MDPI

Article

An Information-Theoretic Perspective on Proper Quaternion Variational Autoencoders

Eleonora Grassucci *, Danilo Comminiello and Aurelio Uncini

Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy; danilo.comminiello@uniroma1.it (D.C.); aurelio.uncini@uniroma1.it (A.U.)
* Correspondence: eleonora.grassucci@uniroma1.it

Abstract: Variational autoencoders are deep generative models that have recently received a great deal of attention due to their ability to model the latent distribution of any kind of input such as images and audio signals, among others. A novel variational autoncoder in the quaternion domain $\mathbb{H}$, namely the QVAE, has been recently proposed, leveraging the augmented second order statics of $\mathbb{H}$-proper signals. In this paper, we analyze the QVAE under an information-theoretic perspective, studying the ability of the $\mathbb{H}$-proper model to approximate improper distributions as well as the built-in $\mathbb{H}$-proper ones and the loss of entropy due to the improperness of the input signal. We conduct experiments on a substantial set of quaternion signals, for each of which the QVAE shows the ability of modelling the input distribution, while learning the improperness and increasing the entropy of the latent space. The proposed analysis will prove that proper QVAEs can be employed with a good approximation even when the quaternion input data are improper.

Keywords: variational autoencoder; generative learning; quaternion neural networks; properness; second-order circularity

1. Introduction

Generative learning models have been recently gained considerable attention due to their surprising performance in producing highly realistic signals of various types [1–4]. They have been successfully employed in a wide variety of applications, such as image-to-image translation [5], image fusion [6], face de-identification [7], natural language generation [8], data augmentation on ancient handwritten characters [9], MRI super-resolution [10], brain tumor growth prediction [11], generative modeling of structured-data [12].

Among such generative methods, the variational autoencoders (VAEs) have been proven to perform stochastic variational inference and learning even for large datasets and intractable posterior distributions [13]. The main advantages of the VAEs rely on their capability of learning smooth latent representations of the input data. This has led VAEs to be used in several fields of applications, including high-quality image generation [14], speech enhancement [15], music style transfer [16], data augmentation [17], 3D scene generation [18], gesture similarity analysis [19], text generation [20] and sequential recommendation [21], among others.

Recent advances on VAEs focus both on theoretical aspects, such as the improvement of the stochastic inference approach [22], and on architectural aspects, such as the use of different types of latent variables to learn local and global structures [23], the definition of hierarchical schemes [24], rather than the use of a multi-agent generator [25]. Among the most recent VAE models, we focus on the quaternion-valued variational autoencoder (QVAE), which exploits the properties of quaternion algebra to improve performance, on one hand, and to significantly reduce the overall number of network parameters [26], on the other hand.

The basic approach on which relies the QVAE is the learning in the quaternion domain, which results in significant advantages in the presence of multidimensional input data (mainly 3D and 4D data) showing some inter-channel correlations. This properties have been widely exploited in shallow learning models, such as linear and nonlinear adaptive filters [27–34]. Another fundamental property of quaternion-valued learning is the Hamilton product, which has recently favored the proliferation of convolutional neural networks in the quaternion domain [35–38]. Due to their capabilities, quaternion-valued learning methods have been applied in several applications, including spoken language understanding [39], color image processing [40,41], 3D audio [42,43], speech recognition [44], image generation [45], quantum mechanics [46], risk diversification [47], gait data analysis [48].

As with any other complex-valued and hypercomplex-valued algebra, the quaternion-valued learning can rely on the properties of the second-order statistics, which are fundamental to characterize the input data based on their correlation. A characterization of correlation is achieved in terms of covariance and pseudo-covariance. In particular, random variables and processes with a vanishing pseudo-covariance are called proper [49–51]. Properness is preserved under affine transformations. Moreover, it should be considered that the multivariate Gaussian density assumes a natural form only for proper random variables. Properness in the quaternion domain, also denoted as $\mathbb{H}$-properness, is based on the vanishing of three different complementary covariance matrices [51]. In the case of possibly improper random variables, we need to take into account the quaternion conjugate and augmented second-order statistics [51–53]. The degree of improperness of a quaternion random vector can be measured in several ways. An interesting approach is to use an improperness measure based on the Kullback–Leibler (KL) divergence between the augmented covariance matrix and its closest proper version in the KL sense [51]. Quaternion second-order statistics have been used in several applications, from independent component analysis to canonical transform [54,55], from linear to nonlinear adaptive filtering [56–58].

Recently, augmented quaternion second-order statistics have been also exploited for the first time for the development of a deep learning model, the QVAE [26]. Deep VAEs should model long-range correlations in data [59] and quaternion layers allow to leverage internal latent relations between input features. The QVAE also leads to a significant reduction in the number of parameters, which may benefit the decoder. A full improper QVAE should be developed to completely exploit correlated multidimensional input data, such as color images. However, on one hand an improper QVAE may be too complex due to the sophisticated structure of the complete covariance matrix. On the other hand, the proper QVAEs may represent a good tradeoff between complexity and performance, even when input data are improper. To this end, an information-theoretic analyses, thanks to its relation with improperness measures, may help in understanding how good such approximation may be in terms of properness [60,61].

In order to understand and demystify the theoretical capabilities of generating data with a proper model when the input data is either proper or improper, in this paper we investigate the proper QVAE from an information-theoretic perspective. To this end, we show how the KL divergence may establish a relation between the loss function of the QVAE and the improperness measure. We exploit the KL-based measure to show the entropy loss due to the improperness of the quaternion random vector input of the QVAE, or equivalently, the mutual information among the quaternion random vector and its involutions over three pure unit quaternions [51,52]. We show how the entropy in input data plays an important role in the generation of improper signals. Furthermore, we illustrate how the improperness can be calibrated in the latent space in order to enhance its differential entropy. Simulation results on artificial signals will prove the ability of the proper QVAE to be adopted with a good approximation for the generation of improper signals. Moreover, results will show how the properness is preserved by the model, even if it is composed of several nonlinearities.

The paper is organized as follows. In Section 2, we briefly introduce the main properties of quaternion algebra that are exploited in quaternion-valued processing. In Section 3, we analyze the quaternion properness and the augmented second-order statistics of quaternion-valued signals, which are then exploited to define the quaternion variational autoencoders. In particular, the proper QVAE is described in Section 4, while in Section 5 we define the relation between the loss function of the QVAE and the properness measures of the involved signals by using the quaternion-valued entropy. Experimental simulations are shown in Section 6, while final conclusions are derived in Section 7.

2. Fundamental Properties of Quaternion Algebra

Quaternions are hypercomplex numbers consisting of four components, a scalar part and three imaginary ones, and elements of $\mathbb{R}^4$: The quaternion domain $\mathbb{H}$ is an extension of the complex domain $\mathbb{C}$. A quaternion is identified as

$$q = q_0 + q_1\hat{\imath} + q_2\hat{\jmath} + q_3\hat{\kappa} = q_0 + \mathbf{q} \tag{1}$$

in which q_0, q_1, q_2, q_3 are real coefficients and $\hat{\imath}$, $\hat{\jmath}$, $\hat{\kappa}$ imaginary units. A quaternion without the scalar part q_0 is named *pure quaternion* and represents a vector in $\mathbb{R}^3$ since the imaginary units play the role of orthonormal basis, being $\hat{\imath} = (1,0,0)$, $\hat{\jmath} = (0,1,0)$, $\hat{\kappa} = (0,0,1)$. The quaternions' peculiarity is the relation among imaginary components which comply with the following properties:

$$\hat{\imath}^2 = \hat{\jmath}^2 = \hat{\kappa}^2 = -1$$
$$\hat{\imath}\hat{\jmath} = \hat{\imath} \times \hat{\jmath} = \hat{\kappa}; \ \hat{\jmath}\hat{\kappa} = \hat{\jmath} \times \hat{\kappa} = \hat{\imath}; \ \hat{\kappa}\hat{\imath} = \hat{\kappa} \times \hat{\imath} = \hat{\jmath} \tag{2}$$
$$\hat{\imath}\hat{\jmath} = -\hat{\jmath}\hat{\imath}; \ \hat{\jmath}\hat{\kappa} = -\hat{\kappa}\hat{\jmath}; \ \hat{\kappa}\hat{\imath} = -\hat{\imath}\hat{\kappa}.$$

The fundamental properties of quaternion operations are then described below.
Addition and subtraction of two quaternions. Addition and subtraction among quaternions q and p are performed component-wise:

$$q \pm p = q_0 \pm p_0 + (q_1 \pm p_1)\hat{\imath} + (q_2 \pm p_2)\hat{\jmath} + (q_3 \pm p_3)\hat{\kappa}. \tag{3}$$

