Applications of Mathematics

Gemma Piella

A framework to combine vector-valued metrics into a scalar-metric: Application to data comparison

Applications of Mathematics, Vol. 68 (2023), No. 2, 143–152

Persistent URL: http://dml.cz/dmlcz/151609

A FRAMEWORK TO COMBINE VECTOR-VALUED METRICS INTO A SCALAR-METRIC: APPLICATION TO DATA COMPARISON

Gemma Piella, Barcelona

Received April 11, 2021. Published online November 22, 2021.

Abstract. Distance metrics are at the core of many processing and machine learning algorithms. In many contexts, it is useful to compute the distance between data using multiple criteria. This naturally leads to consider vector-valued metrics, in which the distance is no longer a real positive number but a vector. In this paper, we propose a principled way to combine several metrics into either a scalar-valued or vector-valued metric. We illustrate our framework by reformulating the popular structural similarity (SSIM) index and a simple case of the Wasserstein distance used for optimal transport.

Keywords: generalized metric; vector-valued metric; scalarization; image comparison; structural similarity index

MSC 2020: 46A40, 94A08

1. Introduction

Data comparison lies at the core of many data processing tasks, such as classification, clustering, data mining or image quality assessment. To quantify similarity between data of any type, a natural approach is to consider them as elements of some metric space, in which distance between points represents the degree of dissimilarity. Metrics induce a topology on the space that satisfies helpful properties for its analysis. Metric spaces are useful for bounding errors, proving convergence, and similarity searching.

Distance metrics are often based on some underlying norm. For example, a distance between two vectors x, y in a real normed vector space (e.g., Euclidean space $\mathbb{R}^n$) can be defined as

$$d(x, y) = \|x - y\|$$

DOI: 10.21136/AM.2021.0090-21

with $\|\cdot\|$ being an l^p-norm with $1 \leqslant p \leqslant \infty$. In such cases, we have that for any two vectors u, v in that space with $\|u\| = \|v\|$:

$$d(x, x + u) = d(x, x + v).$$

An example, inspired by [18], is depicted in Fig. 1. Here, an original signal x is combined with two different signals u and v with the same energy, resulting in two different distorted signals: $x + u$ looks perceptually the same as x, whereas $x + v$ looks different.

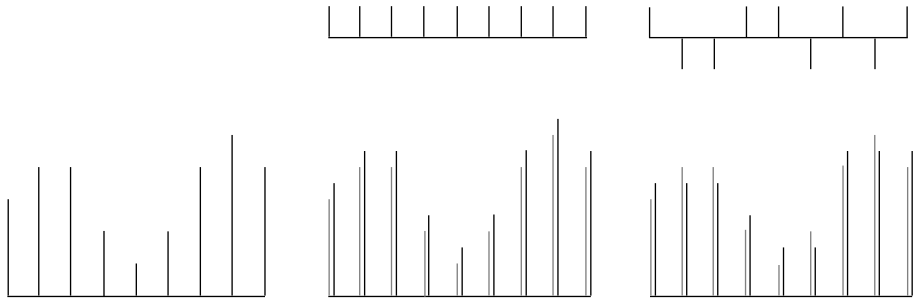

Figure 1. The original signal (bottom left) is added to each distortion signal (top) to generate two distorted signals (bottom middle and right): one 'looks similar' to the original (bottom middle) whereas the other 'looks different'.

We can account mathematically for such phenomena by decomposing signal x into a component along u and a component orthogonal to u:

$$x = \langle x, u \rangle \frac{u}{\|u\|^2} + x',$$

where x' is orthogonal to u. We can then define a new norm, e.g.,

$$\|x\|' = \kappa |\langle x, u \rangle| + \|x'\|, \tag{1.1}$$

where κ is a constant. In this way, by choosing an appropriate decomposition and distance measure able to capture structure that correlates with human perception, signals can be compared in a more perceptual way.

