

Shrinkage Priors on Complex-Valued Circular-Symmetric Autoregressive Processes

Hidemasa Oda¹ and Fumiyasu Komaki², *Member, IEEE*

Abstract—We investigate shrinkage priors on power spectral densities for complex-valued circular-symmetric autoregressive processes. We construct shrinkage predictive power spectral densities, which asymptotically dominate (i) the Bayesian predictive power spectral density based on the Jeffreys prior and (ii) the estimative power spectral density with the maximum likelihood estimator, where the Kullback–Leibler divergence from the true power spectral density to a predictive power spectral density is adopted as a risk. Furthermore, we propose general constructions of objective priors for Kähler parameter spaces by utilizing a positive continuous eigenfunction of the Laplace–Beltrami operator with a negative eigenvalue. We present numerical experiments on a complex-valued stationary autoregressive model of order 1.

Index Terms—Complex-valued Gaussian process, complex-valued signal processing, objective Bayes, shrinkage prior, information geometry, Kähler manifold, α -parallel prior.

I. INTRODUCTION

WE INVESTIGATE the time fluctuation of a single particle having a circular-symmetric distribution in a two-dimensional space. In this situation, the complex plane \mathbb{C} is often used for the representation of the process, and an observation of a single particle at different time points can be represented as a complex-valued vector. A complex-valued random vector Z is called circular-symmetric if, for any constant $\phi \in \mathbb{R}$, the distribution of $e^{\sqrt{-1}\phi}Z$ equals the distribution of Z .

We focus on complex-valued circular-symmetric discrete Gaussian processes, which are defined as complex-valued processes having finite-dimensional marginal distributions that are complex normal distributions. The precise definitions of a complex normal distribution and a complex-valued Gaussian process are given in Section II. The circular symmetry of complex normal distributions with zero mean is of great importance in practical applications [1]. Complex-valued processes are commonly used for directional processes, such as wind, radar, and sonar signals [2]. Furthermore, complex-valued

representations are widely used in diverse fields, such as econometrics [3] and complex-valued neural networks [4].

The power spectral density of a complex-valued Gaussian process is defined as a Fourier transform

$$S(\omega) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_h e^{-\sqrt{-1}h\omega} \quad (1)$$

of the autocovariances $\{\gamma_h\}_{h \in \mathbb{Z}}$ of the process, where $\omega \in [-\pi, \pi]$. We parametrize complex-valued Gaussian processes by complex variables $\theta = (\theta^1, \dots, \theta^p) \in \Theta \subset \mathbb{C}^p$. In other words, we regard the sequence $\{\gamma_h\}_{h \in \mathbb{Z}}$ of autocovariances of the process as functions $\{\gamma_h(\theta)\}_{h \in \mathbb{Z}}$ of the parameter θ . For each $\theta \in \Theta$, we denote the corresponding power spectral density by $S_\theta(\omega)$ or $S(\omega | \theta)$.

We aim to predict the joint distribution of future observations $w = (w_1, w_2, \dots)$ by observing current observations $z^{(N)} = (z_1, \dots, z_N) \in \mathbb{C}^N$ of size N from a complex-valued Gaussian process having the true parameter $\theta_0 \in \Theta$. We suppose that the sample w and the sample $z^{(N)}$ are independent; i.e., the sample w is taken from a different process of the same type or from the same process but a long time after the sample $z^{(N)}$ is taken. Let us consider the problem of constructing a power spectral density $\hat{S}^{(N)}$ that corresponds to the joint distribution of future observations w ; see [5]. The constructed power spectral density $\hat{S}^{(N)}$ is called a predictive power spectral density. More precisely, a predictive power spectral density $\hat{S}^{(N)}(\omega)$ is a function of an observation $z^{(N)}$ for each $\omega \in [-\pi, \pi]$. The goodness of the prediction is evaluated by the risk, which is defined as

$$\begin{aligned} R(\hat{S}^{(N)} | \theta_0) &:= E_{\theta_0} \left[D_{\text{KL}} \left(S_{\theta_0} \parallel \hat{S}^{(N)} \right) \right] \\ &= \int_{\mathbb{C}^N} D_{\text{KL}} \left(S_{\theta_0} \parallel \hat{S}^{(N)} \right) dP_{\theta_0}^{(N)}(z^{(N)}), \end{aligned} \quad (2)$$

where $P_{\theta_0}^{(N)}$ denotes the distribution of $z^{(N)}$ and the Kullback–Leibler divergence between two power spectral densities S_1 and S_2 is defined as

$$D_{\text{KL}}(S_1 \parallel S_2) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(-\log \frac{S_1(\omega)}{S_2(\omega)} - 1 + \frac{S_1(\omega)}{S_2(\omega)} \right) d\omega; \quad (3)$$

see Appendix B. The principal aim of this study is to construct a predictive power spectral density $\hat{S}^{(N)}$ with its risk $R(\hat{S}^{(N)} | \theta)$ being as small as possible for most $\theta \in \Theta$. There is another interpretation for the problem: we estimate the

Manuscript received May 12, 2020; revised February 4, 2021; accepted May 3, 2021. Date of publication May 12, 2021; date of current version July 14, 2021. This work was supported by the Japan Science and Technology Agency (JST) Core Research for Evolutionary Science and Technology (CREST) under Grant JPMJCR1763. (Corresponding author: Hidemasa Oda.)

Hidemasa Oda is with the Department of Mathematical Informatics, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: hidemasa_oda@mist.i.u-tokyo.ac.jp).

Fumiyasu Komaki is with the Department of Mathematical Informatics, The University of Tokyo, Tokyo 113-8656, Japan, and also with the RIKEN Center for Brain Science, Wako 351-0198, Japan (e-mail: komaki@mist.i.u-tokyo.ac.jp).

Communicated by E. Gassiat, Associate Editor for Probability and Statistics. Digital Object Identifier 10.1109/TIT.2021.3079512

underlying true power spectral density of the process under the loss given by (2) based on the Kullback–Leibler divergence.

There are two basic constructions for predictive power spectral densities. The first construction, called the estimative method, is $\hat{S}_{\hat{\theta}^{(N)}}$, where $\hat{\theta}^{(N)} = \hat{\theta}^{(N)}(z^{(N)})$ is an estimator for the true parameter θ_0 . The second construction, called the Bayesian predictive method, is

$$\hat{S}_{\pi}^{(N)}(\omega) := \int_{\Theta} S(\omega | \theta) \pi(\theta | z^{(N)}) d\theta \quad (4)$$

for a possibly improper prior π . Here,

$$\pi(\theta | z^{(N)}) := \frac{p_{\theta}(z^{(N)})\pi(\theta)}{\int_{\Theta} p_{\theta'}(z^{(N)})\pi(\theta') d\theta'} \quad (5)$$

denotes the posterior, given an observation $z^{(N)} \in \mathbb{C}^N$ based on the prior π , and $\int_{\Theta} \pi(\theta | z^{(N)}) d\theta < \infty$ is assumed. The power spectral density $\hat{S}_{\pi}^{(N)}$ is called the Bayesian predictive power spectral density based on the prior π . Herglotz's theorem asserts that a family $\{\gamma_h\}_{h \in \mathbb{Z}}$ of complex numbers parameterized by integers is the family of auto-covariances of a complex-valued stationary process if and only if there exists a spectral distribution function F such that $\gamma_h = \int_{-\pi}^{\pi} e^{i\sqrt{-1}h\omega} dF(\omega)$ and $\{\gamma_h\}_{h \in \mathbb{Z}}$ is positive semi-definite; i.e., $\sum_{i=1}^n \sum_{j=1}^n a_i \gamma_{i-j} \bar{a}_j \geq 0$ for any $n > 0$ and $(a_1, \dots, a_n) \in \mathbb{C}^n$; see Corollary 4.3.1 in [6]. Because

$$\begin{aligned} \hat{\gamma}_{\pi,h}^{(N)} &:= \int_{-\pi}^{\pi} e^{i\sqrt{-1}h\omega} \hat{S}_{\pi}^{(N)}(\omega) d\omega \\ &= \int_{\Theta} \gamma_h(\theta) \pi(\theta | z^{(N)}) d\theta \end{aligned} \quad (6)$$

is a weighted average of $\gamma_h(\theta)$, the process with the predictive power spectral density given by (4) is stationary as long as each process parameterized by $\theta \in \Theta$ is stationary. If a prior π is given and the risk is defined as (2), the Bayesian predictive power spectral density $\hat{S}_{\pi}^{(N)}$ minimizes the Bayes risk

$$r(\hat{S}^{(N)} | \pi) := \int_{\Theta} R(\hat{S}^{(N)} | \theta) \pi(\theta) d\theta \quad (7)$$

among all the predictive power spectral densities $\hat{S}^{(N)}$ as long as the Bayes risk $r(\hat{S}_{\pi}^{(N)} | \pi)$ is finite. Therefore, the remaining problem is to determine and construct an appropriate prior π .

Non-informative priors for time-series models, such as the Jeffreys prior, which is usually improper, have been discussed in previous works [7]–[9]. We propose a proper prior $\pi^{(-1)}$ defined on the complex parameter space $\Theta \subset \mathbb{C}^p$ for the complex-valued stationary autoregressive processes $\text{AR}(p; \mathbb{C})$ of order $p \geq 1$. The Bayesian predictive power spectral density $\hat{S}_{\pi^{(-1)}}^{(N)}$ based on the proposed prior $\pi^{(-1)}$ asymptotically dominates the estimative power spectral density $\hat{S}_{\hat{\theta}^{(N)}}$ with the maximum likelihood estimator $\hat{\theta}^{(N)}$. Moreover, the proposed predictive power spectral density $\hat{S}_{\pi^{(-1)}}^{(N)}$ asymptotically dominates the Bayesian predictive power spectral density $\hat{S}_{\pi_J}^{(N)}$ based on the Jeffreys prior π_J , and the $O(N^{-2})$ term of the risk improvement is constant regardless of $\theta \in \Theta$:

$$R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi^{(-1)}}^{(N)} | \theta) = \frac{2p(p+1)}{N^2} + O(N^{-\frac{5}{2}}), \quad (8)$$

which is summarized as the Main Theorem in Section IV.

An eigenfunction ϕ of the Laplacian (Laplace–Beltrami operator) Δ plays a crucial role in constructing the proposed prior $\pi^{(-1)}$. The Laplacian is a differential operator that does not depend on the specific choice of parameterizations of the parameter space; see Appendix A. This operator transforms a scalar function defined on the parameter space to another scalar function defined on the parameter space. References [10] and [11] stress the importance of the super-harmonicity for the shrinkage effect in the estimation of the mean of a multivariate normal distribution. More generally, it is known that the super-harmonicity of the ratio of the proposed prior to the Jeffreys prior is the key to inducing the shrinkage effect [12]. Another important property in the construction of the proposed prior $\pi^{(-1)}$ is the Kählerness, the generalization of the concept of exponential families, of the complex parameter space Θ . The parameter space of the complex-valued stationary autoregressive moving average processes $\text{ARMA}(p, q; \mathbb{C})$ is shown to be Kähler in [13]. We give a general construction of priors utilizing a positive continuous eigenfunction $\phi > 0$ of the Laplacian Δ with a negative eigenvalue $-K < 0$, i.e., $\Delta\phi = -K\phi < 0$. We define a family of priors $\{\pi^{(\kappa)}\}_{\kappa \in \mathbb{R}}$ by $\pi^{(\kappa)} := \phi^{-\kappa+1} \pi_J$, which are called κ -priors. We prove that if $-1 \leq \kappa < 1$, then $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ asymptotically dominates $\hat{S}_{\pi_J}^{(N)}$. To maximize the worst case of the risk improvement, we propose the κ -prior for $\kappa = -1$ that achieves a constant risk improvement.