Scalar product of two quaternions. As for addition and subtraction, the scalar product is also an element-wise operation:

$$q \cdot p = q_0 p_0 + q_1 p_1 + q_2 p_2 + q_3 p_3. \tag{4}$$

Vector product of two quaternions. Due to the relations among imaginary units in (2) and the non commutative properties of vector product in the hypercomplex domain, a new kind of vector multiplication has to be introduced. The Hamilton product, denoted by $\otimes$, is at the core of quaternion neural networks and it is described by the following equation:

$$\begin{aligned}
q \otimes p &= (q_0 + q_1\hat{\imath} + q_2\hat{\jmath} + q_3\hat{\kappa})(p_0 + p_1\hat{\imath} + p_2\hat{\jmath} + p_3\hat{\kappa}) \\
&= (q_0 p_0 - q_1 p_1 - q_2 p_2 - q_3 p_3) \\
&+ (q_0 p_1 + q_1 p_0 + q_2 p_3 - q_3 p_3)\hat{\imath} \\
&+ (q_0 p_2 - q_1 p_3 + q_2 p_0 + q_3 p_2)\hat{\jmath} \\
&+ (q_0 p_3 + q_1 p_2 - q_2 p_1 + q_3 p_0)\hat{\kappa}.
\end{aligned} \tag{5}$$

The Hamilton product can also be written in concise form:

$$q \otimes p = q_0 p_0 - \mathbf{q} \cdot \mathbf{p} + q_0 \mathbf{p} + p_0 \mathbf{q} + \mathbf{q} \times \mathbf{p},$$

from which it is straightforward obtaining the Hamilton product for pure quaternions, by fixing $q_0 = p_0 = 0$:

$$\mathbf{q} \otimes \mathbf{p} = -\mathbf{q} \cdot \mathbf{p} + \mathbf{q} \times \mathbf{p}.$$

Conjugate of a quaternion. The conjugate of a quaternion number is obtained by subtracting the imaginary units instead of adding them to the scalar part:

$$q^* = q_0 - q_1 \hat{\imath} - q_2 \hat{\jmath} - q_3 \hat{k} = q_0 - \mathbf{q}. \tag{6}$$

Norm of a quaternion. The norm of a quaternion is simply the euclidean norm in $\mathbb{R}^4$, thus $|q| = \sqrt{qq^*} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$.

Quaternion polar form. Quaternions admit an Euler representation, so considering a quaternion q, an angle $\theta \in \mathbb{R}$ and a pure unit quaternion $\nu = \mathbf{q}/\|\mathbf{q}\|$, the polar form is:

$$q = |q|(\cos(\theta) + \nu \sin(\theta)) = |q|e^{\nu\theta}$$

in this setting, $\cos(\theta) = q_0/\|q\|$ and $\sin(\theta) = \|\mathbf{q}\|/\|q\|$.

Quaternion involutions. A quaternion involution is generally defined as a mapping which is its own inverse or self-inverse mapping. Quaternions have infinite involutions that can be generalized by [62]:

$$q^\nu = -\nu q \nu \tag{7}$$

in which q is the quaternion to be involved and ν is the axis of the involution. Interestingly, the scalar part of the quaternion is invariant to any kind of involution while the imaginary parts are reflected across the axis of the involution. Consequently, the conjugate of a quaternion is an involution since the scalar part remains unchanged while the imaginary units are reflected through the axis. Together with the conjugate, there are three crucial involutions we have to consider, which are perpendicular involutions:

$$\begin{aligned} q^{\hat{\imath}} &= -\hat{\imath}q\hat{\imath} = q_0 + q_1\hat{\imath} - q_2\hat{\jmath} - q_3\hat{k} \\ q^{\hat{\jmath}} &= -\hat{\jmath}q\hat{\jmath} = q_0 - q_1\hat{\imath} + q_2\hat{\jmath} - q_3\hat{k} \\ q^{\hat{k}} &= -\hat{k}q\hat{k} = q_0 - q_1\hat{\imath} - q_2\hat{\jmath} + q_3\hat{k} \end{aligned} \tag{8}$$

In the perpendicular involutions, the quaternion is involved across its imaginary units.

3. Augmented Second-Order Statistics and Quaternion Properness

3.1. Statistics for Quaternion Random Signals

In order to understand the behavior of a signal, it is often crucial to study its statistics. A generic Gaussian signal can be completely described by the mean and the covariance matrix. For quaternion random variables, the mean can be easily derived by (3), resulting in another quaternion:

$$\mu_q = \bar{q}_0 + \bar{q}_1\hat{\imath} + \bar{q}_2\hat{\jmath} + \bar{q}_3\hat{k} \tag{9}$$

in which $\bar{q}_\delta$, $\delta = \{0, 1, 2, 3\}$ is the average of each quaternion component. Conversely, deriving second-order statistics for quaternion signals requires more detailed computations due to the interactions among components. As previously demonstrated for complex numbers [49], the covariance matrix is not able to describe the second-order information and the complementary covariance matrix needs to be considered too. Consequently, taking the standard covariance matrix of a quaternion signal into account is not sufficient to cover the complete second-order statistics; thus, the information is augmented through three complementary covariance matrices [52]. In order to build this complete statistical description, we need to introduce the information brought by the perpendicular involutions $\mathbf{q}^{\hat{\imath}}$, $\mathbf{q}^{\hat{\jmath}}$, $\mathbf{q}^{\hat{k}}$ introduced in (8). The matrices to be considered are then:

$$\mathbf{C_{qq}} = \mathrm{E}\{\mathbf{q}\mathbf{q}^{\mathrm{H}}\}, \quad \mathbf{C_{qq}}^{\hat{\imath}} = \mathrm{E}\{\mathbf{q}\mathbf{q}^{\hat{\imath}\mathrm{H}}\}, \quad \mathbf{C_{qq}}^{\hat{\jmath}} = \mathrm{E}\{\mathbf{q}\mathbf{q}^{\hat{\jmath}\mathrm{H}}\}, \quad \mathbf{C_{qq}}^{\hat{k}} = \mathrm{E}\{\mathbf{q}\mathbf{q}^{\hat{k}\mathrm{H}}\} \tag{10}$$

which are, respectively, the standard covariance matrix and the three complementary matrices. Note that $(\cdot)^{\mathrm{H}}$ is the conjugate transpose operator. These matrices are still quaternions, so they comprise a real part and three imaginary units. As an example, we analyze the first matrix $\mathbf{C_{qq}}$. It is composed by the covariance matrices of the quaternion components $\mathbf{q}_0$, $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ as follows:

$$
\begin{aligned}
\mathbf{C_{qq}} = &\left(\mathbf{C_{q_0}} + \mathbf{C_{q_1}} + \mathbf{C_{q_2}} + \mathbf{C_{q_3}}\right) \\
&+ \left(\mathbf{C_{q_1 q_0}} - \mathbf{C_{q_0 q_1}} + \mathbf{C_{q_3 q_2}} - \mathbf{C_{q_2 q_3}}\right)\hat{\imath} \\
&+ \left(\mathbf{C_{q_2 q_0}} - \mathbf{C_{q_0 q_2}} + \mathbf{C_{q_1 q_3}} - \mathbf{C_{q_3 q_1}}\right)\hat{\jmath} \\
&+ \left(\mathbf{C_{q_3 q_0}} - \mathbf{C_{q_0 q_3}} + \mathbf{C_{q_2 q_1}} - \mathbf{C_{q_1 q_2}}\right)\hat{k}.
\end{aligned}
\tag{11}
$$

Similarly, also the matrices $\mathbf{C_{qq^{\hat{\imath}}}}$, $\mathbf{C_{qq^{\hat{\jmath}}}}$, $\mathbf{C_{qq^{\hat{k}}}}$ can be determined and written in the above quaternion form.