The choice of an optimal basis for signal representation and the distance or similarity measure to use remains a challenging problem that is ubiquitous in the field of signal and image processing. Naturally, the question arises of what is optimal, which depends on the signal and task at hand. Indeed, distance metric learning (i.e., choosing an appropriate distance learned from the data) has shown to greatly improve the performance of distance-based algorithms. Recently, several deep learning based models have been explored for metric learning. For example, convolutional neural networks have been used to learn distance metrics for face recognition [12], image classification [16], image segmentation [13], and image quality assessment [9].

On the other hand, in many contexts, it is useful to compute the distance between signals using multiple criteria. This naturally leads to consider vector-valued metrics, in which the distance is no longer a real positive number but a vector. In this paper, rather than focusing on developing a new (dis)similarity measure, we propose a principled way to combine several metrics into either a scalar-valued or vector-valued metric. We show how some well-known (dis)similarity measures can be reformulated within this framework. The groundwork laid here may facilitate future advances in quantitative performance metrics, in particular in the context of signal or image processing when signals should be compared at multiple levels of resolution or using multiple criteria to provide complementary insights. For example, in image quality assessment (see, e.g., [2] for a survey of numerous approaches), different concurrent distortions may require different quality evaluation methods.

2. Generalized metrics

2.1. Formal framework. Let V be a vector space with subspaces V_j, $j = 1, \dots, J$, which together span the entire space. Let P_j be projections on V with range V_j. Note that we do not assume that spaces V_j are disjoint or that projections P_j are orthogonal. We say that the family $\Pi = \{P_j\}_{j=1}^J$ is total if for $x \in V$,

$$P_j x = 0 \quad \forall\, j = 1, \dots, J \Rightarrow x = 0.$$

If Π is total and $\Phi\colon \mathbb{R}^J \to \mathbb{R}_+$ is a norm, we can define a norm $\|\cdot\|_\Pi$ on V as

$$\|x\|_\Pi = \Phi \quad (j \mapsto \|P_j x\|_j), \tag{2.1}$$

where $\|\cdot\|_j$ is a norm on V_j. For example, $\|x\|_\Pi = \sum_{j=1}^J \|P_j x\|_j$ is a norm. The norm Φ is said to be monotone if for any $u, v \in \mathbb{R}_+^J$,

$$\Phi(u + v) \geqslant \Phi(u).$$

This holds, for instance, for all weighted l^p-norms

$$\Phi(u) = \left(\sum_{j=1}^J w_j |u_j|^p \right)^{1/p}$$

with $1 \leqslant p < \infty$ and $w_j > 0$ for all $j = 1, \dots, J$.

The following theorem shows that if Φ is monotone, we can use it to define a distance metric on V from a vector of distance metrics, which is not necessarily the induced distance metric from norm $\|\cdot\|_\Pi$.

Theorem 2.1. *Let $\{V_j\}_{j=1}^J$ be subspaces spanning a vector space V and $\Pi = \{P_j\}_{j=1}^J$, with each $P_j\colon V \mapsto V_j$, being a family of projections that is total. Let Φ be a norm that is monotone on $\mathbb{R}^J$, and $\{d_j\}_{j=1}^J$ be distance metrics on, respectively, $\{V_j\}_{j=1}^J$. Then the mapping $D^\Pi\colon V \times V \mapsto \mathbb{R}_+$ given by*

$$D^\Pi(x,y) = \Phi \quad (j \mapsto d_j(P_j x, P_j y)) \tag{2.2}$$

defines a distance metric on V for all $x, y \in V$.