The remainder of this paper is organized as follows. In Section II, we give the asymptotic expansion of Bayesian predictive power spectral densities $\hat{S}_{\pi}^{(N)}$ for complex-valued autoregressive moving average processes. In Section III, a specific Kähler parameterization for $\text{AR}(p; \mathbb{C})$ is introduced. In Section IV, the Main Theorem (8) is stated. In Section V, we explicitly give the construction of the positive continuous eigenfunction ϕ with a negative eigenvalue $-K = -p(p+1)$ on the Kähler parameter space for $\text{AR}(p; \mathbb{C})$. Furthermore, we show that the family $\{\pi^{(\kappa)}\}_{\kappa \in \mathbb{R}}$ of the proposed κ -priors $\pi^{(\kappa)}$ for $\text{AR}(p; \mathbb{C})$ is a family of α -parallel priors, which was introduced in [14], with $\alpha = (1 - \kappa)/2$. The generalization of the family of the proposed priors $\{\pi^{(\kappa)}\}_{\kappa \in \mathbb{R}}$ and its relation with the α -parallel priors for the i.i.d. case is discussed in Section VI. In Section VII, numerical experiments are reported for the value of the risk differences $N^2(R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi^{(-1)}}^{(N)} | \theta))$ for $\text{AR}(1; \mathbb{C})$.

II. BAYESIAN PREDICTIVE POWER SPECTRAL DENSITIES FOR COMPLEX-VALUED GAUSSIAN PROCESSES

As explained in the Introduction, our aim is to construct a predictive power spectral density $\hat{S}^{(N)}$ after observing a sample $z^{(N)} \in \mathbb{C}^N$ of size N . In this section, we present the asymptotic expansion of Bayesian predictive power spectral densities $\hat{S}_{\pi}^{(N)}$ for complex-valued autoregressive moving average processes. This asymptotic expansion is a basic tool for assessing the performance of the choice of a prior π . In the remainder of this paper, the Einstein notation is assumed. Therefore, the summation is automatically taken over those indices that appear exactly twice, once as a superscript and once as a subscript. The symbols $\alpha, \beta, \gamma, \dots$ run through the

indices $\{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, and the symbols i, j, k, \dots run through the indices $\{1, \dots, p\}$.

We first define the multivariate complex normal distribution. Let $\mu \in \mathbb{C}^N$ and $\Sigma^{(N)}$ be an $N \times N$ complex-valued positive definite Hermitian matrix. Note that the determinant $|\Sigma^{(N)}|$ of the matrix $\Sigma^{(N)}$ is positive. An N -dimensional complex normal distribution (complex-valued circular-symmetric multivariate normal distribution) with mean μ and variance $\Sigma^{(N)}$ is defined by its probability density function:

$$p(z^{(N)} | \mu, \Sigma^{(N)}) := \frac{1}{\pi^N |\Sigma^{(N)}|} e^{-(z^{(N)} - \mu)^* (\Sigma^{(N)})^{-1} (z^{(N)} - \mu)}, \quad (9)$$

where $z^{(N)} = (z^1, \dots, z^N) \in \mathbb{C}^N$ and $z^{(N)*}$ denotes the complex conjugate transpose of $z^{(N)}$; see [1]. The circular symmetry of a complex normal distribution with mean $0 \in \mathbb{C}^N$ is understood from the definition (9). The complex normal distribution with mean $0 \in \mathbb{C}^N$ and the identity matrix of size N as the variance-covariance matrix is called the standard complex normal distribution of size N . If we let $z^i = x^i + \sqrt{-1}y^i$ for $i = 1, \dots, N$, the $2N$ -dimensional real-valued vector $(x^1, \dots, x^N, y^1, \dots, y^N)$ follows the $2N$ -dimensional real-valued multivariate normal distribution with mean $(\Re(\mu), \Im(\mu))$ and the following variance-covariance matrix:

$$\begin{bmatrix} \frac{1}{2} \Re(\Sigma) & -\frac{1}{2} \Im(\Sigma) \\ \frac{1}{2} \Im(\Sigma) & \frac{1}{2} \Re(\Sigma) \end{bmatrix}. \quad (10)$$

Therefore, an N -dimensional complex normal distribution is a special case of a $2N$ -dimensional real normal distribution; however, the opposite is not true.

A complex-valued discrete process $\{Z_t\}_{t \in \mathbb{Z}}$ is called a Gaussian process (complex-valued circular-symmetric discrete Gaussian process) if the tuple $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_N})$ of size N follows a complex normal distribution for any N and any $t_1, t_2, \dots, t_N \in \mathbb{Z}$. A complex Gaussian white noise $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ with variance σ^2 is a Gaussian process and $\frac{1}{\sigma}(\varepsilon_{t_1}, \dots, \varepsilon_{t_N})$ follows a standard complex normal distribution of size N for any N and any $t_1, t_2, \dots, t_N \in \mathbb{Z}$. For a strongly stationary Gaussian process $\{Z_t\}_{t \in \mathbb{Z}}$, we define the autocovariance γ_h of order h as the covariance of Z_{t+h} and \bar{Z}_t . Note that the autocovariances $\{\gamma_h\}_{h \in \mathbb{Z}}$ are complex-valued, and the relation $\overline{\gamma_h} = \gamma_{-h}$ holds for any h . The power spectral density of the process is defined as a Fourier transform (1) of the autocovariances $\{\gamma_h\}_{h \in \mathbb{Z}}$. Because we consider complex-valued processes, power spectral densities are not generally even functions on $[-\pi, \pi]$. For the observation $z^{(N)} = (z_1, \dots, z_N)$ of size N from a Gaussian process with mean zero, let us denote its probability density by $p^{(N)}(z^{(N)})$. The probability density $p^{(N)}(z^{(N)})$ is explicitly calculated as (9) with mean $\mu = 0$ and the following variance-covariance matrix:

$$\Sigma^{(N)} := \begin{bmatrix} \gamma_0 & \gamma_{-1} & \cdots & \gamma_{-N+1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{-N+2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N-1} & \gamma_{N-2} & \cdots & \gamma_0 \end{bmatrix}. \quad (11)$$

As a special case of a strongly stationary Gaussian process with mean zero, we introduce a complex-valued autoregressive moving average (ARMA) process. A complex-valued ARMA process of degree (p, q) is a Gaussian process that satisfies the relation

$$Z_t = - \sum_{i=1}^p a_i Z_{t-i} + \varepsilon_t + \sum_{i=1}^q b_i \varepsilon_{t-i} \quad (12)$$

for all t , where $a_1, \dots, a_p, b_1, \dots, b_q$ are complex-valued coefficients and ε_t is a complex Gaussian white noise with variance σ^2 . We assume that the polynomials $z^p(1 + \sum_{i=1}^p a_i z^{-i})$ and $z^q(1 + \sum_{i=1}^q b_i z^{-i})$ have no common roots in order to ensure identifiability; see Section 3.1 in [6]. We denote the statistical model of complex-valued stationary ARMA processes by $\text{ARMA}(p, q; \mathbb{C})$ in the present paper. If $q = 0$, we call the model a complex-valued stationary autoregressive model and denote it by $\text{AR}(p; \mathbb{C})$. We denote the model of real-valued stationary autoregressive processes of order p by $\text{AR}(p; \mathbb{R})$. The power spectral density (1) of the ARMA model (12) is explicitly given by

$$\frac{\sigma^2}{2\pi} \left| \frac{1 + \sum_{i=1}^q b_i e^{-i\sqrt{-1}\omega}}{1 + \sum_{i=1}^p a_i e^{-i\sqrt{-1}\omega}} \right|^2. \quad (13)$$

In what follows, we prepare the relevant mathematical notation. Suppose Gaussian processes are parameterized by complex parameters $\theta \in \Theta \subset \mathbb{C}^p$. Because we consider complex parameters $\theta \in \Theta \subset \mathbb{C}^p$, we utilize Wirtinger calculus; see Appendix A. Corresponding to the i -th complex coordinate θ^i , there exist Wirtinger derivatives ∂_i and $\bar{\partial}_i$. For simplicity of notation, we set

$$S_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c} := S^{-1}(D_{\alpha_1 \dots \alpha_a} S) S^{-1}(D_{\beta_1 \dots \beta_b} S) \cdots S^{-1}(D_{\gamma_1 \dots \gamma_c} S) \quad (14)$$

for a power spectral density $S = S_\theta$ and indices $\alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, \dots, \gamma_1, \dots, \gamma_c \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, where $D_{\alpha_1 \dots \alpha_a} := \partial_{\alpha_1} \cdots \partial_{\alpha_a}$. For example, $S_{1\bar{1}, 2} = S^{-1}(\partial_1 \bar{\partial}_1 S) S^{-1}(\partial_2 S)$. We also set

$$M_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c} := \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c}(\omega) d\omega \quad (15)$$

and define the quantities $g_{\alpha\beta}$, $T_{\alpha\beta\gamma}$, and $\Gamma_{\alpha\beta\gamma}^{(m)}$ as

$$g_{\alpha\beta} := M_{\alpha, \beta} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\partial_\alpha \log S) (\partial_\beta \log S) d\omega, \quad (16)$$

$$T_{\alpha\beta\gamma} := 2M_{\alpha, \beta, \gamma} = \frac{1}{\pi} \int_{-\pi}^{\pi} (\partial_\alpha \log S) (\partial_\beta \log S) (\partial_\gamma \log S) d\omega, \quad (17)$$

$$\Gamma_{\alpha\beta\gamma}^{(m)} := M_{\alpha\beta, \gamma} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial_\alpha \partial_\beta S}{S} (\partial_\gamma \log S) d\omega \quad (18)$$

for $\alpha, \beta, \gamma \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. The quantities $\Gamma_{\alpha\beta\gamma}^{(m)}$ correspond to the coefficients of the mixture connection (m -connection) $\nabla^{(m)}$; see Appendix I.

