Towards a Statistical Audio Signal Processing Framework Based on the I-Divergence*

○ Hirokazu Kameoka
NTT Communication Science Laboratories

1 Introduction

For many applications in statistical signal processing, a statistical model in which a short-time signal is assumed to be a set of samples drawn from a zero-mean stationary Gaussian process is very often employed. In the frequency domain, this assumption amounts to considering that each frequency component of the signal is generated according to a zero-mean circular complex normal distribution with a different variance. The variance of the distribution in this context is usually called the power spectral density (PSD). The maximum likelihood estimation problem under this statistical model involves determining the PSD estimate from an observed signal, which is shown to be equivalent to minimizing a divergence measure called the Itakura-Saito (IS) divergence between the sample power spectrum and the PSD of the assumed stochastic process [1]. While the solution is obvious when no modeling constraints are imposed on the PSD, it is generally necessary to develop an appropriate optimization algorithm according to this criterion when the PSD is assumed to have a certain structure that can be described using a small number of parameters (for example, using an all-pole model or non-negative matrix factorization model). However, for some classes of parametric models the cost function is sometimes numerically difficult to optimize, due to the highly non-convex nature of the divergence function. For example, as reported in [2], it has been found in practice to be prone to numerical instability and local minima during optimization when applied as a goodness-of-fit criterion for non-negative matrix factorization (NMF).

In this paper, we focus on another divergence measure between two non-negative functions, called the I-divergence [3], which has several remarkable features. Firstly, Csiszár showed in [3] that under non-negativity constraints, the only discrepancy measure consistent with certain fundamental axioms such as locality, regularity and composition-consistency is the I-divergence. Secondly, according to the results obtained in NMF-based single channel source separation tasks under many different model-fitting criteria [4], the use of the I-divergence has been found to provide the best performance. This should indicate that the stochastic process assumption underlying the I-divergence is describing the actual statistics of audio signals well. Thirdly, it is mathematically convenient to derive optimization algorithms for particular classes of parametric spectrum models such as those introduced in [5, 6]. Here, one intriguing question arises: What kind of stochastic process are we implicitly assuming when we choose to use the I-divergence as a spectral distortion measure? The objective of this paper is to answer this question.

2 Review of Itakura-Saito Divergence

In this section, we briefly review the way in which the IS divergence is derived from the stationary Gaussian random process assumption. Let $\boldsymbol{x} = (x_1, \cdots, x_I)^{\mathrm{T}} \in \mathbb{R}^I$ be a real vector of a discrete-time signal, that is assumed to have been drawn from a zero-mean Gaussian random process $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. The Fourier transform of $\boldsymbol{x}$, given by $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{x} \in \mathbb{C}^I$, follows a zero-mean multivariate complex normal distribution with covariance matrix $\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{H}}$, where $\boldsymbol{F} \in \mathbb{C}^{I\times I}$ is the discrete Fourier transform matrix. If we assume stationarity (and circularity), the covariance matrix $\boldsymbol{\Sigma}$ belongs to the class of nonnegative definite symmetric Toeplitz circulant $I \times I$ dimensional matrices. $\boldsymbol{\Sigma}$ is then shown to be exactly diagonalized by $\boldsymbol{F}$ so that we obtain $\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{H}} = \mathrm{diag}(\lambda_1, \cdots, \lambda_I)$ where $\lambda_1, \cdots, \lambda_I$ are the eigenvalues of $\boldsymbol{\Sigma}$, corresponding to the power spectral densities (PSDs). This indicates that each element of $\boldsymbol{z}$, z_i $(1 \leq i \leq I)$, independently follows a zero-mean complex normal distribution with variance λ_i: $z_i \sim \mathcal{N}_{\mathbb{C}}(0, \lambda_i)$. Now, suppose that we are given a Fourier component, z. We consider the maximum likelihood estimation of λ on the basis of the generative model $z \sim \mathcal{N}_{\mathbb{C}}(0, \lambda)$. By differentiating the log-likelihood $L_{\mathcal{N}}(\lambda) = -\log \pi\lambda - |z|^2/\lambda$ with respect to λ and setting the result at zero, we determine the maximum likelihood estimator $\lambda = |z|^2$ and hence we obtain $L_{\mathcal{N}}(|z|^2) \geq L_{\mathcal{N}}(\lambda)$. By subtracting the right-hand side of this inequality from the left-hand side, we obtain a non-negative measure:

$$L_{\mathcal{N}}(|z|^2) - L_{\mathcal{N}}(\lambda) = \frac{|z|^2}{\lambda} - \log\frac{|z|^2}{\lambda} - 1 \geq 0, \qquad (1)$$

which is equal to the IS divergence [1]. This quantity is shown to reach 0 only when $\lambda = |z|^2$.

3 Complex Poisson Distribution

We propose a probability density function of a complex number z, called the complex Poisson distribution, mention some of its properties, and then show that this distribution leads to the I-divergence in the same way that we derived the IS divergence from the complex normal distribution in the previous section.

*I ダイバージェンスに基づく統計的音響信号処理の枠組に向けて, 亀岡弘和 (NTT CS 研)

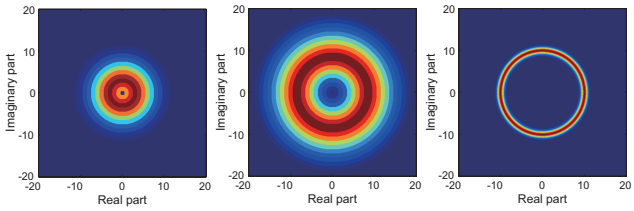

Fig. 1 Illustration of how the density function changes according to the parameter settings: (left) $(\lambda, p) = (5, 1)$, (middle) $(\lambda, p) = (10, 1)$, and (right) $(\lambda, p) = (10^2, 2)$.

Definition 1 (Complex Poisson Distribution)**.** *The complex Poisson distribution is a density function of a circular complex random variable z defined over the support $z \in \mathcal{D} = \{z \in \mathbb{C} \mid |z|^p \in \mathbb{N}\}$ such that*

$$f_z(z; \lambda, p) = \frac{pe^{-\lambda}}{2\pi} \frac{|z|^{p-2}\lambda^{|z|^p}}{(|z|^p)!}, \tag{2}$$

where $\lambda \in \mathbb{R}^{\geq 0}$ and $p \in \mathbb{R}^{>0}$ are the parameters characterizing its distribution. We use the notation

$$z \sim \mathrm{cPois}(\lambda, p), \tag{3}$$

to indicate that a complex-valued random variable z follows a complex Poisson distribution.

Fig. 1 is an illustration of the complex Poisson distributions with different parameter settings. In the following we show several important properties related to the proposed circular distribution.

Property 1. *The integral of f_z over the support $\mathcal{D}$ is*

$$\int_{\mathcal{D}} f_z(z; \lambda, p) \mathrm{d}z = 1. \tag{4}$$

Property 2. *If z follows a complex Poisson distribution $z \sim \mathrm{cPois}(\lambda, p)$, then $k = |z|^p \in \mathbb{N}$ follows a Poisson distribution with mean λ:*

$$k \sim \frac{\lambda^k e^{-\lambda}}{k!}. \tag{5}$$

Property 3. *If z follows a complex Poisson distribution such that $z \sim \mathrm{cPois}(\lambda, p)$, the qth-order moments defined by $\beta_{n;m} := \mathbb{E}[z^n z^{*m}]$ [7], where n and m are natural numbers such that $q = n + m$, are given by*

$$\beta_{n;m} = \begin{cases} \mu'_{q/p} & (n = m) \\ 0 & (n \neq m) \end{cases}, \tag{6}$$

where μ'_ξ denotes the fractional moment of order $\xi > 0$ of the Poisson distribution with mean λ.