P r o o f. If $x = y$, then $d_j(P_j x, P_j y) = 0$ for all $j = 1, \dots, J$, and thus $D^\Pi(x,y) = \Phi(0) = 0$. On the other hand, if $D^\Pi(x,y) = 0$, since Φ is a norm, $d_j(P_j x, P_j y) = 0$ for all $j = 1, \dots, J$. This implies that $P_j x = P_j y$, since d_j is a metric. Thus, $P_j(x-y) = 0$ for all $j = 1, \dots, J$, and since the family $\{P_j\}$ is total, $x = y$. Therefore, $D^\Pi(x,y) = 0 \Leftrightarrow x = y$. The symmetry property is straightforward. The triangle inequality property remains to be proved. Let $d(x,y) \in \mathbb{R}_+^J$ be the vector whose components are d_j. Since $d_j, j = 1, \dots, J$, are metrics, they satisfy the triangular inequality and from the assumption that Φ is monotone, we have that for any $x, y, z \in V$,

$$\Phi(d(x,y)) \leqslant \Phi((d(x,z) + d(z,y)) \leqslant \Phi(d(x,z)) + \Phi(d(z,y)).$$

□

R e m a r k 2.1. Set $\varrho_j(x,y) := d_j(P_j x, P_j y)$. Then the family $\{\varrho_j\}_{j=1}^J$ is a separating family of pseudometrics, i.e., for any $x, y \in V$, if $x \neq y$, then $\varrho_j(x,y) > 0$ for some $j \in \{1, \dots, J\}$.

Norm $\Phi\colon \mathbb{R}^J \to \mathbb{R}_+$ can be seen as a scalarization mapping of the vector of distance metrics:

$$d(x,y) = (d_1(P_1 x, P_1 y), \dots, d_J(P_J x, P_J y)). \tag{2.3}$$

Scalarization methods are an important tool in the study of vector optimization. They allow combining different measures into a single one, which in turn allows a simple and total ordering among vectors. However, for certain applications, it may be desirable to work directly with a vector-valued function d. In this case, the standard ordering $\leqslant$ of real numbers is replaced by a partial ordering $\preceq$ defined on space $\mathbb{R}_+^J$ (or another vector space). For example, for $a, b \in \mathbb{R}_+^J$, the coordinate-wise ordering is defined as: $a \preceq b \Leftrightarrow b - a \in \mathbb{R}_+^J$ (i.e., $a_j \leqslant b_j$ for all $j = 1, \dots, J$). Thus, unlike in the real case, distances of pairs of elements may not be comparable. We need the order to be total so that any two elements are comparable (i.e., either $a \preceq b$ or $b \preceq a$).

2.2. Vector metrics. The concept of (distance) metric can be extended by replacing the set of real numbers $\mathbb{R}$ in the codomain of the metric by an arbitrary partially ordered set. Such metric spaces have been defined under various names

(vector-valued metric space, generalized metric space, K-metric space, cone metric space, topological ordered vector cone metric space, etc). Huang et al. [8] popularized the notion of cone metric space, where $\mathbb{R}$ is replaced by Banach spaces. Next, we present a short review on the concepts of cone metric spaces.

Definition 2.1. An ordered vector space E is defined as a vector space endowed with a partial ordering $\preceq$ if for all $x, y, z \in E$ it satisfies

(1) $x \preceq y \Longrightarrow x + z \preceq y + z$;
(2) $x \preceq y \Longrightarrow \alpha x \preceq \alpha y$ for any $\alpha \in \mathbb{R}_+$.

Any partial order $\preceq$ defines a convex cone $C \subset E$ by $C = \{x \in E \mid 0 \preceq x\}$, and any convex cone defines a partial ordering by $x \preceq y \Leftrightarrow y - x \in C$ for all $x, y \in C$. For example, the cone representing the component-wise order in $E = \mathbb{R}^n$ is the positive orthant $\mathbb{R}^n_+$.

Definition 2.2. Let E be a real Banach space equipped with the partial ordering $\preceq$ with respect to cone $C \subset E$. A cone metric on a set V is a mapping $d\colon V \times V \mapsto C$ satisfying for all $x, y, z \in V$:

(1) $d(x, y) \succeq 0$ and $d(x, y) = 0 \Leftrightarrow x = y$;
(2) $d(x, y) = d(y, x)$;
(3) $d(x, y) \preceq d(x, z) + d(z, y)$.

Then (V, d, E) is called a cone metric space.