The complex-valued $2p \times 2p$ matrix $[g_{\alpha\beta}]$ is called the Fisher information matrix, which naturally induces the metric on the complex parameter space Θ . The inner product of two functions $N_1 = N_1(\omega | \theta)$ and $N_2 = N_2(\omega | \theta)$ defined on $[-\pi, \pi]$ at $\theta \in \Theta$ is defined as

$$\langle N_1, N_2 \rangle_\theta := \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{N_1(\omega | \theta) N_2(\omega | \theta)}{S(\omega | \theta)} d\omega, \quad (19)$$

and the norm $\|N\|_\theta$ of a function $N = N(\omega | \theta)$ at $\theta \in \Theta$ is defined as $\|N\|_\theta^2 := \langle N, N \rangle_\theta$; see also [5]. This form (16) of the Fisher information matrix was introduced in [15] for real-valued time-series analysis. For real-valued processes, the constant 4π , rather than 2π , is usually used in the denominator in (16); see [5], [15]–[17]. On the other hand, the constant 2π is used for signal processing; see [13], [18]. For complex-valued processes, it is natural to use the constant 2π as in (16) because it yields $g_{\alpha\beta} = \frac{1}{N} E[-\partial_\alpha \partial_\beta l^{(N)}] + O(N^{-1})$, where $l^{(N)}$ denotes the log-likelihood (56); see Proposition 4.2 in Appendix D. Let us denote by $[g^{\alpha\beta}]$ the inverse matrix of the Fisher information matrix $[g_{\alpha\beta}]$, i.e., $g_{\alpha\gamma} g^{\gamma\beta} = \delta_\alpha^\beta$ for the Kronecker delta δ_α^β . The prior defined as the square root of the determinant of the $2p \times 2p$ complex-valued matrix $[g_{\alpha\beta}]$ is called the Jeffreys prior and denoted by π_J in the present paper.

Asymptotic expansions of Bayesian predictive power spectral densities play an important role in the present paper. Let us fix a possibly improper prior π and assume that $\int_{\Theta} p_\theta(z^{(N)}) \pi(\theta) d\theta$ is finite for any $z^{(N)} \in \mathbb{C}^N$ and that the Bayesian predictive power spectral density (4) exists for any $\omega \in [-\pi, \pi]$. The asymptotic expansion of a Bayesian predictive power spectral density (4) of a complex-valued ARMA process around the maximum likelihood estimator $\hat{\theta} = \hat{\theta}^{(N)}(z^{(N)})$ is

$$\begin{aligned} \hat{S}_\pi^{(N)}(\omega) &= S(\omega | \hat{\theta}) \\ &+ \frac{1}{N} \left(G_\pi^{(N)}(\omega | \hat{\theta}) + H^{(N)}(\omega | \hat{\theta}) \right) \\ &+ O_P(N^{-\frac{3}{2}}), \end{aligned} \quad (20)$$

where functions $G_\pi^{(N)}$ and $H^{(N)}$ represent the parallel and orthogonal parts of the quantity $N(\hat{S}_\pi^{(N)} - S_{\hat{\theta}^{(N)}})$, respectively; see Appendix F. Functions $G_\pi^{(N)}$ and $H^{(N)}$ are explicitly given by

$$\begin{aligned} G_\pi^{(N)}(\omega | \theta) &:= g^{\alpha\beta}(\omega | \theta) \left(\partial_\alpha \log \frac{\pi}{\pi_J}(\theta) + \frac{1}{2} T_\alpha(\omega | \theta) \right) \partial_\beta S(\omega | \theta), \end{aligned} \quad (21)$$

$$\begin{aligned} H^{(N)}(\omega | \theta) &:= \frac{1}{2} g^{\alpha\beta}(\omega | \theta) \left(\partial_\alpha \partial_\beta S(\omega | \theta) - \Gamma_{\alpha\beta\gamma}^{(m)}(\theta) \partial_\gamma S(\omega | \theta) \right), \end{aligned} \quad (22)$$

where $\Gamma_{\alpha\beta\gamma}^{(m)}(\theta) := \Gamma_{\alpha\beta\delta}^{(m)}(\theta) g^{\delta\gamma}(\theta)$ and $T_\alpha(\theta) := T_{\alpha\beta\gamma}(\theta) g^{\beta\gamma}(\theta)$. Note first that $G_\pi^{(N)}$ and $H^{(N)}$ are orthogonal in the sense that $\langle G_\pi^{(N)}, H^{(N)} \rangle_\theta = 0$ for any $\theta \in \Theta$. Note also that, while the parallel part $G_\pi^{(N)}$ may depend on the choice of prior π , the orthogonal part $H^{(N)}$ is independent of the choice; see [12] for more details. See Appendix F for the proof of the expansion in (20).

The Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$ minimizes the Bayes risk (7) among all the predictive power spectral densities $\hat{S}^{(N)}$ as long as the Bayes risk of $\hat{S}_\pi^{(N)}$ is finite; see Appendix B. Therefore, once we have a prior π , we can calculate the predictive power spectral density $\hat{S}_\pi^{(N)}$ that minimizes the Bayes risk (7). The only remaining problem is to find a reasonable prior π .

III. KÄHLER PARAMETER SPACES FOR COMPLEX-VALUED AUTOREGRESSIVE PROCESSES

Let us consider a family $\{S_\theta\}_{\theta \in \Theta}$ of power spectral densities of complex-valued stationary ARMA processes, where $\Theta \subset \mathbb{C}^p$ is a complex parameter space. If the Fisher information matrix $[g_{\alpha\beta}]$ of the process satisfies the relations $g_{ij} = g_{\bar{i}\bar{j}} = 0$, $g_{i\bar{j}} = g_{\bar{j}i} = \overline{g_{\bar{j}i}} = \overline{g_{i\bar{j}}}$ for all $i, j = 1, \dots, p$ and the relations $\partial_i g_{j\bar{k}} = \partial_{\bar{j}} g_{i\bar{k}}$, $\partial_{\bar{i}} g_{j\bar{k}} = \partial_{\bar{k}} g_{j\bar{i}}$ for all $i, j, k = 1, \dots, p$, we say that the complex parameter space Θ is Kähler; see also Appendix A. The Kählerness of the complex parameter space Θ plays an important role in the construction of priors.

A specific complex parameter space $\Theta \subset \mathbb{C}^{p+q}$ for complex-valued stationary autoregressive moving average processes ARMA($p, q; \mathbb{C}$) was shown to be Kähler in [13]. We focus on the specific Kähler parameter space Θ for complex-valued stationary autoregressive processes AR($p; \mathbb{C}$). We examine the power spectral densities of AR($p; \mathbb{C}$) of the form

$$S(\omega) = \frac{1}{2\pi \left| \prod_{i=1}^p (1 - \xi^i e^{-\sqrt{-1}\omega}) \right|^2}, \quad (23)$$

where complex parameters $\xi = (\xi^1, \dots, \xi^p)$ are roots of the polynomial $z^p (1 + \sum_{i=1}^p a_i z^{-i})$ of the formal variable z and $\sigma^2 = 1$ is assumed. From the stationarity condition, we assume that $|\xi^i| < 1$ for any $i = 1, \dots, p$.

We define the parameter space $\tilde{\Theta}_1 \subset \mathbb{C}^p$ as

$$\begin{aligned} \tilde{\Theta}_1 &:= \{(\xi^1, \dots, \xi^p) \in \mathbb{C}^p \mid |\xi^i| < 1 \text{ for any } i = 1, \dots, p\} \\ &= U \times U \times \dots \times U, \end{aligned}$$

where U is the open unit disk in the complex plane \mathbb{C} . In this specific parameterization $\xi = (\xi^1, \dots, \xi^p)$, the center $0 = (0, \dots, 0) \in \mathbb{C}^p$ corresponds to the white noise process. Because we wish to ignore the measure-zero subset, where the denominator of (23) has multiple roots, we restrict our attention to the dense subset

$$\Theta_1 := \left\{ (\xi^1, \dots, \xi^p) \in \tilde{\Theta}_1 \mid \xi^i \neq \xi^j \text{ for any } i, j = 1, \dots, p \right\} \quad (24)$$

of the original parameter space $\tilde{\Theta}_1$. The parameter space Θ_1 is a complex manifold of complex dimension p because the set of

complex variables $\xi = (\xi^1, \dots, \xi^p)$ yields a local coordinate of the space, and the space Θ_1 is open as a topological space with the boundary $\partial\Theta_1$. In particular, Θ_1 is relatively compact but not compact.

For the specific parameterization $\xi = (\xi^1, \dots, \xi^p)$ defined in (23) for $\text{AR}(p, \mathbb{C})$, the Fisher information matrix is explicitly given by

$$g_{ij} = g_{\bar{i}\bar{j}} = 0, \quad g_{\bar{i}j} = g_{j\bar{i}} = \overline{g_{j\bar{i}}} = \overline{g_{\bar{i}j}} = \frac{1}{1 - \xi^i \bar{\xi}^j} \quad (25)$$

for $i, j = 1, \dots, p$. Therefore, the complex parameter space Θ_1 is Kähler. This is a very important property of the complex parameter space Θ_1 for analyzing the super-harmonic property of priors. The $p \times p$ complex-valued matrix $[g_{i\bar{j}}]$ is positive definite if and only if the denominator of (23) has no multiple roots. Note that, on the parameter space Θ_1 , there exists the inverse $[g^{i\bar{j}}]$ of the Fisher information matrix, which is necessary to define the Laplacian on the parameter space; see Appendix A.

For a Kähler parameter space, the Jeffreys prior is the determinant of the $p \times p$ complex-valued Hermitian matrix $[g_{i\bar{j}}]$; see Appendix A. For the specific parameterization $\xi = (\xi^1, \dots, \xi^p)$ defined in (23) for $\text{AR}(p, \mathbb{C})$, the Jeffreys prior π_J is explicitly given by

$$\pi_J(\xi) = \frac{\prod_{1 \leq i < j \leq p} |\xi^i - \xi^j|^2}{\prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j)}; \quad (26)$$

see also [13]. The Jeffreys prior (26) for $\text{AR}(p, \mathbb{C})$ is continuous in the parameter space $\tilde{\Theta}_1 = U \times \dots \times U$. It vanishes if and only if the denominator of (23) has multiple roots. Thus, the Jeffreys prior is strictly positive on the parameter space Θ_1 . Moreover, it diverges at the boundary $\partial\tilde{\Theta}_1$ of the parameter space $\tilde{\Theta}_1$ and defines an improper prior on $\tilde{\Theta}_1$.

IV. MAIN THEOREM

Let us consider a family $\{S_\theta\}_{\theta \in \Theta}$ of power spectral densities of complex-valued stationary processes, where $\Theta \subset \mathbb{C}^p$ is a complex parameter space. Our objective is to construct a predictive power spectral density $\hat{S}^{(N)}$, the risk $R(\hat{S}^{(N)})$ of which is kept as small as possible. We say that a predictive power spectral density $\hat{S}_1^{(N)}$ dominates a predictive power spectral density $\hat{S}_2^{(N)}$ if $R(\hat{S}_1^{(N)} | \theta) \leq R(\hat{S}_2^{(N)} | \theta)$ for any $\theta \in \Theta$ and the strict inequality holds for some θ .