The composition and the combination of the standard covariance matrix and of the complementary covariance matrices give rise to the augmented covariance matrix $\tilde{\mathbf{C}}_{\mathbf{qq}}$ which recovers the complete second-order information of the augmented quaternion vector $\tilde{\mathbf{q}} = \left[\mathbf{q}, \mathbf{q}^{\hat{\imath}}, \mathbf{q}^{\hat{\jmath}}, \mathbf{q}^{\hat{k}}\right]$ [52]. The matrices in (10) have interesting properties: each component of the matrix is symmetric but the component corresponding to the involved imaginary unit is, instead, skew-symmetric. These characteristics are included in the Hermitian property, according to which $\mathbf{C_{qq^{\hat{\imath}}}}$ is $\hat{\imath}$-Hermitian, $\mathbf{C_{qq^{\hat{\jmath}}}}$ is $\hat{\jmath}$-Hermitian and $\mathbf{C_{qq^{\hat{k}}}}$ is $\hat{k}$-Hermitian. The derived augmented covariance matrix follows.

$$
\tilde{\mathbf{C}}_{\mathbf{qq}} = \mathrm{E}\left\{\tilde{\mathbf{q}}\tilde{\mathbf{q}}^{\mathrm{H}}\right\} = \begin{bmatrix}
\mathbf{C_{qq}} & \mathbf{C_{qq^{\hat{\imath}}}} & \mathbf{C_{qq^{\hat{\jmath}}}} & \mathbf{C_{qq^{\hat{k}}}} \\
\mathbf{C_{qq^{\hat{\imath}}}^{H}} & \mathbf{C_{q^{\hat{\imath}}q^{\hat{\imath}}}} & \mathbf{C_{q^{\hat{\imath}}q^{\hat{\jmath}}}} & \mathbf{C_{q^{\hat{\imath}}q^{\hat{k}}}} \\
\mathbf{C_{qq^{\hat{\jmath}}}^{H}} & \mathbf{C_{q^{\hat{\jmath}}q^{\hat{\imath}}}} & \mathbf{C_{q^{\hat{\jmath}}q^{\hat{\jmath}}}} & \mathbf{C_{q^{\hat{\jmath}}q^{\hat{k}}}} \\
\mathbf{C_{qq^{\hat{k}}}^{H}} & \mathbf{C_{q^{\hat{k}}q^{\hat{\imath}}}} & \mathbf{C_{q^{\hat{k}}q^{\hat{\jmath}}}} & \mathbf{C_{q^{\hat{k}}q^{\hat{k}}}}
\end{bmatrix}.
\tag{12}
$$

By means of the mean and augmented covariance matrix and following [52], it is now possible to define a generic quaternion Gaussian distribution $p(\tilde{\mathbf{q}}) = p\left(\mathbf{q}, \mathbf{q}^{\hat{\imath}}, \mathbf{q}^{\hat{\jmath}}, \mathbf{q}^{\hat{k}}\right)$, which is Gaussian if its components are jointly normal:

$$
p(\tilde{\mathbf{q}}) = \frac{1}{(\pi/2)^{2N}\det\left(\tilde{\mathbf{C}}_{\mathbf{qq}}\right)^{1/2}}\exp\left\{-\frac{1}{2}\left(\tilde{\mathbf{q}} - \tilde{\boldsymbol{\mu}}_q\right)^{\mathrm{H}}\tilde{\mathbf{C}}_{\mathbf{qq}}^{-1}\left(\tilde{\mathbf{q}} - \tilde{\boldsymbol{\mu}}_q\right)\right\},
\tag{13}
$$

where the mean $\tilde{\boldsymbol{\mu}}_q$ is the mean of the augmented quaternion $\tilde{\mathbf{q}}$ and N the number of samples.

3.2. $\mathbb{H}$-Properness

In the previous subsection, we describe the complete second-order information of a quaternion signal by means of the augmented covariance matrix which is comprised of the covariance matrices of the quaternion with its perpendicular involutions. Due to the four degrees of freedom in quaternions, the $\mathbb{H}$ domain includes manifold properness definitions that are employed to characterize the relation of $\mathbf{q}$ with $\mathbf{q}^{\hat{\imath}}$, $\mathbf{q}^{\hat{\jmath}}$, $\mathbf{q}^{\hat{k}}$ in some particular cases.

The lighter condition is described in the $\mathbb{R}$-properness, which holds for a signal that is uncorrelated with one of its perpendicular involutions, leading to the cancellation of either one of the complementary covariance submatrices $\mathbf{C_{qq^{\hat{\imath}}}}$, $\mathbf{C_{qq^{\hat{\jmath}}}}$, $\mathbf{C_{qq^{\hat{k}}}}$. A slightly harder constraint is brought by the $\mathbb{C}$-properness which claims the nullification of two of the three complementary covariance submatrices. A composition of the two already defined properties gives rise to the strongest properness condition, the $\mathbb{H}$-properness. The latter holds when each complementary covariance matrix vanishes, that is $\mathbf{C_{qq^{\hat{\imath}}}} = \mathbf{C_{qq^{\hat{\jmath}}}} = \mathbf{C_{qq^{\hat{k}}}} = \mathbf{0}$. However, according to [52], an $\mathbb{H}$-proper variable has also several other features that are reported in Table 1. An $\mathbb{H}$-proper signal is then uncorrelated with its perpendicular involutions and, according to the second property in Table 1, it also has uncorrelated components $\mathbf{q}_0$, $\mathbf{q}_1$, $\mathbf{q}_2$ $\mathbf{q}_3$ which have equal variance σ^2. From these

assumptions, we can derive the corresponding augmented covariance matrix for $\mathbb{H}$-proper signals which will be symmetric and positive definite:

$$\widehat{\mathbf{C}}_{\mathbf{q}\mathbf{q}} = \mathrm{E}\left\{\tilde{\mathbf{q}}\tilde{\mathbf{q}}^{\mathrm{H}}\right\} = \begin{bmatrix} \mathbf{C}_{\mathbf{q}\mathbf{q}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\imath}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\jmath}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{q}^{\hat{k}}\mathbf{q}^{\hat{k}}} \end{bmatrix} = 4\sigma^2\mathbf{I}. \tag{14}$$

Table 1. Fundamental properties of $\mathbb{H}$-proper random variables.

$\mathrm{E}\{\mathbf{q}_\delta^2\} = \mathrm{E}\{\mathbf{q}_\epsilon^2\} = \sigma^2$	$\forall \delta, \epsilon = \{0,1,2,3\}$
$\mathrm{E}\{\mathbf{q}_\delta \mathbf{q}_\epsilon\} = \mathbf{0}$	$\forall \delta, \epsilon = \{0,1,2,3\}, \delta \neq \epsilon$
$\mathrm{E}\{\mathbf{q}\mathbf{q}\} = -2\sigma^2$	$\forall \delta = \{0,1,2,3\}$
$\mathrm{E}\{|\mathbf{q}|^2\} = 4\sigma^2$	$\forall \delta = \{0,1,2,3\}$

The structure of the matrix $\widehat{\mathbf{C}}_{\mathbf{q}\mathbf{q}}$ is invariant to any linear transformation $\nu = \mathbf{T}^{\mathrm{H}}\mathbf{q}$ where $\mathbf{T}$ is a quaternion weight matrix [54], thus ensuring that the quaternion properness is preserved by linear layers in neural networks.

As for improper signals in Section 3.1, a probability density function can be derived for $\mathbb{H}$-proper signals too. In this framework, the distribution is described by the covariance matrix in (14) which is equal to $4\sigma^2\mathbf{I}$ and by the mean of the quaternion component in (9):

$$p(\tilde{\mathbf{q}}) = \frac{1}{(2\pi\sigma^2)^{2N}} \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{q} - \boldsymbol{\mu}_{\mathbf{q}}\right)^{\mathrm{H}}\left(\mathbf{q} - \boldsymbol{\mu}_{\mathbf{q}}\right)\right\} \tag{15}$$

In order to evaluate the properness of an augmented zero-mean quaternion random vector, a simple and fast coefficient was proposed in [32]. It indicates the correlation of the signal with its perpendicular involutions through the complementary covariance matrices. Note that the coefficient is bounded in $[0,1]$ and it is 0 in the case of $\mathbb{H}$-proper signals, while it is equal to 1 for improper ones. It is defined as

$$I_\alpha = \left|\frac{\mathbf{C}_{\mathbf{q}\mathbf{q}^\alpha}}{\mathbf{C}_{\mathbf{q}\mathbf{q}}}\right|, \qquad \alpha = \hat{\imath}, \hat{\jmath}, \hat{k}. \tag{16}$$

In our experiments, since we are dealing with $\mathbb{H}$-proper signals, we can just report the average of the three coefficients, simply denoted as I. Conversely, if we evaluate $\mathbb{R}$ or $\mathbb{C}$ proper signals, we should consider each coefficient independently.

4. The $\mathbb{H}$-Proper Variational Autoencoder

Originally introduced in [26] (the implementation of the QVAE is available online at: https://github.com/eleGAN23/QVAE, accessed on 1 July 2021), the quaternion variational autoencoder (QVAE) is a generative model which learns and controls the latent distribution of an n-dimensional $\mathbb{H}$-proper quaternion input. Once the distribution is learnt, then the model is able to generate new samples from it thanks to the probabilistic relation between the two quaternion spaces of the input and of the latent distribution.

