistant Professor with the School of Mathematics, Georgia Institute of Technology, Atlanta. In 1986, he returned to the Department of Applied Mathematics, University of Waterloo as an NSERC University Fellow. Since 1994, he has been a Professor with the department. He is a coeditor (with M. Barnsley and D. Saupe) of the book *Fractals in Multimedia* (2001) and a coauthor (with H. Kunze, D. La Torre, and F. Mendivil) of the book *Fractal-Based Methods in Analysis* to be published very shortly by Springer Verlag. His original research interests were in mathematical physics, particularly quantum mechanics. At Georgia Institute of Technology, he became involved with dynamical systems, fractal geometry, and fractal image coding. Since then, he has worked on the mathematical formulation and solution of inverse problems for fractal-based approximation. More recently, he has been involved in problems of mathematical imaging, including denoising, superresolution, image self-similarity, the representation of images in various spaces (e.g., multifunctions), and the mathematical formulation of the structural similarity index.



Zhou Wang (S'97–A'01–M'02) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, Austin, in 2001.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He has more than 100 publications related to his fields of interest with more than 8000 citations (Google Scholar). He is an author of *Modern Image Quality Assessment* (Morgan and Claypool, 2006). His research interests include image processing, coding,

and quality assessment, computational vision and pattern analysis, multimedia communications, and biomedical signal processing.

Dr. Wang has served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–present), the IEEE SIGNAL PROCESSING LETTERS (2006–2010), and *Pattern Recognition* (2006–present), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2007–2009), *European Association for Signal Processing Journal of Image and Video Processing* (2009–2010), and *Signal, Image and Video Processing* (2011–present). He was a recipient of the 2009 IEEE Signal Processing Best Paper Award, the International Conference on Image Processing 2008 IBM Best Student Paper Award (as a senior author), and the 2009 Ontario Early Researcher Award.