Suppose that the parameter space Θ is Kähler and there exists a positive continuous eigenfunction ϕ of the Laplacian (Laplace–Beltrami operator; see (53) in Appendix A) $\Delta := 2g^{i\bar{j}}\partial_i\partial_{\bar{j}}$ with a negative eigenvalue $-K$ globally defined on Θ . We define a family of priors $\{\pi^{(\kappa)}\}_{\kappa \in \mathbb{R}}$, called κ -priors, as $\pi^{(\kappa)} := \phi^{-\kappa+1}\pi_J$, where π_J denotes the Jeffreys prior. We state that, with a suitable choice of $\kappa \in \mathbb{R}$, the Bayesian predictive power spectral density $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ based on the proposed prior $\pi^{(\kappa)}$ asymptotically dominates the Bayesian predictive power spectral density $\hat{S}_{\pi_J}^{(N)}$ based on the Jeffreys prior π_J . We have proved the following theorem, which is particularly for κ -priors for complex-valued ARMA processes.

Theorem 4.1 (Main Theorem): Let $\{S_\theta\}_{\theta \in \Theta}$ be a family of power spectral densities parameterized by a Kähler parameter

space $\Theta \subset \mathbb{C}^p$. We suppose that the inverse $[g^{i\bar{j}}]$ of the Fisher information matrix $[g_{i\bar{j}}]$ exists on Θ and, hence, the Laplacian Δ is properly defined on Θ . We also suppose that there exists a positive continuous eigenfunction ϕ of the Laplacian Δ with a negative eigenvalue $-K$ globally defined on Θ . Let $\pi_1 := \pi^{(\kappa_1)}$ and $\pi_2 := \pi^{(\kappa_2)}$ be two κ -priors for $\kappa_1, \kappa_2 \in \mathbb{R}$, and assume that Bayesian predictive power spectral densities $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ exist for $\kappa = \kappa_1, \kappa_2$. Further assume that there exist the asymptotic expansions (20) of the Bayesian predictive power spectral densities around the maximum likelihood estimator. Then, we have

$$\begin{aligned} R(\hat{S}_{\pi_1}^{(N)} | \theta) - R(\hat{S}_{\pi_2}^{(N)} | \theta) &= \frac{(\kappa_1 - \kappa_2)K + (\kappa_1 - \kappa_2)(\kappa_1 + \kappa_2)g^{i\bar{j}}(\partial_i \log \phi)(\partial_{\bar{j}} \log \phi)}{N^2} \\ &\quad + O(N^{-\frac{5}{2}}) \end{aligned}$$

for $\theta \in \Theta$.

The proof of the Main Theorem is largely aided by the form (20) of the asymptotic expansion of the Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$ around the maximum likelihood estimator $\hat{\theta}^{(N)}$. The comprehensive proof of the Main Theorem is given in Appendix H. Note that the specific construction of the Kähler parameter space (24) for $\text{AR}(p; \mathbb{C})$ is unrelated to this theorem. This theorem is a very general theorem that always holds as long as the parameter space Θ is Kähler. The importance of the Kähler property of parameter spaces in statistics and the generalization of the Main Theorem for the i.i.d. case are discussed in Section VI. The metric $g_{i\bar{j}}$ and its inverse $g^{i\bar{j}}$ of a Kähler manifold are explained in Appendix A.

Setting $\pi_1 := \pi^{(+1)} = \pi_J$ in Theorem 4.1, we can easily observe that if $-1 \leq \kappa < 1$, then $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ asymptotically dominates $\hat{S}_{\pi_J}^{(N)}$.

Corollary 4.1: Let $\pi := \pi^{(\kappa)}$ be a κ -prior on a Kähler parameter space $\Theta \subset \mathbb{C}^p$. We have

$$\begin{aligned} R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi^{(\kappa)}}^{(N)} | \theta) &= \frac{(1 - \kappa)K + (1 - \kappa^2)g^{i\bar{j}}(\partial_i \log \phi)(\partial_{\bar{j}} \log \phi)}{N^2} + O(N^{-\frac{5}{2}}). \end{aligned}$$

Therefore, if $-1 \leq \kappa < 1$, then $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ asymptotically dominates $\hat{S}_{\pi_J}^{(N)}$.

Recall that $g^{i\bar{j}}(\partial_i \log \phi)(\partial_{\bar{j}} \log \phi) \geq 0$ because the Hermitian matrix $[g^{i\bar{j}}]$ is positive definite. Therefore, to maximize the worst case of the risk improvement, we propose the κ -prior for $\kappa = -1$. When $\kappa = -1$, the Bayesian predictive power spectral density $\hat{S}_{\pi^{(-1)}}^{(N)}$ achieves constant risk improvement: $R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi^{(-1)}}^{(N)} | \theta) = 2K/N^2 + O(N^{-\frac{5}{2}})$. Equation (8) is a special case of Corollary 4.1 for $\text{AR}(p; \mathbb{C})$ and $\kappa = -1$, where $K = p(p+1)$. Section V discusses the existence of the positive continuous eigenfunction ϕ with a negative eigenvalue $-K = -p(p+1)$ on the parameter space Θ_1 for $\text{AR}(p; \mathbb{C})$.

V. SUPER-HARMONIC PRIORS ON $\text{AR}(p; \mathbb{C})$

In this section, we prove the existence of the positive continuous eigenfunction ϕ of the Laplacian Δ with eigenvalue $-K = -p(p+1)$ for $\text{AR}(p; \mathbb{C})$. Furthermore, for

the $\text{AR}(p; \mathbb{C})$ model, we show that the Bayesian predictive power spectral density $\hat{S}_{\pi^{(-1)}}^{(N)}$ based on the (-1) -prior $\pi^{(-1)}$ asymptotically dominates the estimative power spectral density $S_{\hat{\theta}^{(N)}}$ with the maximum likelihood estimator $\hat{\theta}^{(N)}$. This is another reason why we propose the (-1) -prior $\pi^{(-1)}$ for the case of $\text{AR}(p; \mathbb{C})$.

The eigenfunction ϕ for $\text{AR}(p; \mathbb{C})$ is defined as

$$\phi(\xi) := \prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j) \quad (27)$$

for $\xi = (\xi^1, \dots, \xi^p) \in \tilde{\Theta}_1 = U \times \dots \times U$. The function ϕ is the inverse of the determinant $|\Sigma^{(N)}|$ of the variance-covariance matrix $\Sigma^{(N)}$ of size $N \geq p$ for $\text{AR}(p; \mathbb{C})$; see Appendix G. The function ϕ is a real-valued continuous function defined globally on the parameter space $\tilde{\Theta}_1$. Moreover, it is positive on $\tilde{\Theta}_1$ and is 0 at the boundary $\partial\tilde{\Theta}_1$ of $\tilde{\Theta}_1$. Note also that the function ϕ has its maximum at the white noise process.

The κ -prior $\pi^{(\kappa)}$ for $\text{AR}(p; \mathbb{C})$ is

$$\begin{aligned} \pi^{(\kappa)} &:= \phi^{-\kappa+1} \pi_J \\ &= \left(\prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j) \right)^{-\kappa} \left(\prod_{1 \leq i < j \leq p} |\xi^i - \xi^j|^2 \right), \end{aligned} \quad (28)$$

where π_J is the Jeffreys prior (26) for $\text{AR}(p; \mathbb{C})$. The κ -prior $\pi^{(\kappa)}$ for $\text{AR}(p; \mathbb{C})$ is proper if $\kappa < 1$ and improper if $\kappa \geq 1$ on the parameter space $\tilde{\Theta}_1$; see Appendix G. In particular, the Jeffreys prior $\pi_J = \pi^{(+1)}$ is improper on the parameter space $\tilde{\Theta}_1$. The Bayesian predictive power spectral densities $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ for $\text{AR}(p; \mathbb{C})$ based on the κ -prior $\pi^{(\kappa)}$ exists if $\kappa < 2$ and $N \geq p$; see Appendix G.

Before proving $\Delta\phi = -p(p+1)\phi$, we introduce a useful lemma.

Lemma 5.1: We have

$$\partial_i \log \phi = \sum_{j=1}^p \frac{-\bar{\xi}^j}{1 - \xi^i \bar{\xi}^j} = -g_{i\bar{j}} \bar{\xi}^j, \quad (29)$$

$$\partial_{\bar{j}} \log \phi = \sum_{i=1}^p \frac{-\xi^i}{1 - \xi^i \bar{\xi}^j} = -g_{i\bar{j}} \xi^i, \quad (30)$$

and

$$\xi^i \partial_i \log \pi_J = \bar{\xi}^j \partial_{\bar{j}} \log \pi_J = \frac{1}{2} p(p-1) + \xi^i \bar{\xi}^j g_{i\bar{j}}. \quad (31)$$

Proof: To prove the third equation, use the identity $\sum_{i=1}^p \sum_{j=1, j \neq i}^p \frac{\xi^i}{\xi^i - \bar{\xi}^j} = \frac{1}{2} p(p-1)$. ■

Using Lemma 5.1, we see that ϕ is, in fact, an eigenfunction of the Laplacian with eigenvalue $-K = -p(p+1)$.

Proposition 5.1:

$$\Delta\phi = -p(p+1)\phi. \quad (32)$$

Proof: Because the parameter space is Kähler, we can use the formula (53) for its definition of the Laplacian. Direct

computation shows

$$\begin{aligned} \Delta\phi &= 2g^{i\bar{j}} \partial_i \partial_{\bar{j}} \phi = 2g^{i\bar{j}} \partial_i (-g_{k\bar{j}} \xi^k \phi) \\ &= 2g^{i\bar{j}} (-\partial_i g_{k\bar{j}}) \xi^k \phi - g_{k\bar{j}} (\partial_i \xi^k) \phi - g_{k\bar{j}} \xi^k (\partial_i \phi) \\ &= 2g^{i\bar{j}} (-\partial_k g_{i\bar{j}}) \xi^k \phi - g_{i\bar{j}} \phi + g_{k\bar{j}} g_{i\bar{l}} \xi^k \bar{\xi}^l \phi \\ &= -2 \xi^k (\partial_k \log \pi_J) \phi - 2g^{i\bar{j}} g_{i\bar{j}} \phi + 2g_{k\bar{l}} \xi^k \bar{\xi}^l \phi \\ &= -p(p-1)\phi - 2p\phi = -p(p+1)\phi, \end{aligned}$$

where we have used the Kählerness (52), the Jacobi formula (51), and $g^{i\bar{j}} g_{i\bar{j}} = p$. ■

As stated in Corollary 4.1, $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ asymptotically dominates $\hat{S}_{\pi^{(+1)}}^{(N)} = \hat{S}_{\pi_J}^{(N)}$ if $-1 \leq \kappa < 1$. For $\text{AR}(2; \mathbb{C})$, the specific prior $\psi := (1 - \xi^1 \bar{\xi}^2)(1 - \xi^2 \bar{\xi}^1)(1 - |\xi^1|^2)(1 - |\xi^2|^2)$ is introduced as a super-harmonic prior in [13]. This prior ψ is the special case of Corollary 4.1 for $(p, \kappa) = (2, 0)$. For $\text{AR}(p; \mathbb{R})$ with $p \geq 2$, a similar but slightly different prior is presented in [19]. This prior corresponds to the $\kappa = 0$ case for a positive eigenfunction of the Laplacian with eigenvalue $-K = -p(p-1)$.