Similar to the variational autoencoder (VAE) for real-valued distributions in [13], the QVAE grasps the mapping of the original quaternion input $\mathbf{x} \in \mathbb{H}$ into the latent space described by a prior distribution $p_{\boldsymbol{\theta}}(\epsilon)$, with $\epsilon \in \mathbb{H}$. Then, it learns the conditional distribution $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ of the input with respect to the latent vector $\mathbf{z} \in \mathbb{H}$, that is the prior distribution adjusted with the statistics learnt by the encoder. However, due to the intractability of the mapping distribution $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$, the QVAE introduces an approximation $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ to encode the input into the latent vector, as in its real-valued counterpart [13]. The parameters of this recognition model (or inference model) are then optimized to match the true intractable

conditional distribution as $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \approx p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$. By employing this approximation, it is possible to express the marginal likelihood

$$
\begin{aligned}
\log(p_{\boldsymbol{\theta}}(\mathbf{x})) &= \lambda \mathcal{D}_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \| p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})) + \mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x}) \\
&= -\lambda \mathcal{D}_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \| p_{\boldsymbol{\theta}}(\epsilon)) + \mathrm{E}_q\{\log(p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}))\}
\end{aligned}
$$

where $\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x})$ is the so called *evidence lower bound* (ELBO) [13] with respect to the parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, which is a lower bound on the log-likelihood of the data distribution [63]. $\mathcal{D}_{\mathrm{KL}}(\cdot)$ is the KL divergence between the $\mathbb{H}$-proper centered isotropic Gaussian variable $p_{\boldsymbol{\theta}}(\epsilon) \sim \mathcal{N}(\mathbf{0}, \widehat{\mathbf{C}}_{\epsilon\epsilon})$ and the $\mathbb{H}$-proper Gaussian $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, \widehat{\mathbf{C}}_{\mathbf{zz}})$, with $\widehat{\mathbf{C}}_{\mathbf{zz}}$ taking the form of (14) and $\widehat{\mathbf{C}}_{\epsilon\epsilon} = 4\mathbf{I}\sigma^2$. The KL divergence in the quaternion domain takes the following form:

$$
\begin{aligned}
\mathcal{D}_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \| p_{\boldsymbol{\theta}}(\epsilon)) = \frac{1}{2}\Bigg(&\mathrm{Tr}\left\{ \widehat{\mathbf{C}}_{\epsilon\epsilon}^{-\frac{1}{2}} \widehat{\mathbf{C}}_{\mathbf{zz}} \widehat{\mathbf{C}}_{\epsilon\epsilon}^{-\frac{1}{2}} \right\} \\
&+ \boldsymbol{\mu}_{\mathbf{z}}^{\mathrm{H}} \widehat{\mathbf{C}}_{\epsilon\epsilon}^{-1} \boldsymbol{\mu}_{\mathbf{z}} - N + \log\left(\frac{\det\left(\widehat{\mathbf{C}}_{\epsilon\epsilon}\right)}{\det\left(\widehat{\mathbf{C}}_{\mathbf{zz}}\right)} \right) \Bigg).
\end{aligned} \tag{17}
$$

The optimization of the QVAE cost function (17) goes through the minimization of the KL divergence, through which the QVAE forces the recognition model to be as close as possible to the prior distribution by operating on the statistics $\boldsymbol{\mu}_{\mathbf{z}}$ and $\widehat{\mathbf{C}}_{\mathbf{zz}}$.

5. Kullback–Leibler Divergence as Entropy Loss Due to Improperness

The KL divergence in (17) can be indeed seen as a measure of the improperness of the recognition model. As defined in [51], the measure is described by:

$$
\mathcal{P} = \min_{\widehat{\mathbf{C}}_{\epsilon\epsilon} \in \mathcal{C}} \mathcal{D}_{\mathrm{KL}}\left(\widehat{\mathbf{C}}_{\mathbf{zz}} \| \widehat{\mathbf{C}}_{\epsilon\epsilon} \right) \tag{18}
$$

where $\mathcal{C}$ is the set of all the augmented covariance matrices respecting one of the quaternion propernesses introduced in Section 3.2. This measure estimates the distance of the recognition model $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ with an augmented covariance matrix $\widehat{\mathbf{C}}_{\mathbf{zz}}$ from the $\mathbb{H}$-proper isotropic Gaussian variable $p_{\boldsymbol{\theta}}(\epsilon)$ described by the augmented covariance matrix $\widehat{\mathbf{C}}_{\epsilon\epsilon}$, by assuming $\boldsymbol{\mu}_{\mathbf{z}} = 0$.

Since we are considering distributions with zero mean, the first term in the second line of the KL Equation (17) vanishes. Furthermore, Equation (18) minimizes the divergence over the set $\mathcal{C}$; thus, in the $\mathbb{H}$-proper case, the matrix which minimizes (18) is the diagonal covariance matrix defined in (14). Employing these considerations and with some computations, we can reduce the improperness measure to:

$$
\mathcal{P} = -\frac{1}{2} \log\left(\det\left\{ \widehat{\mathbf{C}}_{\epsilon\epsilon}^{-\frac{1}{2}} \widehat{\mathbf{C}}_{\mathbf{zz}} \widehat{\mathbf{C}}_{\epsilon\epsilon}^{-\frac{1}{2}} \right\} \right). \tag{19}
$$

If the signal for which we want to measure the improperness, i.e., the recognition model $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$, is $\mathbb{H}$-proper, then the measure will be close to 0 since $\widehat{\mathbf{C}}_{\mathbf{zz}}$ will be close to $\widehat{\mathbf{C}}_{\epsilon\epsilon}$ in the KL sense. Conversely, if the signal is improper, $\mathcal{P}$ will be far from 0.

By defining the improperness measure through the KL, it is possible to leverage the relation among the divergence and the entropy and exploring the mutual information of the signal. Therefore, we can use the definition of the differential entropy for a generic quaternion vector [51,52] to compute the entropy of the recognition model $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$:

$$\mathrm{H}\big(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})\big) = 2N\log\Big(\frac{\pi e}{2}\Big) + \frac{1}{2}\log\Big(\det\big(\widehat{\mathbf{C}}_{\mathbf{zz}}\big)\Big)$$
$$= \log\Big(\Big(\frac{\pi e}{2}\Big)^{2N}\det\big(\widehat{\mathbf{C}}_{\mathbf{zz}}\big)^{\frac{1}{2}}\Big). \tag{20}$$

The entropy expression can be simplified for isotropic $\mathbb{H}$-proper variables such as the prior distribution $p_{\boldsymbol{\theta}}(\epsilon)$, by replacing the value of the determinant from (14), which is equal to $(4\sigma^2)^{4N}$.

$$\mathrm{H}(p_{\boldsymbol{\theta}}(\epsilon)) = \log\Big(\Big(\frac{\pi e}{2}\Big)^{2N}\Big(\big(4\sigma^2\big)^{4N}\Big)^{\frac{1}{2}}\Big)$$
$$= 2N\Big(\log\big(2\pi e\sigma^2\big)\Big). \tag{21}$$

In [52], the authors proved that the differential entropy is maximum for $\mathbb{H}$-proper signals. Thus, (21) is an upper bound for the general entropy definition (20), hence: $\mathrm{H}\big(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})\big) \leq \mathrm{H}(p_{\boldsymbol{\theta}}(\epsilon))$.

Then, according to [51], the improperness measure $\mathcal{P}$ can be rewritten through the difference between the entropy of the generic augmented quaternion and the entropy of the isotropic variable:

$$\mathcal{P}_{\mathrm{H}} = \mathrm{H}(p_{\boldsymbol{\theta}}(\epsilon)) - \mathrm{H}\big(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})\big). \tag{22}$$

This formulation elegantly evaluates the loss of entropy that is due to the improperness of the recognition model. Therefore, we can measure the properness of the recognition model and compute the QVAE entropy loss due to the measure of properness (or improperness) of $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$. Thus, when optimizing the QVAE cost function during training, together with the ELBO minimization, through the KL divergence we are jointly minimize also the difference between the Gaussian distributions $p_{\boldsymbol{\phi}}(\epsilon)$ and $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and the entropy loss of the latter.

Since the KL under the information theoretic perspective allows the QVAE to deal also with improper signals, although the QVAE is originally set up for $\mathbb{H}$-proper distributions, by means of an approximation it can be employed also for improper random variables. Let us consider an incoming improper signal $\mathbf{x} \in \mathbb{H}$, so both the recognition model $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and the generative model $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ are represented by improper distributions too. We leave instead unchanged the prior distribution $p_{\boldsymbol{\theta}}(\epsilon)$, which is the centered isotropic $\mathbb{H}$-proper distribution. On the training stage, the KL divergence of the QVAE will move the improper distribution towards the $\mathbb{H}$-proper prior distribution. While the distribution learning is well performed by the model, the KL takes care of increasing the entropy by slightly reducing the improperness. However, the overall original (improper) structure of the signal is preserved by the QVAE which just aims at enhancing the entropy of the recognition model, while leaving free the generative model.

6. Experimental Results

In this section, we report the sets of experiments we performed for testing the reliability of the improperness measures and the performances of the $\mathbb{H}$-proper variational autoencoder in learning proper as well as improper distributions.