We focus on the particular case where this Banach space is $\mathbb{R}^J_+$, with the coordinate wise order $x \preceq y \Leftrightarrow x_j \leqslant y_j$ for all $j = 1, \ldots, J$. This cone metric is also called (J-dimensional) vector-valued metric. Note that the component-wise ordering corresponds to giving the same importance to all vector components. This could be too restrictive in some applications in which it is convenient to assign priorities to some of the components; e.g., by imposing a lexicographic ordering $x \preceq y \Leftrightarrow \exists j > 0$ so that for all $i < j$ we have that $x_i = y_i$ and $x_j < y_j$. That is, $x \preceq y$ if one of the terms $x_j < y_j$ and all preceding terms are equal. This ordering yields the familiar sorting used in dictionaries. It is complete in the sense that any two elements in $\mathbb{R}^J$ are comparable (i.e., either $x \succeq y$ or $y \succeq x$). The lexicographic cone of $\mathbb{R}^J$ is defined as the set of all vectors whose first nonzero coordinate (if any) is positive: $C_{\text{lex}} = \{0\} \cup \{x \in \mathbb{R}^J \mid x_1 = \ldots = x_k = 0,\ x_{k+1} > 0 \text{ for some } k \in \{1, \ldots, J-1\}\}$.

It has been shown that a cone metric space is metrizable [1], [10]. In other words, a cone metric d can be scalarized using a suitable mapping Φ so that $D(x, y) = \Phi(d(x, y))$ results in a distance metric in the usual scalar-valued sense, i.e., $D\colon V \times V \mapsto \mathbb{R}_+$, that induces the same topology as the cone metric topology induced by d. Theorem 2.1 provides a particular procedure of scalarization.

Example 2.1. Function $d\colon \mathbb{R}^J \times \mathbb{R}^J \mapsto \mathbb{R}^J_+$ given by

$$d(x,y) = (\alpha_1|x_1 - y_1|, \ldots, \alpha_J|x_J - y_J|)$$

with $\alpha_j \in \mathbb{R}_+$ (for all $j = 1, \ldots, J$) is an J-dimensional vector-valued metric. The (possibly weighted) l^1-norm of this vector defines a scalar distance metric

$$D(x,y) = \sum_{j=1}^{J} w_j \alpha_j |x_j - y_j|$$

with $w_j > 0$ for all $j = 1, \ldots, J$. Note that this is a particular case of Theorem 2.1, where Φ is the weighted l^1-norm, P_j is the projection on coordinate j, and $d_j(a,b) = |a-b|$ for any $a, b \in \mathbb{R}$.

Example 2.2. Consider an image x whose domain is divided into (possibly overlapping) square windows $w \in W$, and define projection $P_w x = x|_w$, where $x|_w$ denotes the image x evaluated on window w. Denote the range by V_w and decompose it as

$$V_w = V_{w,0} \oplus V_{w,1},$$

where $V_{w,0}$ is the space spanned by u_w, the function identically one on w, and $\oplus$ denotes direct sum. The projection on $V_{w,0}$ gives the average $x_{w,0} = \bar{x}_w$, whereas the projection on $V_{w,1}$ gives the zero-mean image $x_{w,1} = x_w - \bar{x}_w$.

We can compare two images x, y by measuring their distance in these two projections for each windowed region. That is, for each V_w, define the vector-value metric $d_w\colon V \times V \mapsto \mathbb{R}^2_+$ as

$$d_w(x_w, y_w) = (d_{w,0}(\bar{x}_w, \overline{y}_w), d_{w,1}(x_w - \bar{x}_w, y_w - \overline{y}_w)),$$

where $d_{w,i}$, $i = 0, 1$ is a distance metric on $V_{w,i}$, and define a distance metric over the whole image as

$$D^{\Pi}(x,y) = \Phi_w(w \mapsto \Phi_i(i \mapsto d_{w,i}(x_{w,i}, y_{w,i})).$$

For example, Φ_i (for $i = 0, 1$) could be the l^2-norm and Φ_w a weighted l^1-norm, so that it computes the mean or assigns a spatially varying weight to different windows:

$$D^{\Pi}(x,y) = \frac{1}{|W|} \sum_{w \in W} (d^2_{w,0}(\bar{x}_w, \overline{y}_w) + d^2_{w,1}(x_w - \bar{x}_w, y_w - \overline{y}_w))^{1/2},$$

where $|W|$ is the cardinality of W, i.e., the number of windows.