We show that the Bayesian predictive power spectral density $\hat{S}_{\pi^{(-1)}}^{(N)}$ asymptotically dominates the estimative power spectral density $S_{\hat{\theta}^{(N)}}$ with the maximum likelihood estimator $\hat{\theta}^{(N)}$. Let us fix the true parameter $\theta_0 \in \Theta$ and denote the maximum likelihood estimator by $\hat{\theta}^{(N)} := \hat{\theta}^{(N)}(z^{(N)})$ for the observation $z^{(N)} \in \mathbb{C}^N$. According to [5], if we fortunately find a prior π such that $G_{\pi}^{(N)} = 0$, then we have

$$R(S_{\hat{\theta}^{(N)}} | \theta_0) - R(\hat{S}_{\pi}^{(N)} | \theta_0) = \frac{1}{2N^2} \|H^{(N)}\|_{\theta_0}^2 + O(N^{-\frac{5}{2}}). \quad (33)$$

For the specific parametrization $\xi \in \Theta_1 \subset \mathbb{C}^p$ for $\text{AR}(p; \mathbb{C})$ defined in (23), direct computation shows

$$G_{\pi^{(\kappa)}}^{(N)} = 2g^{i\bar{j}} \left(\partial_i \log \frac{\pi^{(\kappa)}}{\pi_J} + \frac{1}{2} T_i \right) \partial_{\bar{j}} S = -2(\kappa+1) \bar{\xi}^j \partial_{\bar{j}} S \quad (34)$$

for the κ -prior $\pi^{(\kappa)} := \phi^{-\kappa+1} \pi_J$. Thus, if $\kappa = -1$, then $G_{\pi}^{(N)} = 0$. Therefore, the Bayesian predictive power spectral density $\hat{S}_{\pi^{(-1)}}^{(N)}$ asymptotically dominates the estimative power spectral density $S_{\hat{\theta}^{(N)}}$ with the maximum likelihood estimator $\hat{\theta}^{(N)}$.

VI. GENERALIZATION OF THE MAIN THEOREM

Although the present study mainly focuses on the complex Gaussian process, the Main Theorem is valid for the i.i.d. case as long as the complex parameter space is Kähler.

Kählerness is a complexified concept of the exponential family. Consider a family of probability density functions $\{p_{\theta}\}_{\theta \in \Theta}$ of the exponential family parameterized by real parameters $\theta = (\theta^1, \dots, \theta^p) \in \Theta \subset \mathbb{R}^p$. We may assume that the probability density function is of the form $p_{\theta}(x) = \exp(\theta^i x_i - \Psi(\theta))$. We know that the Fisher information matrix is given by $g_{ij} = \partial_i \partial_j \Psi$. The Kähler parameter space Θ is the generalization of the exponential family in the sense that there exists, at least locally, a function \mathcal{K} on Θ , called a Kähler potential, such that $g_{i\bar{j}} = \partial_i \partial_{\bar{j}} \mathcal{K}$. If there exists a Kähler

potential \mathcal{K} on Θ , then it is easy to see that the definition of Kählerness (52) holds. The converse is also true; see [13], [20].

We define the risk of a Bayesian predictive distribution for the i.i.d. case. Consider a family of probability density functions $\{p_\theta\}_{\theta \in \Theta}$ parameterized by complex parameters $\theta = (\theta^1, \dots, \theta^p) \in \Theta \subset \mathbb{C}^p$, where Θ is Kähler. The sample space \mathcal{Z} of this model may be any subset of \mathbb{R}^r or \mathbb{C}^r . Let π be a possibly improper prior for this model. Consider an i.i.d. sample $z^{(N)} = (z_1, \dots, z_N) \in \mathcal{Z}^N$ of size N from the distribution at $\theta \in \Theta$. The predictive distribution $\hat{p}_\pi^{(N)}(z) := \int_\Theta p_\theta(z) \pi(\theta | z^{(N)}) d\theta$ for $z \in \mathcal{Z}$ is called the Bayesian predictive distribution based on the prior π . The risk of $\hat{p}_\pi^{(N)}$ is defined as

$$\begin{aligned} R(\hat{p}_\pi^{(N)} | \theta) &:= E_\theta \left[D_{\text{KL}} \left(p_\theta \parallel \hat{p}_\pi^{(N)} \right) \right] \\ &= \int_{\mathcal{Z}^N} D_{\text{KL}} \left(p_\theta \parallel \hat{p}_\pi^{(N)} \right) dP_\theta^{(N)}(z^{(N)}) \\ &= \int_{\mathcal{Z}^N} \left(\int_{\mathcal{Z}} p_\theta(z) \log \frac{p_\theta(z)}{\hat{p}_\pi^{(N)}(z)} dz \right) dP_\theta^{(N)}(z^{(N)}), \end{aligned}$$

where $dP_\theta^{(N)}(z^{(N)}) := (\prod_{t=1}^N p_\theta(z_t) dz_t)$.

We prepare the relevant mathematical notation. We set

$$\begin{aligned} K_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c} \\ := \int_{\mathcal{Z}} (D_{\alpha_1 \dots \alpha_a} \log p(z)) (D_{\beta_1 \dots \beta_b} \log p(z)) \\ \dots (D_{\gamma_1 \dots \gamma_c} \log p(z)) P(dz) \end{aligned} \quad (35)$$

for a probability density function $p = p_\theta$ and indices $\alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, \dots, \gamma_1, \dots, \gamma_c \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, where $P(dz) := p(z) dz$. We define the quantities $g_{\alpha\beta}$, $T_{\alpha\beta\gamma}$, and $\Gamma_{\alpha\beta\gamma}^{(m)}$ as

$$\begin{aligned} g_{\alpha\beta} &:= K_{\alpha,\beta} \\ &= E [(\partial_\alpha \log p)(\partial_\beta \log p)], \end{aligned} \quad (36)$$

$$\begin{aligned} T_{\alpha\beta\gamma} &:= K_{\alpha,\beta,\gamma} \\ &= E [(\partial_\alpha \log p)(\partial_\beta \log p)(\partial_\gamma \log p)], \end{aligned} \quad (37)$$

$$\begin{aligned} \Gamma_{\alpha\beta\gamma}^{(m)} &:= K_{\alpha\beta,\gamma} + K_{\alpha,\beta\gamma} \\ &= E [(\partial_\alpha \partial_\beta p)(\partial_\gamma p)], \end{aligned} \quad (38)$$

respectively, for $\alpha, \beta, \gamma \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. The relation to the former definitions of the quantities $g_{\alpha\beta}$, $T_{\alpha\beta\gamma}$, and $\Gamma_{\alpha\beta\gamma}^{(m)}$, namely (16), (17), and (18), is given in Proposition 4.2 in Appendix D; they coincide with each other up to $O(1)$ terms for complex-valued stationary ARMA processes.

We have the same form of the asymptotic expansion of $\hat{p}_\pi^{(N)}$ as (20) for the i.i.d. case; see [21]. Consider a positive continuous eigenfunction ϕ of the Laplacian Δ with a negative eigenvalue $-K$, i.e., $\Delta\phi = -K\phi < 0$. Then, we can construct the κ -prior by $\pi^{(\kappa)} := \phi^{-\kappa+1} \pi_J$ for $\kappa \in \mathbb{R}$, where π_J is the Jeffreys prior of this model. As the proof of the Main Theorem (Proposition 4.1) only depends on the form (20) of the asymptotic expansion of the Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$, Theorem 4.1 also holds for the risk difference $R(\hat{p}_{\pi_1}^{(N)} | \theta) - R(\hat{p}_{\pi_2}^{(N)} | \theta)$ for the i.i.d. case.

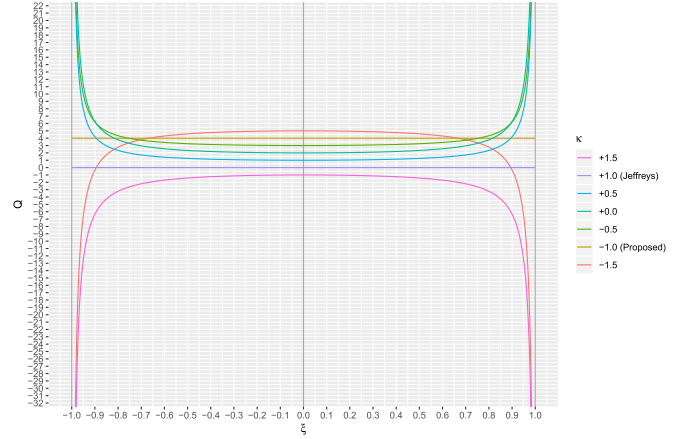


Fig. 1. Expected pointwise limit $Q^{(\kappa)}(\xi)$ for the risk difference $Z^{(\kappa)}(\xi)$ for $\text{AR}(1; \mathbb{C})$, where $\xi \in (-1, 1) \subset U$.

Therefore, our proposal is to use the prior $\pi^{(-1)} := \phi^{-2} \pi_J$, where ϕ is an eigenfunction of the Laplacian with the smallest negative eigenvalue.

The construction $\pi^{(\kappa)} := \phi^{-\kappa+1} \pi_J$ of a family of κ -priors is related to other types of objective priors. For example, if $\partial_i \phi$ is proportional to $T_i := T_{ik\bar{j}} g^{k\bar{j}} + T_{i\bar{j}k} g^{k\bar{j}}$ on the parameter space, the family $\{\pi^{(\kappa)}\}_{\kappa \in \mathbb{R}}$ of κ -priors is, in fact, a family of α -parallel priors; see Appendix I. In general, α -parallel priors do not always exist. Therefore, the existence of a family of α -parallel priors suggests some statistical property in the statistical model; see [14].

VII. NUMERICAL EXPERIMENTS FOR RISK DIFFERENCES

We consider the $\text{AR}(1; \mathbb{C})$ case of $z^t = \xi z^{t-1} + \varepsilon$ in this section. The parameter space is the open unit disk $U = \{\xi \in \mathbb{C} \mid |\xi| < 1\}$. The κ -prior for $\text{AR}(1; \mathbb{C})$ is $\pi^{(\kappa)}(\xi) := (1 - |\xi|^2)^{-\kappa}$, where $\xi \in U$. Recall that $\kappa = +1$ corresponds to the improper Jeffreys prior $\pi_J = (1 - |\xi|^2)^{-1}$, which is mentioned as a reference prior in [7] for $\text{AR}(1, \mathbb{R})$. On the other hand, $\kappa = -1$ corresponds to the proposed proper prior $\pi^{(-1)} = (1 - |\xi|^2)$, which is the inverse of the Jeffreys prior, and is also mentioned in [8] for the $\text{AR}(1; \mathbb{R})$ case. In fact, the inverse of the Jeffreys prior for $\text{AR}(1; \mathbb{R})$ is a maximal data information prior (MDIP) for $\text{AR}(1; \mathbb{R})$; see [8]. Note that for $p \geq 2$, the proposed proper prior $\pi^{(-1)}$ is not the inverse of the Jeffreys prior $\pi^{(+1)}$ for $\text{AR}(p; \mathbb{C})$.