6.1. Proper and Improper Test Signals

We consider two proper signals and two improper signals and we evaluate the improperness coefficient (16), the differential entropy (20), the improperness measure (18) and the entropy loss (22). We expect that the proper signals have improperness measure and coefficients close to 0, as well as the entropy loss. On the contrary, the entropy should be greater than for the corresponding improper signals.

As input signals, we consider:

1. $\mathbb{H}$-proper signal **x** with independent Gaussian components defined as:

$$\text{P1:} \quad \mathbf{x}_\delta \sim \mathcal{N}(0.03, 1)$$

with $\delta = \{0, 1, 2, 3\}$.

2. $\mathbb{H}$-proper filtered signal **x** from a colored Gaussian noise $\eta[n]$ as [31]:

$$\text{P2:} \quad \mathbf{x}[n] = b\mathbf{x}[n-1] + \frac{\sqrt{1-b^2}}{\sqrt{4}}\eta[n]$$

with $b = 0.5$.

3. Improper signal **x** from a quaternion **q** with components:

$$\mathbf{q}_0 = 0, \quad \mathbf{q}_1 \sim \mathcal{N}(0, 1), \quad \mathbf{q}_2 \sim \mathcal{N}(0, 1), \quad \mathbf{q}_3 \sim \mathcal{N}(0, 1)$$

and composed as:

$$\text{I1:} \quad \mathbf{x} = \mathbf{z}e^{\mathbf{q}}, \quad \text{with } \mathbf{z} \sim \mathcal{N}(0.03, 1)$$

4. Improper filtered signal **x** from a Gaussian noise $\eta[n]$ as:

$$\text{I2:} \quad \mathbf{x}_\delta[n] = \sqrt{1 - b_\delta^2} + (1 - b_\delta)\eta[n]$$

with $b = [0.75, 0.78, 0.82, 0.85]$ and $\delta = \{0, 1, 2, 3\}$.

Mild improper signals (i.e., with I_α different from 0 and to 1) can be generated by varying the coefficient b in the improper signal I2. As an example, by setting $b = [0.01, 0.2, 0.75, 0.99]$, it is possible to obtain a mild-improper signal with $0.4 < I_\alpha < 0.7$. However, we want to stress the $\mathbb{H}$-proper QVAE with completely improper cases that are the extreme and most difficult cases. For this reason, we only consider improper signals with $I_\alpha = 1$.

We compute the measures defined in Section 5 for each of these signal by using the Quaternion Toolbox for MATLAB® (S. J. Sangwine and N. Le Bihan. Quaternion Toolbox for MATLAB®, Version 2 with support for Octonions, 2013. First public release 2005. Software library available online at: http://qtfm.sourceforge.net/, accessed on 1 July 2021) and report the results in Table 2. As we expect, the proper signals have low improperness coefficients, meaning that the complementary covariances are sparse, while the coefficients for the improper signals are reasonably equal to 1. On the other hand, the differential entropy is low and negative for the improper signals, while it is highest for the proper ones, proving that (21) is an upper bound for the generic differential entropy. Moreover, the entropy loss is almost 0 for proper signals, confirming that there is no loss of entropy due to the improperness. On the contrary, the improper signals come at the cost of a high entropy drop.

Table 2. In order to prove the measures and the assumptions presented in Section 5, we compute the average improperness coefficient I (16), the differential entropy H (20), the improperness measure $\mathcal{P}$ (19) and entropy loss $\mathcal{P}_\text{H}$ (22) for different kinds of $\mathbb{H}$-proper and improper signals. The entropy H is highest for the $\mathbb{H}$-proper signals, while the improperness measure and coefficients are close to 0. As is clear from $\mathcal{P}$ and $\mathcal{P}_\text{H}$, there is a low entropy loss due to the improperness in the first two signals, while it is high for the last two, improper, random vectors.

| Signal	I	**H**	$\mathcal{P}$	$\mathcal{P}_\text{H}$
P1. (proper)	0.0393	4.0177	0.0049	0.0049
P2. (proper)	0.0528	5.0174	0.0103	0.0103
I1. (improper)	1.0000	−51.6040	50.9298	51.0691
I2. (improper)	0.9999	−45.4454	51.3121	51.1843

6.2. Proper and Improper Signals for an $\mathbb{H}$-Proper QVAE

We consider a plain QVAE with a quaternion encoder network and a quaternion decoder network. The first model is comprised of a stack of five quaternion fully connected layers (QFC) with an increasing number of weights $[32, 64, 128, 256, 512]$, as in [26,64] and split ReLU activation function as the PixelHVAE in [65]. The statistics of the latent distribution are learnt through two additionally quaternion linear layers with number of weights equal to 4 times the latent dimension, which we fix to 25. The decoder has a mirrored structure, and thus quaternion fully connected layers are piled up as $[512, 256, 128, 64, 32]$, with an additional refiner layer at the end of the stack. We do not consider including the quaternion batch normalization [38,45] since it can introduce randomness which may affect the correct learning of the distribution statistics [24,66]. For every experiment, the prior distribution is a centered isotropic $\mathbb{H}$-proper quaternion Gaussian distribution, as described in Section 4. We employ a training signal with 1000 samples, a validation set of 500 and a test set of 1000 without mini batches and we perform 5k iterations with Adam having a learning rate equal to 0.0005. As in the original QVAE [26], the cost function is a weighted sum of a quaternion mean squared error as defined in [37], and the KL divergence of Section 5, for which we fix the scale parameter λ equal to 0.1. Figure 1 shows the architecture of the QVAE we consider in our tests.
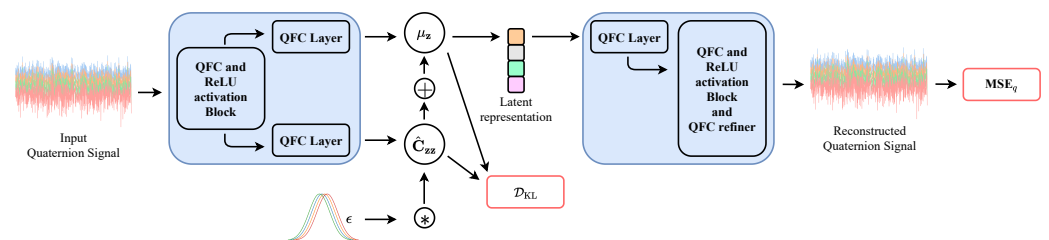

Figure 1. Plain QVAE architecture. The input signal can be either proper or improper. The encoder network is composed of quaternion fully connected layers (QFC) and it learns the statistics of the original distribution and employs them to build the latent representation, on which the KL divergence acts. The representation is then passed to the decoder network which reconstruct the input signal, evaluating the result by means of a quaternion MSE loss (MSE_q).

For each experiment, we evaluate the performance of the model in the reconstruction task and generation task as well as on the ability of preserving the properness or the improperness of the incoming signal. In order to assess the generation ability of the model, in Figure 2, we plot the generated components and the complete signal sampling from the model trained on a proper signal and on an improper signal. While the components $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ of the improper signal show different variances, the ones from the proper signal has the same variance according to the properties in Table 1. Figure 3 reports instead the original and reconstructed distributions for the proper and improper case, in the first and in the second line, respectively, with an MSE equal to 0.1161 for the proper signal and to 0.01293 for the improper one. In the figure, the first four columns represent the four components $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$, while the last one displays the distribution of the complete signal $\mathbf{x}$. The $\mathbb{H}$-QVAE is able to perfectly reconstruct the input distribution even in the improper case, grasping the different variances in each component and learning the complete signal. Thus, the QVAE is effective even if approximated for improper distributions, for which it is capable of learning the correct statistics of each component.
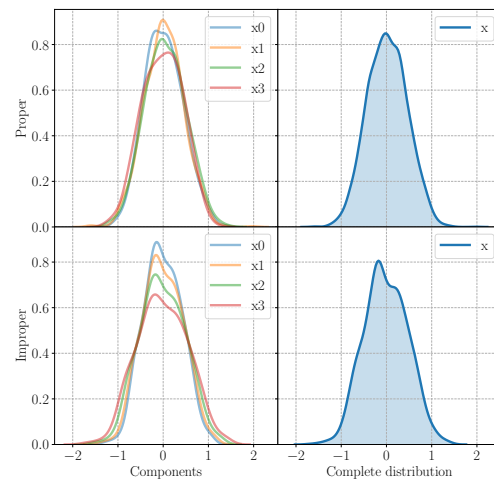
Figure 2. Generated samples from the QVAE trained on the proper signal P1 (first line), and on the improper signal I2 (second line). On the left are the generated components, on the right are the complete distribution of the signal. While $\mathbb{H}$-proper components have the same variance, the improper ones have different variance values, respecting the properties of the original signal.
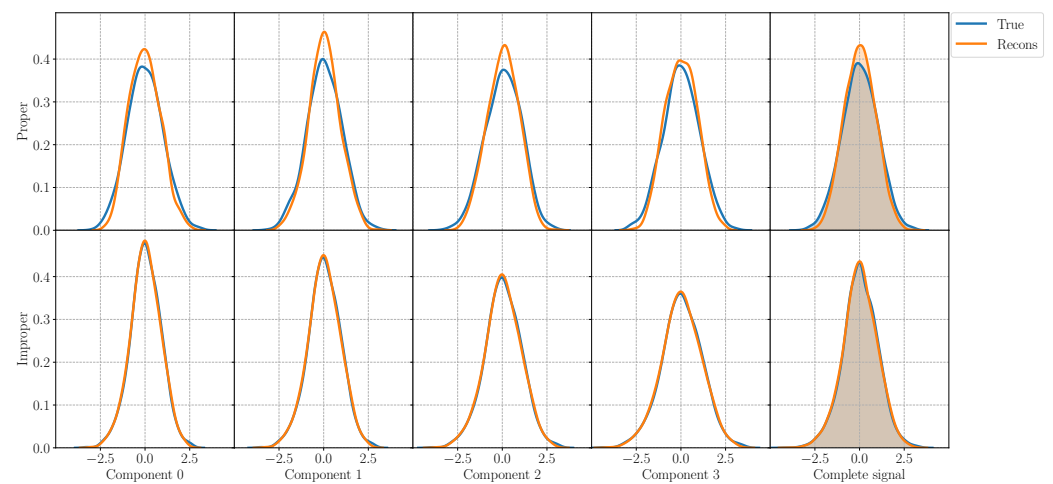
Figure 3. Original and reconstructed distributions from the QVAE trained with the proper signal P1 in the first line, and with improper signal I2 in the second line. The QVAE is able to perfectly learn the different variance values for each of the four components, thus fully reconstructing the distribution of the input signal.