3. Application to data comparison

3.1. SSIM revisited. We show that the well-known structural similarity index (SSIM) proposed by Wang et al. [17], [19] can be reformulated within the framework in Section 2.1. We first briefly review the SSIM index.

Given two images x and y in $\mathbb{R}^n_+$, let $\bar{x}$ denote the mean of x, σ_x^2 be the variance of x, and σ_{xy} be the covariance between x and y. The SSIM index is defined as

$$\mathrm{SSIM}(x,y) = \frac{2\bar{x}\overline{y} + c_1}{\bar{x}^2 + \overline{y}^2 + c_1} \cdot \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \cdot \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \tag{3.1}$$

where c_1, c_2, and c_3 are small constants that account for the saturation effects of the visual system at low luminance and contrast, and which guarantee stability when the denominators are close to zero. The first component in (3.1) measures the average luminance distortion and has a dynamic range of $[0,1]$. The second component measures the contrast (variance) distortion and its range is also $[0,1]$. The third factor in (3.1) is the correlation coefficient between x and y, referred to as structural term. The maximum value $\mathrm{SSIM}(x,y) = 1$ is achieved when x and y are identical. Thus, SSIM is a measure for the similarity of two images and takes values between -1 and 1.

Since image signals are generally non-stationary, it is appropriate to measure the SSIM index over local regions and then combine the different results into a single measure. In [17] the authors propose to use a sliding window approach: starting from the top-left corner of the two images x, y, a sliding window of fixed size moves pixel by pixel over the entire image until the bottom-right corner is reached. For each window w, the local quality index $\mathrm{SSIM}(x, y \mid w)$ is computed for the image patches whose pixels lie in the sliding window w. Finally, the overall image quality index is computed by pooling, e.g., averaging, all local quality indices:

$$\mathrm{SSIM}(x,y) = \frac{1}{|W|} \sum_{w \in W} \mathrm{SSIM}(x, y \mid w), \tag{3.2}$$

where W is the family of all windows and $|W|$ is the cardinality of W.

Wang et al. [19] compared (under several types of distortions) their quality index with existing image measures, such as the mean squared error (MSE), as well as with subjective evaluations. Their main conclusion was that SSIM outperforms MSE due to the index's ability of measuring structural distortions, in contrast to MSE, which is highly sensitive to the l^2-norm of errors. Over the years, SSIM has gained widespread popularity as a tool to assess the quality of images or as an objective function in optimization problems, and several variants of the SSIM index have been proposed, e.g., using a multiscale [20] or a gradient-based [5] formulation, or including texture [11]. SSIM has even been used as a loss function for deep learning systems [21], and it is included in platforms such as Tensorflow.

SSIM is not a distance metric. However, as shown in [3], it can be reformulated by considering cone-metric distances in which the cone is the positive orthant. Given $x, y \in \mathbb{R}^n_+$ and choosing $c_3 = c_2/2$, the SSIM index (3.1) can be decomposed as

$$\mathrm{SSIM}(x, y) = S_1(x, y)S_2(x, y),$$

where $S_1(x, y)$ is the luminance distortion term defined as

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

and $S_2(x, y)$ is the product of the contrast and structural distortion terms:

$$S_2(x, y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}.$$

Then $d_1(x, y) = \sqrt{1 - S_1(x, y)}$ and $d_2(x, y) = \sqrt{1 - S_2(x, y)}$ are distance metrics on $\mathbb{R}^n$, and $d = (d_1, d_2) \in \mathbb{R}^2_+$ is a cone-metric distance on $\mathbb{R}^n$. Thus,