Corollary 4.1 for $\text{AR}(1; \mathbb{C})$ now becomes

$$R(\hat{S}_{\pi_J}^{(N)} | \xi) - R(\hat{S}_{\pi^{(\kappa)}}^{(N)} | \xi) = \frac{1}{N^2} Q^{(\kappa)}(\xi) + O(N^{-\frac{5}{2}}), \quad (39)$$

where $Q^{(\kappa)}(\xi) := 2(1 - \kappa) + (1 - \kappa^2) \frac{|\xi|^2}{1 - |\xi|^2}$ is the expected pointwise limit of the normalized risk difference $Z^{(\kappa)}(\xi) := N^2 \left(R(\hat{S}_{\pi_J}^{(N)} | \xi) - R(\hat{S}_{\pi^{(\kappa)}}^{(N)} | \xi) \right)$. Figure 1 shows the function $Q^{(\kappa)}(\xi)$. As stated in Corollary 4.1, if $-1 \leq \kappa \leq 1$, then $Q^{(\kappa)}(\xi) \geq 0$. In particular, if $\kappa = -1$, then the risk difference $Z^{(\kappa)}(\xi)$ asymptotically achieves the constant $Q^{(-1)}(\xi) = 2K = 4$.

The results of numerical experiments for $Z^{(\kappa)}(\xi)$ with $N = 30$, $N = 60$, and $N = 120$ are shown in Figure 2, 3, and 4,

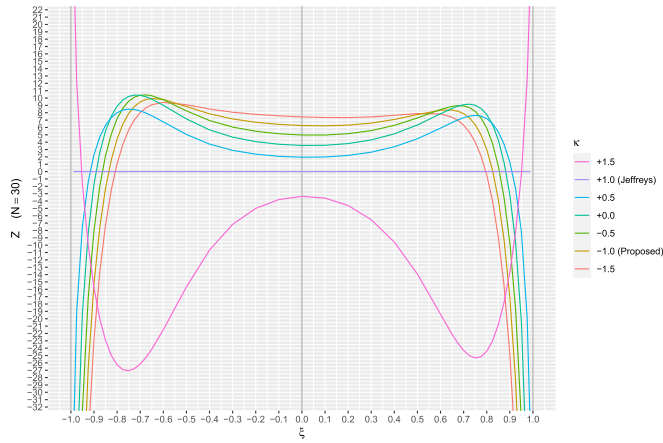


Fig. 2. Numerical experiment for the risk difference $Z^{(\kappa)}(\xi)$ for AR(1; \mathbb{C}) with $N = 30$, where $\xi \in (-1, 1) \subset U$. The proposed $\hat{S}_{\pi^{(-1)}}^{(N)}$ dominates the baseline $\hat{S}_{\pi_J}^{(N)}$ for $\xi \in [-0.825, +0.825]$.

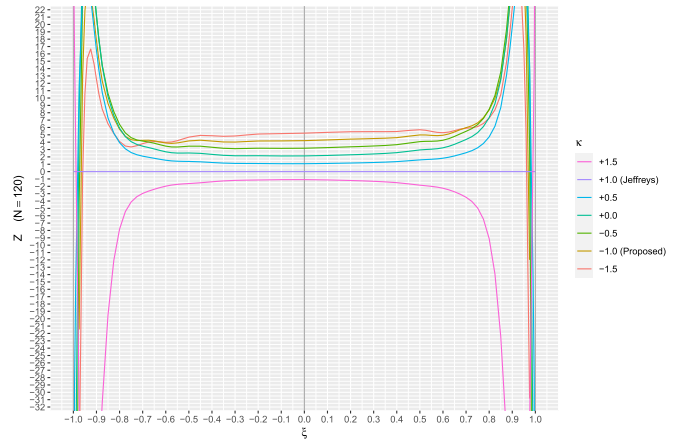


Fig. 4. Numerical experiment for the risk difference $Z^{(\kappa)}(\xi)$ for AR(1; \mathbb{C}) with $N = 120$, where $\xi \in (-1, 1) \subset U$. The proposed $\hat{S}_{\pi^{(-1)}}^{(N)}$ dominates the baseline $\hat{S}_{\pi_J}^{(N)}$ for $\xi \in [-0.960, +0.960]$.

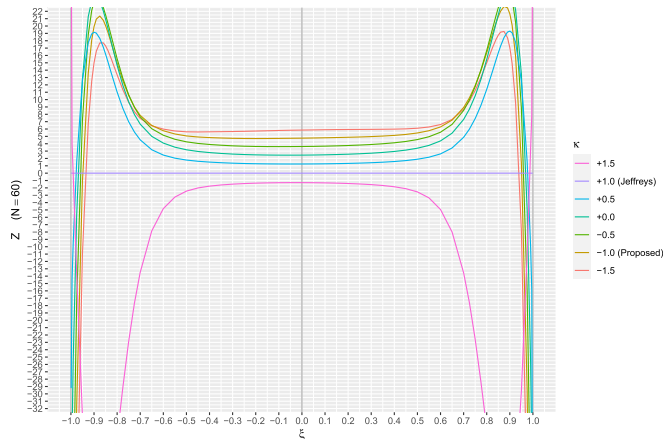


Fig. 3. Numerical experiment for the risk difference $Z^{(\kappa)}(\xi)$ for AR(1; \mathbb{C}) with $N = 60$, where $\xi \in (-1, 1) \subset U$. The proposed $\hat{S}_{\pi^{(-1)}}^{(N)}$ dominates the baseline $\hat{S}_{\pi_J}^{(N)}$ for $\xi \in [-0.925, +0.925]$.

respectively, where the Monte Carlo method is used for evaluating the value of (4). These Figures show that $Z^{(\kappa)}(\xi)$ asymptotically achieves $Q^{(\kappa)}(\xi)$, but the rate of convergence may depend on $\xi \in U$. It appears that the convergence is not uniform on U .

For simplicity, we intentionally omitted mathematical details for some notations in the preceding sections. In Appendix A, we introduce complex differential analysis and define the Laplacian as the differential operator on a Kähler parameter space. In Appendix B, we derive and justify the risk for a predictive power spectral density for a complex-valued Gaussian process. In Appendix C, we present the complexified version of tensorial Hermite polynomials. In Appendix D, we relate the theory of power spectral densities to the theory of probability densities by utilizing the asymptotic expansions of the expectation of the derivatives of the log likelihood. In Appendix E, we present the asymptotic expansion of the risk of the estimative power spectral density with the maximum likelihood estimator. In Appendix F, we provide the asymptotic

expansion of the Bayesian predictive power spectral density around the maximum likelihood estimator. In Appendix G, we prove the existence of Bayesian predictive power spectral densities for AR(p ; \mathbb{C}). In Appendix H, we prove the Main Theorem. In Appendix I, we show the relation between κ -priors and α -parallel priors.

APPENDIX A WIRTINGER CALCULUS

In this appendix, we introduce an elegant equivalent formulation of usual differential calculus, called Wirtinger calculus or $\mathbb{C}\mathbb{R}$ calculus [2]. Let us consider a complex-valued function f defined on \mathbb{C}^p . A function defined on the domain \mathbb{C}^p is always regarded as one defined on the domain \mathbb{R}^{2p} . For the i -th complex coordinate $z^i = x^i + \sqrt{-1}y^i$ in \mathbb{C}^p , the Wirtinger derivatives ∂_i and $\bar{\partial}_i$ are defined as linear partial differential operators of the first order:

$$\partial_i := \frac{1}{2} \left(\frac{\partial}{\partial x^i} - \sqrt{-1} \frac{\partial}{\partial y^i} \right), \quad \bar{\partial}_i := \frac{1}{2} \left(\frac{\partial}{\partial x^i} + \sqrt{-1} \frac{\partial}{\partial y^i} \right), \quad (40)$$

where $\frac{\partial}{\partial x^i}$ and $\frac{\partial}{\partial y^i}$ denote the usual partial differential operators on \mathbb{R}^{2p} . The symbol $\bar{\partial}_i$ is occasionally expressed as $\bar{\partial}_{\bar{i}}$.

We should mention that although the variables z^i and \bar{z}^i are not independent, the derivatives ∂_i and $\bar{\partial}_i$ are independent differential operators in the complexified tangent space of $\mathbb{C}^p = \mathbb{R}^{2p}$. In fact, direct computation shows that the set $\{\partial_1, \dots, \partial_p, \bar{\partial}_1, \dots, \bar{\partial}_p\}$ forms a basis of the complexified tangent space of $\mathbb{C}^p = \mathbb{R}^{2p}$.

The Wirtinger derivatives are not the partial derivatives in usual differential calculus; however, Wirtinger calculus inherits most of the properties of usual differential calculus.

The most fascinating property inherited by Wirtinger calculus from usual differential calculus is its chain rule property:

$$\partial_i(f \circ g) = \sum_{j=1}^q (\partial_i g^j)(\partial_j f \circ g) + \sum_{\bar{j}=1}^q (\partial_i \bar{g}^{\bar{j}})(\bar{\partial}_{\bar{j}} f \circ g), \quad (41)$$

$$\partial_i(f \circ g) = \sum_{j=1}^q (\partial_i g^j) (\partial_j f \circ g) + \sum_{j=1}^q (\partial_i \bar{g}^j) (\partial_j f \circ g) \quad (42)$$

for $i = 1, \dots, p$, where $f: \mathbb{C}^q \rightarrow \mathbb{C}$ and $g = (g^1, \dots, g^q): \mathbb{C}^p \rightarrow \mathbb{C}^q$.

Another important property of Wirtinger calculus is its summation rule. For a complex vector $\lambda = (\lambda^1, \dots, \lambda^p) \in \mathbb{C}^p$ and a complex-valued function f defined on \mathbb{C}^p ,

$$\sum_{i=1}^p \lambda^i \partial_i f + \sum_{j=1}^p \bar{\lambda}^j \partial_j f = \sum_{k=1}^p \Re(\lambda^k) \frac{\partial f}{\partial x^k} + \sum_{k=1}^p \Im(\lambda^k) \frac{\partial f}{\partial y^k}, \quad (43)$$

where the function f is regarded as a function defined on \mathbb{R}^{2p} on the right-hand side of the equation.

As the Einstein notation is assumed throughout this paper, the summation is automatically taken over those indices that appear exactly twice, once as a superscript and once as a subscript. Therefore, when the Einstein notation is used, the left-hand side of (43) is denoted by $\lambda^\alpha \partial_\alpha f$, if α runs through the indices $\{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, or sometimes by $\lambda^i \partial_i f + \bar{\lambda}^j \partial_j f$, if i, j run through the indices $\{1, \dots, p\}$, where $\bar{\lambda}^j$ represents the complex conjugate of λ^j . In this paper, we attempt to use the symbols $\alpha, \beta, \gamma, \dots$ when they run through the indices $\{1, \dots, p, \bar{1}, \dots, \bar{p}\}$ and to use the symbols i, j, k, \dots when they run through the indices $\{1, \dots, p\}$.