Table 3 presents the results of the improperness and entropy analysis on the reconstructed and generated signals on the test set. As for the original samples in Table 2, we compute the improperness coefficient (16), the differential entropy (20), the improperness measure (18) and the entropy loss (19). It is worth noting that the QVAE trained on $\mathbb{H}$-proper signals still generates and reconstruct $\mathbb{H}$-proper signals, as measured by the improperness measures and the entropy loss. Moreover, the entropy of this kind of signal is higher than the entropy of the improper samples. While we already knew that $\mathbb{H}$-properness is preserved by linear transformations, this result asserts that it is maintained also for non-linear transformations brought by activation functions in neural networks, adding an important milestone in the study of $\mathbb{H}$-proper signals.

Another interesting result comes when the QVAE is trained on improper signals. While the improperness coefficients of both reconstructed and generated signals are close to the coefficients of their respective original signals in Table 2, indicating the improperness of these random variable, the entropy values and losses values are intriguing. Indeed, the differential entropy is increased with respect to the input signal (originally, ≈ -51 and

≈ -45), approaching values from ≈ -10 up to 0, meaning that the $\mathbb{H}$-proper structure of the QVAE works towards an increase in entropy and an improperness reduction in the input signal. Confirming this, the loss of entropy $\mathcal{P}_H$ was drastically reduced with respect to the original signal, ranging from approximately 8 and 13 while the input entropy loss is around ≈ 51.

Hence, the $\mathbb{H}$-proper QVAE can be approximated to learn improper distributions, on which it preserves the original structure of the signal while increasing the entropy of the latent space, at the cost of a limited improperness reduction. Although the slight modification on the distribution, the QVAE is still able to catch the correct statistics and to perform a good generation as well as a precise reconstruction, as Figures 2 and 3 show.

To further assess the performance of the QVAE and provide a comparison with its real-valued counterpart, we test a real-valued VAE on the same signal benchmarks. We expect that the real-valued VAE performs similar to the QVAE in the proper case, but that the quaternion model outperforms the real-valued one when dealing with improper signals. Indeed, thanks to the quaternion algebra properties, the QVAE is able to learn intra-component correlations while the VAE do not catch them. Thus, in cases with no correlation among components i.e., proper signals, the two models should have similar performances. On the contrary, in the improper case where the correlation among components is higher the QVAE gains advantages from the quaternion algebra and obtains better performances. Results and details are reported in the Appendix A. However, thanks to the quaternion algebra properties, the QVAE has just 25% of free parameters with respect to its real-valued counterpart.

Table 3. Average improperness coefficient I (16), differential entropy H (20), improperness measure $\mathcal{P}$ (19) and entropy loss $\mathcal{P}_H$ (22) for different outputs from the QVAE (Recons stands for the reconstructed signal, Sample for the generated one) with different kinds of $\mathbb{H}$-proper and improper signals. The QVAE preserves the structure of the signal, learning the properness and improperness nature of the input.

Signal	Out	I	H	$\mathcal{P}$	$\mathcal{P}_H$
P1. (proper)	Recons	0.0287	5.1519	0.0030	0.0029
	Sample	0.0537	2.3859	0.0096	0.0095
P2. (proper)	Recons	0.0640	3.4614	0.0124	0.0123
	Sample	0.1099	-0.2705	0.0348	0.0347
I1. (improper)	Recons	0.9989	0.3749	9.9700	9.9700
	Sample	0.9930	-3.6337	8.1208	8.1208
I2. (improper)	Recons	0.9998	-8.1100	13.0767	13.0767
	Sample	0.9964	-10.7147	12.1469	12.1469

7. Conclusions

In this paper, we have conducted a thorough analysis on a $\mathbb{H}$-proper quaternion-valued variational autoencoder (QVAE) under an information-theoretic perspective, investigating the entropy loss of the model due to the measure of improperness in the input signal. We have proved that the QVAE can be approximated for improper signals, for which it increases the entropy of the latent space by slightly reducing the improperness. Moreover, it is worth noting that, in each of our experiments, the properness is preserved also for nonlinear transformations brought by activation functions and not only for simple quaternion linear operations. As a conclusion, the QVAE is able to learn the correct distribution of both $\mathbb{H}$-proper and improper random variables, while the measure of improperness, as well as the entropy of the latent space, can be calibrated through the KL divergence in the cost function.

## Appendix A. Evaluation of the Real-Valued VAE on Proper and Improper Signals

In order to further assess the performance of the $\mathbb{H}$-proper QVAE, we perform a set of experiments with a real-valued VAE on the same benchmarks. In order to make a fair comparison, the hyperparameters are the same as for the QVAE, while the architecture is built with real-valued blocks. Moreover, the KL divergence is the standard divergence in [13]. However, due to the algebra in real domain, the VAE has $506,316$ parameters while the QVAE has just $127,980$, thus the proposed model in the quaternion domain has exactly the 25% of parameters with respect to its real-valued counterpart. Thus, the QvAE is able to learn the improperness of the input signal and to generate or to reconstruct precise signals while requiring $1/4$ parameters and memory. Moreover, the QVAE is able to perfectly reconstruct the improper signal, as shown in Figure 3, while the real-valued VAE makes some errors. These results can be seen in Figures A1 and A2. As we expect, the QVAE outperform the real-valued VAE when dealing with correlated signals i.e., improper cases, since the proposed method leverages the quaternion algebra properties, including the Hamilton product, to learn intra-components relations, while the real-valued VAE is not able to catch them. Thus, while for proper signals the performances between VAE and QVAE are similar, in the improper cases the QVAE has better performances thanks to the Hamilton product in its core blocks. Finally, the improperness measures and corresponding entropy loss for each of the signals considered are reported in Table A1.

Table A1. Average improperness coefficient *I* (16), differential entropy H (20), improperness measure $\mathcal{P}$ (19) and entropy loss $\mathcal{P}_\text{H}$ (22) for different output from the real-valued VAE (Recons stands for the reconstructed signal, Sample for the generated one) with different kinds of $\mathbb{H}$-proper and improper signals. As the QVAE, also the VAE, which is however 75% larger in terms of free parameters, is able to learn the improperness of the input.