$$\mathrm{SSIM}(x, y) = (1 - d_1^2(x, y))(1 - d_2^2(x, y)).$$

Moreover, it can be shown that $d_1(x, y) = d_1(\bar{x}, \overline{y})$ and $d_2(x, y) = d_2(x - \bar{x}, y - \overline{y})$, as in Example 2.2. In particular,

$$\text{(3.3a)} \qquad d_1(x, y) = \frac{|\bar{x} - \overline{y}|}{\sqrt{\bar{x}^2 + \overline{y}^2 + c_1}} = d_1(\bar{x}, \overline{y}),$$

$$\text{(3.3b)} \qquad d_2(x, y) = \frac{\|(x - \bar{x}) - (y - \overline{y})\|}{\sqrt{\|x - \bar{x}\|^2 + \|y - \overline{y}\|^2 + (n - 1)c_2}} = d_2(x - \bar{x}, y - \overline{y}),$$

which can be seen as normalized root mean square errors. This is interesting, since it suggests a link with MSE, which was already noticed experimentally and theoretically by Horé and Ziou [7], and Dosselman and Yang [6]. Indeed, S_2 essentially computes the MSE between normalized patches (i.e., using $(x - \bar{x})/\sigma_x$ instead of x). Such a normalization—similar to a divisive normalisation [15]—accounts for the contrast masking effect in the visual system.

3.2. Wasserstein distance between Gaussian distributions revisited. For some image processing tasks, it is more relevant to compare probability density functions rather than the intensity levels themselves. One way of doing so is by Monge-Kantarovich optimal transport and the associated Wasserstein distance metric [14]. Here, we focus on the simple case where the probability distributions are

Gaussians and reformulate the 2-Wasserstein distance using the proposed framework in Section 2.1.

Let $X \sim (\mu_x, \Sigma_x)$, $Y \sim (\mu_y, \Sigma_y)$ be two Gaussian distributions on $\mathbb{R}^n$, $n \geqslant 1$. The 2-Wasserstein distance between distributions X and Y has a closed-form expression

$$W_2^2(X, Y) = \|\mu_x - \mu_y\|^2 + d_B^2(\Sigma_x, \Sigma_y), \tag{3.4}$$

where d_B is the Bures metric proposed in quantum information geometry [4] and is defined as

$$d_B^2(A, B) := \operatorname{tr}(A + B - 2(A^{1/2}BA^{1/2})^{1/2})$$

for semi-definite positive matrices A, B, where $\operatorname{tr}(\cdot)$ is the trace operator. We can rewrite the Wasserstein distance in (3.4) as

$$W_2(X, Y) = \Phi(d(X, Y)) = \Phi(d_1(\mu_x, \mu_y), d_2(\Sigma_x, \Sigma_y)),$$

where Φ is the l^2-norm, d_1 is the l^2-distance, and d_2 corresponds to the Bures distance. Furthermore, $d_1(\mu_x, \mu_y) = d_1(P_1 x, P_1 y)$ with P_1 being the expectation operator, and $d_2(\Sigma_x, \Sigma_y) = d_2(P_2 x, P_2 y)$ with P_2 being the covariance operator. If the covariance matrices are diagonal, i.e., $\Sigma_x = \operatorname{diag} \sigma_x, \Sigma_y = \operatorname{diag} \sigma_y$, the Bures distance reduces to the Hellinger metric, i.e., $d_2(\sigma_x, \sigma_y) = \|\sqrt{\sigma_x} - \sqrt{\sigma_y}\|$.

4. Conclusions

In this work, we propose to use vector-valued distance metrics to simultaneously consider several metrics, and we provide a principled way to combine them into a single-valued metric. This naturally arises in many contexts of signal and image processing tasks, where signals are compared at multiple levels of resolution or using multiple criteria to provide complementary insights, or when dealing with images of higher dimensions, such as color and hyperspectral images. We illustrate the framework by reformulating the popular SSIM index for image comparison and the 2-Wasserstein distance between Gaussian distributions.