Consider a positive definite metric $g_{\alpha\beta}$ on \mathbb{C}^p , i.e., $g_{\alpha\beta} = g_{\beta\alpha}$ for any $\alpha, \beta \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, and

$$g_{\alpha\beta} \lambda^\alpha \lambda^\beta = g_{ij} \lambda^i \lambda^j + g_{i\bar{j}} \lambda^i \bar{\lambda}^j + g_{i\bar{j}} \bar{\lambda}^i \lambda^j + g_{\bar{i}\bar{j}} \bar{\lambda}^i \bar{\lambda}^j > 0 \quad (44)$$

for any $\lambda \in \mathbb{C}^p \setminus \{0\}$. Let us denote by $[g^{\alpha\beta}]$ the inverse matrix of the $2p \times 2p$ matrix $[g_{\alpha\beta}]$, i.e., $g_{\alpha\gamma} g^{\gamma\beta} = \delta_\alpha^\beta$ for the Kronecker delta δ_α^β .

For the i -th complex coordinate $z^i = x^i + \sqrt{-1} y^i$ in \mathbb{C}^p , the derivatives ∂^i and $\bar{\partial}^i$ are defined by

$$\partial^i := g^{ij} \partial_j + g^{i\bar{j}} \partial_{\bar{j}}, \quad (45)$$

$$\bar{\partial}^i := g^{\bar{i}j} \partial_j + g^{\bar{i}\bar{j}} \partial_{\bar{j}}. \quad (46)$$

The derivatives ∂^i and $\bar{\partial}^i$ are also simply defined by $\partial^\alpha = g^{\alpha\beta} \partial_\beta$ for $\alpha \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. The symbol $\bar{\partial}^i$ is sometimes denoted by $\bar{\partial}^i$. For later use, let us define the differential operators $D_{\alpha_1 \dots \alpha_a}$ and $D^{\alpha_1 \dots \alpha_a}$ by

$$D_{\alpha_1 \dots \alpha_a} := \partial_{\alpha_1} \dots \partial_{\alpha_a}, \quad (47)$$

$$D^{\alpha_1 \dots \alpha_a} := g^{\alpha_1 \beta_1} \dots g^{\alpha_a \beta_a} \partial_{\beta_1} \dots \partial_{\beta_a}, \quad (48)$$

respectively, for $\alpha_1, \dots, \alpha_a \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. In particular, $D_\alpha = \partial_\alpha$ and $D^\alpha = \bar{\partial}^\alpha$.

A metric $g_{\alpha\beta}$ is called Hermitian, if

$$g_{ij} = g_{i\bar{j}} = 0, \quad g_{i\bar{j}} = g_{\bar{j}i} = \overline{g_{j\bar{i}}} = \overline{g_{\bar{i}j}} \quad (49)$$

for all $i, j = 1, \dots, p$; see Section 8.4 in [20]. If the metric $g_{\alpha\beta}$ is Hermitian, the square of the distance of the infinitesimal complex vector ds is given by

$$ds^2 = g_{\alpha\beta} ds^\alpha ds^\beta = 2 g_{i\bar{j}} ds^i ds^{\bar{j}}. \quad (50)$$

If the metric $g_{\alpha\beta}$ is Hermitian, we only need to consider one-fourth of the $2p \times 2p$ complex-valued matrix $[g_{\alpha\beta}]$, namely the $p \times p$ Hermitian matrix $[g_{i\bar{j}}]$. A complex manifold with a Hermitian metric is called a Hermitian manifold. The Jacobi formula for the Hermitian manifold is

$$g^{i\bar{j}} \partial_k g_{i\bar{j}} = \partial_k \log \pi_J, \quad g^{i\bar{j}} \bar{\partial}_k g_{i\bar{j}} = \bar{\partial}_k \log \pi_J, \quad (51)$$

where π_J is the determinant of the $p \times p$ Hermitian matrix $[g_{i\bar{j}}]$, i.e., the square root of the determinant of the $2p \times 2p$ matrix $[g_{\alpha\beta}]$; see Section 8.4 in [20].

A Hermitian manifold with a metric $g_{i\bar{j}}$ is called a Kähler manifold, if

$$\partial_i g_{j\bar{k}} = \partial_j g_{i\bar{k}}, \quad \partial_i g_{j\bar{k}} = \partial_{\bar{k}} g_{j\bar{i}} \quad (52)$$

for all $i, j, k = 1, \dots, p$; see Section 8.5 in [20].

The Laplacian (Laplace–Beltrami operator) on a Kähler manifold is

$$\Delta = \partial^\alpha \partial_\alpha = g^{\alpha\beta} \partial_\alpha \partial_\beta = 2 g^{i\bar{j}} \partial_i \bar{\partial}_j, \quad (53)$$

which does not hold in general for the usual Riemannian manifold. From (53), we have

$$\frac{\Delta \phi^a}{\phi^a} = a \frac{\Delta \phi}{\phi} + 2a(a-1) g^{i\bar{j}} (\partial_i \log \phi) (\bar{\partial}_j \log \phi) \quad (54)$$

for $a \in \mathbb{R}$, which is a useful formula for calculating $\Delta \phi^a$.

APPENDIX B

KULLBACK–LEIBLER DIVERGENCE BETWEEN POWER SPECTRAL DENSITIES

This appendix derives and justifies the form (2) of risk $R(\hat{S} | \theta)$ for a predictive power spectral density \hat{S} for a complex-valued Gaussian process, and it explains why the Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$ minimizes the Bayes risk (7) given an observation $z^{(N)} \in \mathbb{C}^N$ and a prior π .

Let $W^{(N)}$ be a circulant matrix and $D^{(N)}$ be a diagonal matrix defined by

$$W^{(N)} := \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix},$$

$$D^{(N)} := \begin{pmatrix} e^{2\pi i \frac{1}{N}} & 0 & \dots & 0 \\ 0 & e^{2\pi i \frac{2}{N}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{2\pi i \frac{N}{N}} \end{pmatrix},$$

respectively. We have a relation $W^{(N)} U^{(N)} = U^{(N)} D^{(N)}$, where $U^{(N)} := [U_{st}^{(N)}]$ is a unitary matrix defined by $U_{st}^{(N)} := \frac{1}{\sqrt{N}} e^{2\pi i \frac{st}{N}}$. Set $\Lambda(z) := \sum_{h=-\infty}^{\infty} \gamma_h z^{-h}$, where z is a formal variable. Note that $S(\omega) = \frac{1}{2\pi} \Lambda(e^{i\omega})$ and $\Lambda(W^{(N)}) = U^{(N)} \Lambda(D^{(N)}) U^{(N)*}$. Suppose $\Lambda(z)$ has a Laurent expansion on $|z| = 1$. Then, $(\Lambda(z))^{-1} = 1/\Lambda(z)$ is defined on the neighborhood of $|z| = 1$. If the autocovariance γ_t decreases exponentially, we have $\Lambda(W^{(N)}) \approx \Sigma^{(N)}$ for

large N because the (s, t) -th element of the matrix $\Lambda(W^{(N)})$ is approximated as

$$[\Lambda(W^{(N)})]_{st} = \sum_{k=-\infty}^{\infty} \gamma_{(s-t)+kN} \approx \gamma_{(s-t)} = [\Sigma^{(N)}]_{st}. \quad (55)$$

With this approximation, the log likelihood

$$l^{(N)}(z^{(N)}) = -N \log \pi - \log |\Sigma^{(N)}| - z^* (\Sigma^{(N)})^{-1} z \quad (56)$$

of the observation $z = z^{(N)}$ from a complex-valued Gaussian process with mean $0 \in \mathbb{C}$ is approximated as

$$\begin{aligned} l^{(N)}(z^{(N)}) &\approx -N \log \pi - \log |\Lambda(W)| - z^* (\Lambda(W))^{-1} z \\ &= NC - \sum_{n=1}^N \log S \left(2\pi \frac{n}{N} \right) - \sum_{n=1}^N \frac{I \left(2\pi \frac{n}{N} \right)}{S \left(2\pi \frac{n}{N} \right)}, \end{aligned} \quad (57)$$

where I denotes the empirical power spectral density (periodogram) defined by $I \left(2\pi \frac{n}{N} \right) := \frac{1}{2\pi} |\tilde{z}_n|^2$ with $\tilde{z}_n := (U^* z)_n = \frac{1}{\sqrt{N}} \sum_{s=1}^N e^{-2\pi i s \frac{n}{N}} z_s$ and C is a constant independent of N and S . For a more detailed explanation of (57), see [15] for real-valued stationary processes and [16] for real-valued ARMA processes.

Suppose the variance-covariance matrix (11) is parameterized by complex parameters $\theta \in \Theta \subset \mathbb{C}^p$, i.e., the autocovariances $\{\gamma_h\}_{h \in \mathbb{Z}}$ are parametrized by $\theta \in \Theta$. Its power spectral density (1) is denoted by $S_\theta(\omega)$ or $S(\omega | \theta)$ for $\theta \in \Theta$. For $\theta \in \Theta$, we denote the corresponding probability distribution, probability density function, and log likelihood of the observation $z^{(N)} \in \mathbb{C}^N$ by $P_\theta^{(N)}$, $p_\theta^{(N)}$, and $l_\theta^{(N)}$, respectively.

For $\theta_1, \theta_2 \in \Theta$, the KL-divergence $D_{\text{KL}}(P_{\theta_1} \| P_{\theta_2})$ of the distributions P_{θ_2} from the distribution P_{θ_1} is approximated as

$$\begin{aligned} D_{\text{KL}}(P_{\theta_1} \| P_{\theta_2}) &= \int_{\mathbb{C}^p} \left(l_{\theta_1}^{(N)}(z^{(N)}) - l_{\theta_2}^{(N)}(z^{(N)}) \right) dP_{\theta_1}(z^{(N)}) \\ &\approx E_{\theta_1} \left[- \sum_{n=1}^N \log \frac{S_{\theta_1} \left(2\pi \frac{n}{N} \right)}{S_{\theta_2} \left(2\pi \frac{n}{N} \right)} \right. \\ &\quad \left. - \sum_{n=1}^N \left(\frac{I \left(2\pi \frac{n}{N} \right)}{S_{\theta_1} \left(2\pi \frac{n}{N} \right)} - \frac{I \left(2\pi \frac{n}{N} \right)}{S_{\theta_2} \left(2\pi \frac{n}{N} \right)} \right) \right] \\ &\approx \frac{N}{2\pi} \int_{-\pi}^{\pi} \left(- \log \frac{S_{\theta_1}(\omega)}{S_{\theta_2}(\omega)} - 1 + \frac{S_{\theta_1}(\omega)}{S_{\theta_2}(\omega)} \right) d\omega \\ &= ND_{\text{KL}}(S_{\theta_1} \| S_{\theta_2}), \end{aligned}$$

where $D_{\text{KL}}(S_{\theta_1} \| S_{\theta_2})$ is the KL-divergence (3) between power spectral densities, which has previously been discussed in the literature [13], [18] in the context of signal processing. On the other hand, in the literature [5], [17] on real-valued processes, 4π instead of 2π is used in the denominator of (3). However, as we have explained, the constant in the denominator of (3) for the complex-valued process should be 2π . Note also that for any power spectral densities S_1 and S_2 , because $-\log x - 1 + x \geq 0$ for any $x \geq 0$, we have $D_{\text{KL}}(S_1 \| S_2) \geq 0$ in general, and $D_{\text{KL}}(S_1 \| S_2) = 0$ if and

only if $S_1(\omega) = S_2(\omega)$ for $\omega \in [-\pi, \pi]$ almost everywhere. The asymptotic expansion

$$\begin{aligned} - \log \frac{1}{1+x} - 1 + \frac{1}{1+x} \\ = \frac{1}{2}x^2 - \frac{2}{3}x^3 + \frac{3}{4}x^4 - \frac{4}{5}x^5 + O(x^6) \end{aligned} \quad (58)$$

is a useful formula for calculating the value of $D_{\text{KL}}(S \| S + dS)$.