Signal	Out	*I*	H	$\mathcal{P}$	$\mathcal{P}_\text{H}$
P1. (proper)	Recons	0.031	5.4547	0.0031	0.0030
	Sample	0.0697	5.4498	0.0148	0.0147
P2. (proper)	Recons	0.0396	4.8138	0.0050	0.0050
	Sample	0.9879	-21.2687	6.7205	6.7205
I1. (improper)	Recons	0.9983	-8.9158	14.0350	14.0350
	Sample	0.9995	-8.2952	16.1035	16.1035
I2. (improper)	Recons	0.9984	-9.0598	14.2591	14.2591
	Sample	0.9977	-9.2681	13.7240	13.7240
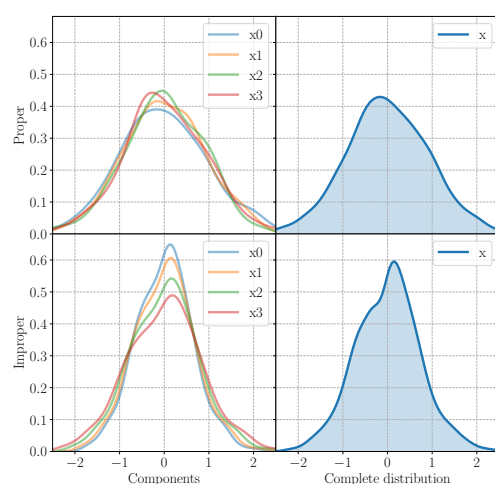
Figure A1. Generated samples from the real-valued VAE trained on the proper signal P1 and on the improper signal I2. These model generates good samples (upper figure), even if the variance is larger with respect to the QVAE.
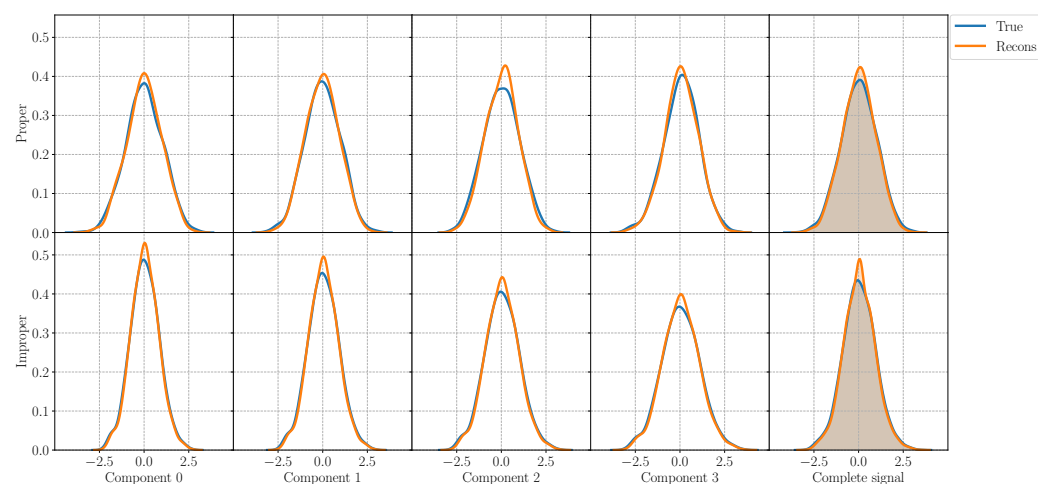
Figure A2. Reconstructed distributions from the real-valued VAE on signal P1 (proper) in the upper line and I2 (improper) in the lower line. Although the larger number of parameters, the VAE commits some errors in both the signal for the reconstruction task, while the QVAE is able instead to perfectly reconstruct original signals. The MSE values are 0.1237 for the proper signal and 0.0191 for the improper one.

References

1. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems (NIPS)*; MIT Press: Cambridge, MA, USA, 2014; Volume 2, pp. 2672–2680.
2. Rezende, D.J.; Metz, L.; Chintala, S. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In Proceedings of International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 1278–1286.
3. Bond-Taylor, S.; Leach, A.; Long, Y.; Willcocks, C.G. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *arXiv* **2021**, arXiv:2103.04922v1.
4. Ruiz-Garcia, A.; Schmidhuber, J.; Palade, V.; Took, C.C.; Mandic, D. Deep neural network representation and Generative Adversarial Learning. *Neural Netw.* **2021**, *139*, 199–200. [CrossRef]
5. Fu, Y.; Ma, J.; Guo, X. Unsupervised Exemplar-Domain Aware Image-to-Image Translation. *Entropy* **2021**, *23*, 565. [CrossRef]
6. Hou, J.; Zhang, D.; Wu, W.; Ma, J.; Zhou, H. A Generative Adversarial Network for Infrared and Visible Image Fusion Based on Semantic Segmentation. *Entropy* **2021**, *23*, 376. [CrossRef]
7. Lin, J.; Li, Y.; Yang, G. FPGAN: Face de-identification method with generative adversarial networks for social robots. *Neural Netw.* **2021**, *133*, 132–147. [CrossRef] [PubMed]

8. Dong, L.; Yang, N.; Wang, W.; Wei, F.; Liu, X.; Wang, Y.; Gao, J.; Zhou, M.; Hon, H.W. Unified Language Model Pre-training for Natural Language Understanding and Generation. *arXiv* **2019**, arXiv:1905.03197.

9. Grassucci, E.; Scardapane, S.; Comminiello, D.; Uncini, A. Flexible Generative Adversarial Networks with Non-Parametric Activation Functions. In *Progress in Artificial Intelligence and Neural Systems. Smart Innovation, Systems and Technologies*; Springer: Singapore, 2021; Volume 184.

10. Sanchez, I.; Vilaplana, V. Brain MRI Super-Resolution Using 3D Generative Adversarial Networks. *arXiv* **2018**, arXiv:1812.11440v1.

11. Elazab, A.; Wang, C.; Gardezi, S.J.S.; Bai, H.; Hu, Q.; Wang, T.; Chang, C.; Lei, B. GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images. *Neural Netw.* **2020**, *132*, 321–332. [CrossRef] [PubMed]

12. Kim, J.; Yoo, J.; Lee, J.; Hong, S. SetVAE: Learning Hierarchical Composition for Generative Modeling of Set-Structured Data. *arXiv* **2021**, arXiv:2103.15619v1.

13. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2014**, arXiv:1312.6114v10.

14. Razavi, A.; van den Oord, A.; Vinyals, O. Generating Diverse High-Fidelity Images with VQ-VAE-2. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 14866–14876.

15. Sadeghi, M.; Alameda-Pineda, X. Robust Unsupervised Audio-Visual Speech Enhancement Using a Mixture of Variational Autoencoders. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 7534–7538.

16. Wu, S.L.; Yang, Y.H. MuseMorphose: Full-Song and Fine-Grained Music Style Transfer with Just One Transformer VAE. *arXiv* **2021**, arXiv:2105.04090v1.

17. Chadebec, C.; Thibeau-Sutre, E.; Burgos, N.; Allassonnière, S. Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder. *arXiv* **2021**, arXiv:2105.00026v1.

18. Kosiorek, A.R.; Strathmann, H.; Zoran, D.; Moreno, P.; Schneider, R.; Mokrá, S.; Rezende, D.J. NeRF-VAE: A Geometry Aware 3D Scene Generative Model. *arXiv* **2021**, arXiv:2104.00587v1.

19. Stewart, K.; Danielescu, A.; Supic, L.; Shea, T.; Neftci, E. Gesture Similarity Analysis on Event Data Using a Hybrid Guided Variational Auto Encoder. *arXiv* **2021**, arXiv:2104.00165v1.

20. Qian, D.; Cheung, W.K. Enhancing Variational Autoencoders with Mutual Information Neural Estimation for Text Generation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 4047–4057.

21. Xie, Z.; Liu, C.; Zhang, Y.; Lu, H.; Wang, D.; Ding, Y. Adversarial and Contrastive Variational Autoencoder for Sequential Recommendation. *arXiv* **2021**, arXiv:2103.10693v1.

22. Gallego, V.; Ríos Insua, D. Variationally Inferred Sampling through a Refined Bound. *Entropy* **2021**, *23*, 123. [CrossRef]

23. Sadeghi, H.; Andriyash, E.; Vinci, W.; Buffoni, L.; Amin, M.H. PixelVAE++: Improved PixelVAE with Discrete Prior. *arXiv* **2019**, arXiv:1908.09948v1.

24. Vahdat, A.; Kautz, J. NVAE: A Deep Hierarchical Variational Autoencoder. *arXiv* **2020**, arXiv:2007.03898v1.

25. Zhao, H.; Li, T.; Xiao, Y.; Wang, Y. Improving Multi-Agent Generative Adversarial Nets with Variational Latent Representation. *Entropy* **2020**, *22*, 1055. [CrossRef] [PubMed]

26. Grassucci, E.; Comminiello, D.; Uncini, A. A Quaternion-Valued Variational Autoencoder. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021.

27. Bülow, T.; Sommer, G. Hypercomplex Signal—A Novel Extension of the Analystic Signal to the Multidimensional Case. *IEEE Trans. Signal Process.* **2001**, *49*, 2844–2852. [CrossRef]

28. Took, C.C.; Mandic, D.P. The Quaternion LMS Algorithm for Adaptive Filtering of Hypercomplex Processes. *IEEE Trans. Signal Process.* **2009**, *57*, 1316–1327. [CrossRef]

29. Paul, T.K.; Ogunfunmi, T. A Kernel Adaptive Algorithm for Quaternion-Valued Inputs. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 2422–2439. [CrossRef] [PubMed]

30. Ortolani, F.; Comminiello, D.; Scarpiniti, M.; Uncini, A. Frequency Domain Quaternion Adaptive Filters: Algorithms and Convergence Performance. *Signal Process.* **2017**, *136*, 69–80. [CrossRef]

31. Ortolani, F.; Comminiello, D.; Scarpiniti, M.; Uncini, A. Advances in Hypercomplex Adaptive Filtering for 3D Audio Processing. In Proceedings of the IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 29 May–2 June 2017; pp. 1125–1130.