References

[1] *A. Amini-Harandi, M. Fakhar*: Fixed point theory in cone metric spaces obtained via the scalarization method. Comput. Math. Appl. *59* (2010), 3529–3534. zbl MR doi

[2] *S. Athar, Z. Wang*: A comprehensive performance evaluation of image quality assessment algorithms. IEEE Access *7* (2019), 140030–140070. doi

[3] *D. Brunet, E. R. Vrscay, Z. Wang*: On the mathematical properties of the structural similarity index. IEEE Trans. Image Process. *21* (2012), 1488–1499. zbl MR doi

[4] *D. Bures*: An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras. Trans. Am. Math. Soc. *135* (1969), 199–212. zbl MR doi
[5] *G.-H. Chen, C.-L. Yang, S.-L. Xie*: Gradient-based structural similarity for image quality assessment. IEEE International Conference on Image Processing. IEEE, Piscataway, 2006, pp. 2929–2932. doi
[6] *R. Dosselmann, X. D. Yang*: A comprehensive assessment of the structural similarity index. Signal Image Video Process. *5* (2011), 81–91. doi
[7] *A. Horé, D. Ziou*: Image quality metrics: PSNR vs. SSIM. 20th International Conference on Pattern Recognition. IEEE, Piscataway, 2010, pp. 2366–2369. doi
[8] *L.-G. Huang, X. Zhang*: Cone metric spaces and fixed point theorems of contractive mappings. J. Math. Anal. Appl. *332* (2007), 1468–1476. zbl MR doi
[9] *L. Kang, P. Ye, Y. Li, D. Doermann*: Convolutional neural networks for no-reference image quality assessment. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Piscataway, 2014, pp. 1733–1740. doi
[10] *M. Khani, M. Pourmahdian*: On the metrizability of cone metric spaces. Topology Appl. *158* (2011), 190–193. zbl MR doi
[11] *C. Li, A. C. Bovik*: Content-partitioned structural similarity index for image quality assessment. Signal Process., Image Commun. *25* (2010), 517–526. doi
[12] *Y. Li, G. Wang, L. Nie, Q. Wang, W. Tan*: Distance metric optimization driven convolutional neural network for age invariant face recognition. Pattern Recognition *75* (2018), 51–62. doi
[13] *Z. Ma, S. Zhou, X. Wu, H. Zhang, W. Yan, S. Sun, J. Zhou*: Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning. Phys. Med. Biol. *64* (2019), Article ID 025005. doi
[14] *C. Villani*: Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften 338. Springer, Berlin, 2009. zbl MR doi
[15] *M. J. Wainwright, O. Schwartz, E. P. Simoncelli*: Natural image statistics and divisive normalization. Probabilistic Models of the Brain: Perception and Neural Function. MIT Press, Cambridge, 2002, pp. 203–222.
[16] *C. Wang, G. Peng, B. De Baets*: Deep feature fusion through adaptive discriminative metric learning for scene recognition. Inf. Fusion *63* (2020), 1–12. doi
[17] *Z. Wang, A. C. Bovik*: A universal image quality index. IEEE Signal Process. Lett. *9* (2002), 81–84. doi
[18] *Z. Wang, A. C. Bovik, L. Lu*: Why is image quality assessment so difficult? IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, Piscataway, 2002, pp. IV-3313–IV-3316. doi
[19] *Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli*: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. *13* (2004), 600–612. doi
[20] *Z. Wang, E. P. Simoncelli, A. C. Bovik*: Multiscale structural similarity for image quality assessment. 37th Asilomar Conference on Signals, Systems & Computers. IEEE, Piscataway, 2003, pp. 1398–1402. doi
[21] *H. Zhao, O. Gallo, I. Frosio, J. Kautz*: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging *3* (2017), 47–57. doi

Author's address: *Gemma Piella*, Universitat Pompeu Fabra, Tanger 122, 08018 Barcelona, Spain, e-mail: gemma.piella@upf.edu.