Example 1. *It follows from Property 3 that the mean $\mathbb{E}[z]$ and the variance $\mathbb{V}[z]$ of $z \sim \mathrm{cPois}(\lambda, p)$ are*

$$\mathbb{E}[z] = 0, \quad \mathbb{V}[z] = \begin{cases} \lambda^2 + \lambda & (p = 1) \\ \lambda & (p = 2) \end{cases}. \tag{7}$$

Property 4. *The maximum likelihood estimation of λ under the generative model $z \sim \mathrm{cPois}(\lambda, p)$ is equivalent to the problem of minimizing the I-divergence between $|z|^p$ and λ.*

Proof: This can be easily confirmed by considering only the terms involving λ in the log-likelihood, $L_{\mathcal{P}}(\lambda) = -\lambda + |z|^p \log \lambda$, which is maximized when $\lambda = |c|^p$. Thus, $L_{\mathcal{P}}(|z|^p) \geq L_{\mathcal{P}}(\lambda)$. Subtracting the right-hand side of this inequality from the left-hand side gives a non-negative measure

$$L_{\mathcal{P}}(|z|^p) - L_{\mathcal{P}}(\lambda) = |z|^p \log \frac{|z|^p}{\lambda} - (|z|^p - \lambda) \geq 0, \tag{8}$$

which is equal to the I-divergence. □

4 Construction of Stationary Process

Given a set consisting of Fourier components $\boldsymbol{z} = (z_1, \ldots, z_I)^{\mathrm{T}}$, let us assume that each component has been generated independently according to $z_i \sim \mathrm{cPois}(\lambda_i, p)$. The probability of $\boldsymbol{z}$ being generated is

$$f_{\boldsymbol{z}}(\boldsymbol{z}; \lambda, p) = \prod_{i=1}^{I} \frac{pe^{-\lambda_i}}{2\pi} \frac{|z_i|^{p-2}\lambda_i^{|z_i|^p}}{(|z_i|^p)!}, \tag{9}$$

where $\lambda = \{\lambda_i\}_{1 \leq i \leq I}$. We can easily show from what kind of probability distribution the time-domain signal, $\boldsymbol{x} = \boldsymbol{F}^{\mathrm{H}}\boldsymbol{z}$, is supposed to have been generated under this assumption. Since the (inverse) Fourier transform is a unitary transform, we know that $|\det \boldsymbol{F}| = 1$. The density function of $\boldsymbol{x}$ can thus be described in terms of $f_{\boldsymbol{z}}$ such that $f_{\boldsymbol{x}}(\boldsymbol{x}; \lambda, p) = f_{\boldsymbol{z}}(\boldsymbol{F}\boldsymbol{x}; \lambda, p)$.

It is important to note that when $p = 2$, λ corresponds to the PSD of the assumed stochastic process (see Example 1), in which case the maximum likelihood estimation of λ can be understood as the problem of fitting the PSD to the sample power spectrum $\{|z_i|^2\}_{1 \leq i \leq I}$ under the I-divergence criterion.

5 Conclusion

In this paper, we proposed a new probability density function of a circular complex random variable, mentioned some of its properties, and showed that this distribution can be used to construct a stochastic process that supports the use of the I-divergence as a well-founded spectral distortion measure. By using this stochastic process, we should be able to build a well-defined audio signal processing framework, allowing for the introduction of all kinds of spectrum models which behave well with respect to the I-divergence criterion.

Acknowledgement

The author would like to thank Dr. Naonori Ueda (NTT) for his fruitful discussions.

References

[1] Itakura, Ph.D. thesis, Nagoya University, 1972.
[2] Bertin et al., Proc. ICASSP'09, pp. 840–843, 2009.
[3] Csiszár, Ann. Stat., **19**(4), pp. 2032–2066, 1991.
[4] FitzGerald et al., Proc. ISSC'09, 2009.
[5] Kameoka, Ph.D. thesis, University of Tokyo, 2007.
[6] Kameoka, IEICE Tech. Rep., SP2010-74, pp. 29–34, 2010.
[7] Amblard et al., Signal Process., **53**(1), pp. 1–13, 1996.