32. Xiang, M.; Kanna, S.; Mandic, D.P. Performance Analysis of Quaternion-Valued Adaptive Filters in Nonstationary Environments. *IEEE Trans. Signal Process.* **2018**, *66*, 1566–1579. [CrossRef]

33. Ogunfunmi, T.; Safarian, C. The Quaternion Stochastic Information Gradient Algorithm for Nonlinear Adaptive Systems. *IEEE Trans. Signal Process.* **2019**, *67*, 5909–5921. [CrossRef]

34. Comminiello, D.; Scarpiniti, M.; Parisi, R.; Uncini, A. Frequency-Domain Adaptive Filtering: From Real to Hypercomplex Signal Processing. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 7745–7749.

35. Gaudet, C.; Maida, A. Deep Quaternion Networks. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018.

36. Parcollet, T.; Ravanelli, M.; Morchid, M.; Linarès, G.; Trabelsi, C.; De Mori, R.; Bengio, Y. Quaternion Recurrent Neural Networks. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019; pp. 1–19.
37. Parcollet, T.; Morchid, M.; Linarès, G. A Survey of Quaternion Neural Networks. *Artif. Intell. Rev.* **2019**, *53*, 2957–2982. [CrossRef]
38. Vecchi, R.; Scardapane, S.; Comminiello, D.; Uncini, A. Compressing Deep-Quaternion Neural Networks with Targeted Regularisation. *CAAI Trans. Intell. Technol.* **2020**, *5*, 172–176. [CrossRef]
39. Parcollet, T.; Morchid, M.; Linarès, G. Deep Quaternion Neural Networks for Spoken Language Understanding. In Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 504–511.
40. Parcollet, T.; Morchid, M.; Linarès, G. Quaternion Convolutional Neural Networks for Heterogeneous Image Processing. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8514–8518.
41. Yin, Q.; Wang, J.; Luo, X.; Zhai, J.; Jha, S.K.; Shi, Y. Quaternion Convolutional Neural Network for Color Image Classification and Forensics. *IEEE Access* **2019**, *7*, 20293–20301. [CrossRef]
42. Comminiello, D.; Lella, M.; Scardapane, S.; Uncini, A. Quaternion Convolutional Neural Networks for Detection and Localization of 3D Sound Events. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8533–8537.
43. Ricciardi Celsi, M.; Scardapane, S.; Comminiello, D. Quaternion Neural Networks for 3D Sound Source Localization in Reverberant Environments. In Proceedings of the IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), Espoo, Finland, 21–24 September 2020; pp. 1–6.
44. Parcollet, T.; Zhang, Y.; Morchid, M.; Trabelsi, C.; Linarès, G.; De Mori, R.; Bengio, Y. Quaternion Convolutional Neural Networks for End-to-End Automatic Speech Recognition. *arXiv* **2018**, arXiv:1806.07789.
45. Grassucci, E.; Cicero, E.; Comminiello, D. Quaternion Generative Adversarial Networks. *arXiv* **2021**, arXiv:2104.09630v1.
46. Danielewski, M.; Sapa, L. Foundations of the Quaternion Quantum Mechanics. *Entropy* **2020**, *22*, 1424. [CrossRef]
47. Sugitomo, S.; Maeta, K. Quaternion Valued Risk Diversification. *Entropy* **2020**, *22*, 390. [CrossRef] [PubMed]
48. Szczęsna, A. Quaternion Entropy for Analysis of Gait Data. *Entropy* **2019**, *21*, 79. [CrossRef] [PubMed]
49. Neeser, F.; Massey, J. Proper complex random processes with applications to information theory. *IEEE Trans. Inf. Theory* **1993**, *39*, 1293–1302. [CrossRef]
50. Le Bihan, N. The Geometry of Proper Quaternion Random Variables. *Signal Process.* **2017**, *138*, 106–116. [CrossRef]
51. Vìa, J.; Ramìrez, D.; Santamarìa, I. Properness and Widely Linear Processing of Quaternion Random Vectors. *IEEE Trans. Inf. Theory* **2010**, *56*, 3502–3515. [CrossRef]
52. Took, C.C.; Mandic, D.P. Augmented Second-Order Statistics of Quaternion Random Signals. *Signal Process.* **2011**, *91*, 214–224. [CrossRef]
53. Navarro-Moreno, J.; Fernández-Alcalá, R.M.; Ruiz-Molina, J.C. A Quaternion Widely Linear Model for Nonlinear Gaussian Estimation. *IEEE Trans. Signal Process.* **2014**, *62*, 6414–6424. [CrossRef]
54. Vìa, J.; Palomar, D.P.; Vielva, L.; Santamarìa, I. Quaternion ICA From Second-Order Statistics. *IEEE Trans. Signal Process.* **2011**, *59*, 1586–1600. [CrossRef]
55. Zhang, Y.N.; Li, B.Z. Generalized Uncertainty Principles for the Two-Sided Quaternion Linear Canonical Transform. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4594–4598.
56. Ortolani, F.; Comminiello, D.; Uncini, A. The Widely Linear Block Quaternion Least Mean Square Algorithm For Fast Computation In 3D Audio Systems. In Proceedings of the IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri sul Mare, Italy, 13–16 September 2016.
57. Ogunfunmi, T.; Safarian, C. A Quaternion Kernel Minimum Error Entropy Adaptive Filter. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4149–4153.
58. Xia, Y.; Xiang, M.; Li, Z.; Mandic, D.P. *Adaptive Learning Methods for Nonlinear System Modeling*; Chapter Echo State Networks for Multidimensional Data: Exploiting Noncircularity and Widely Linear; Elsevier: Amsterdam, The Netherlands, 2018; pp. 267–288.
59. Vahdat, A.; Andriyash, E.; Macready, W.G. Undirected Graphical Models as Approximate Posteriors. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 9680–9689.
60. Smith, J.D. Some Observations on the Concepts of Information-Theoretic Entropy and Randomness. *Entropy* **2001**, *3*, 1–11. [CrossRef]
61. Santamaria, I.; Crespo, P.M.; Lameiro, C.; Schreier, P.J. Information-Theoretic Analysis of a Family of Improper Discrete Constellations. *Entropy* **2018**, *20*, 45. [CrossRef] [PubMed]
62. Ell, T.A.; Sangwine, S.J. Quaternion Involutions and Anti-Involutions. *Comput. Math. Appl.* **2007**, *53*, 137–143. [CrossRef]
63. Kingma, D.P.; Welling, M. An Introduction to Variational Autoencoders. *Found. Trends Mach. Learn.* **2019**, *12*, 307–392. [CrossRef]
64. Hou, X.; Shen, L.; Sun, K.; Qiu, G. Deep Feature Consistent Variational Autoencoder. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 1133–1141.

65. Tomczak, J.M.; Welling, M. VAE with a VampPrior. In Proceedings of the 21st International Conference on Artificial Intelligence and Statistics, Playa Blanca, Lanzarote, Spain, 9–11 April 2018; pp. 1214–1223.

66. Maaløe, L.; Fraccaro, M.; Liévin, V.; Winder, O. BIVA: A Very Deep Hierarchy of Latent Variables for Generative Modeling. *arXiv* **2019**, arXiv:1902.02102.

Short Biography of Authors

Eleonora Grassucci is a PhD student in Information and Communication Technologies with the Department of Information Engineering, Electronics and Telecommunications at Sapienza University of Rome, Italy. She earned a Master Degree in Data Science at Sapienza University of Rome.

Danilo Comminiello is an Associate Professor with the Department of Information Engineering, Electronics and Telecommunications at Sapienza University of Rome. His current research interests deep learning and computational intelligence methodologies. Danilo Comminiello is a member of the IEEE Machine Learning for Signal Processing Technical Committee. He is one of the editors of the book "Adaptive Learning Methods for Nonlinear Modeling" (D. Comminiello and J. C. Principe, eds.), Elsevier, 2018.

Aurelio Uncini is a Full Professor in Circuits Theory with the Department of Information Engineering, Electronics and Telecommunications at Sapienza University of Rome, Italy. He is author of more than 250 papers in the field of optimization algorithms for circuits design, signal processing, neural networks and audio signal processing. His present research interests include also linear and nonlinear adaptive filters, blind signal processing, adaptive audio and array processing, machine learning for signal processing.