For a possibly improper prior π , the Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$ minimizes the Bayes risk (7) among all the predictive power spectral densities $\hat{S}^{(N)}$ if $r(\hat{S}_\pi^{(N)} | \pi) < +\infty$; see [22]. In fact,

$$\begin{aligned} r(\hat{S}^{(N)} | \pi) - r(\hat{S}_\pi^{(N)} | \pi) &= \int_{\Theta} \int_{\mathbb{C}^p} \left(D_{\text{KL}}(S_\theta \| \hat{S}^{(N)}) - D_{\text{KL}}(S_\theta \| \hat{S}_\pi^{(N)}) \right) \\ &\quad dP_\theta^{(N)} d\Theta \\ &= \int_{\Theta} \int_{\mathbb{C}^p} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(- \log \frac{\hat{S}_\pi^{(N)}}{\hat{S}^{(N)}} - \frac{S_\theta}{\hat{S}_\pi^{(N)}} + \frac{S_\theta}{\hat{S}^{(N)}} \right) d\omega \right) \\ &\quad dP_\theta^{(N)} d\Theta \\ &= \int_{\mathbb{C}^p} D_{\text{KL}}(\hat{S}_\pi^{(N)} \| \hat{S}^{(N)}) m_\pi^{(N)}(z^{(N)}) dz^{(N)} \geq 0 \end{aligned}$$

for any predictive power spectral density $\hat{S}^{(N)}$, where $dP_\theta^{(N)} := p_\theta^{(N)}(z^{(N)}) dz^{(N)}$, $d\Theta := \pi(\theta) d\theta$, and $m_\pi^{(N)}(z^{(N)}) := \int_{\Theta} p_\theta^{(N)}(z^{(N)}) \pi(\theta) d\theta$ is the marginal distribution of the observation $z^{(N)} \in \mathbb{C}^N$ based on the prior π .

APPENDIX C TENSORIAL HERMITE POLYNOMIALS

Tensorial Hermite polynomials, as introduced in [23], are very useful tools for calculating Edgeworth expansions. Here, we present the complexified version of tensorial Hermite polynomials to calculate (78).

Consider a metric $g_{\alpha\beta}$ on \mathbb{C}^p . We define a complex-valued function ϕ on \mathbb{C}^p by $\phi(\lambda) := \frac{1}{G} e^{-\frac{1}{2} g_{\alpha\beta} \lambda^\alpha \lambda^\beta}$ for $\lambda = (\lambda^1, \dots, \lambda^p) \in \mathbb{C}^p$, where α, β run through the indices $\{1, \dots, p, \bar{1}, \dots, \bar{p}\}$ and $G := \int_{\mathbb{C}^p} e^{-\frac{1}{2} g_{\alpha\beta} \lambda^\alpha \lambda^\beta} d\lambda$ is the normalization factor that gives $\int_{\mathbb{C}^p} \phi(\lambda) d\lambda = 1$. We assume that the metric $g_{\alpha\beta}$ is positive definite so that $\lim_{|\lambda| \rightarrow \infty} |\phi(\lambda)| = 0$. If the metric $g_{\alpha\beta}$ is Hermitian, the normalization factor G reduces to the product of π^p and the determinant of the $p \times p$ Hermitian matrix $[g^{i\bar{j}}]$. However, to obtain a general result, we do not assume that the metric $g_{\alpha\beta}$ is Hermitian in this appendix.

We define the complex-valued tensorial Hermite polynomial $h^{\alpha_1 \dots \alpha_a}$ for $\alpha_1, \dots, \alpha_a \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$ by the identity

$$(-1)^k D^{\alpha_1 \dots \alpha_a} \phi(\lambda) = h^{\alpha_1 \dots \alpha_a}(\lambda) \phi(\lambda), \quad (59)$$

where the differential operator $D^{\alpha_1 \dots \alpha_a}$ is defined by (48). For example, $h^{\alpha_1}(\lambda) = \lambda^{\alpha_1}$, and $h^{\alpha_1 \alpha_2}(\lambda) = \lambda^{\alpha_1} \lambda^{\alpha_2} - g^{\alpha_1 \alpha_2}$.

Following a procedure similar to that in [23], we obtain

$$\int_{\mathbb{C}^p} h^{\alpha_1 \cdots \alpha_a}(\lambda) h^{\beta_1 \cdots \beta_b}(\lambda) \phi(\lambda) d\lambda = \begin{cases} a! g^{\alpha_1 \beta_1} \cdots g^{\alpha_a \beta_a} & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$$

for $\alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, where the parentheses around the indices $\alpha_1 \beta_1 \cdots \alpha_a \beta_a$ imply the symmetrization of the indices $\alpha_1, \dots, \alpha_a$; i.e., $(\alpha_1 \beta_1 \cdots \alpha_a \beta_a) := (1/a!) \sum_{\sigma} \alpha_{\sigma(1)} \beta_1 \cdots \alpha_{\sigma(a)} \beta_a$, where σ runs through all the permutations of the indices $\alpha_1, \dots, \alpha_a$. For example,

$$\int_{\mathbb{C}^p} \lambda^{\alpha_1} \lambda^{\alpha_2} \phi(\lambda) d\lambda = \int_{\mathbb{C}^p} (h^{\alpha_1 \alpha_2} + g^{\alpha_1 \alpha_2}) \phi(\lambda) d\lambda = g^{\alpha_1 \alpha_2}, \quad (60)$$

and

$$\begin{aligned} & \int_{\mathbb{C}^p} \lambda^{\alpha_1} \lambda^{\alpha_2} \lambda^{\alpha_3} \lambda^{\alpha_4} \phi(\lambda) d\lambda \\ &= \int_{\mathbb{C}^p} (h^{\alpha_1 \alpha_2} h^{\beta_1 \beta_2} + \lambda^{\alpha_1} \lambda^{\alpha_2} g^{\beta_1 \beta_2} + g^{\alpha_1 \alpha_2} \lambda^{\beta_1} \lambda^{\beta_2} \\ & \quad - g^{\alpha_1 \alpha_2} g^{\beta_1 \beta_2}) \phi(\lambda) d\lambda \\ &= g^{\alpha_1 \beta_1} g^{\alpha_2 \beta_2} + g^{\alpha_2 \beta_1} g^{\alpha_1 \beta_2} + g^{\alpha_1 \alpha_2} g^{\beta_1 \beta_2} \end{aligned} \quad (61)$$

because $\int h^{\alpha_1 \alpha_2}(\lambda) h^{\beta_1 \beta_2}(\lambda) \phi(\lambda) d\lambda = 2g^{(\alpha_1 \beta_1) (\alpha_2 \beta_2)} = g^{\alpha_1 \beta_1} g^{\alpha_2 \beta_2} + g^{\alpha_2 \beta_1} g^{\alpha_1 \beta_2}$.

APPENDIX D

ASYMPTOTIC EXPANSION OF THE EXPECTATION OF THE DERIVATIVES OF THE LOG LIKELIHOOD

Here, we show the asymptotic expansions of the expectation of the derivatives of the log likelihood. These asymptotic expansions relate the theory of power spectral densities to the theory of probability densities.

Consider a variance-covariance matrix Σ of size N parameterized by complex parameters $\theta = (\theta^1, \dots, \theta^p) \in \Theta \subset \mathbb{C}^p$. We set the matrix

$$\Sigma_{\alpha_1 \cdots \alpha_a, \beta_1 \cdots \beta_b, \gamma_1 \cdots \gamma_c} := \Sigma^{-1} (D_{\alpha_1 \cdots \alpha_a} \Sigma) \Sigma^{-1} (D_{\beta_1 \cdots \beta_b} \Sigma) \cdots \Sigma^{-1} (D_{\gamma_1 \cdots \gamma_c} \Sigma).$$

For example, $\Sigma_{\bar{1}, 2} = \Sigma^{-1} (\partial_1 \partial_{\bar{1}} \Sigma) \Sigma^{-1} (\partial_2 \Sigma)$. Direct computation shows that

$$\begin{aligned} \partial_{\alpha} l &= z^* (\Sigma_{\alpha}) \Sigma^{-1} z - \text{tr}(\Sigma_{\alpha}), \\ \partial_{\alpha} \partial_{\beta} l &= -z^* (\Sigma_{\alpha, \beta} + \Sigma_{\beta, \alpha}) \Sigma^{-1} z + z^* (\Sigma_{\alpha \beta}) \Sigma^{-1} z \\ & \quad - \text{tr}(-\Sigma_{\alpha, \beta} + \Sigma_{\alpha \beta}), \\ \partial_{\alpha} \partial_{\beta} \partial_{\gamma} l &= z^* (\Sigma_{\alpha, \beta, \gamma} + \Sigma_{\alpha, \gamma, \beta} + \Sigma_{\beta, \alpha, \gamma} \\ & \quad + \Sigma_{\beta, \gamma, \alpha} + \Sigma_{\gamma, \alpha, \beta} + \Sigma_{\gamma, \beta, \alpha}) \Sigma^{-1} z \\ & \quad - z^* (\Sigma_{\alpha, \beta \gamma} + \Sigma_{\beta, \gamma \alpha} + \Sigma_{\gamma, \alpha \beta} \\ & \quad + \Sigma_{\alpha \beta, \gamma} + \Sigma_{\beta \gamma, \alpha} + \Sigma_{\gamma \alpha, \beta}) \Sigma^{-1} z \\ & \quad + z^* (\Sigma_{\alpha \beta \gamma}) \Sigma^{-1} z \\ & \quad - \text{tr}(\Sigma_{\alpha, \beta, \gamma} + \Sigma_{\alpha, \gamma, \beta} - \Sigma_{\alpha, \beta \gamma} \\ & \quad - \Sigma_{\beta, \alpha \gamma} - \Sigma_{\gamma, \alpha \beta} + \Sigma_{\alpha \beta \gamma}), \end{aligned}$$

where $l = l_{\theta}^{(N)}(z^{(N)})$ is the log likelihood (56) of the observation $z^{(N)} \in \mathbb{C}^N$ from the complex-valued Gaussian process at the parameter $\theta \in \Theta \subset \mathbb{C}^p$.

As $E[zz^*] = \Sigma$, we have

$$\begin{aligned} \frac{1}{N} E_{\theta} [\partial_{\alpha} l_{\theta}^{(N)}] &= 0, \\ \frac{1}{N} E_{\theta} [\partial_{\alpha} \partial_{\beta} l_{\theta}^{(N)}] &= -\text{tr} \Sigma_{\alpha, \beta}, \\ \frac{1}{N} E_{\theta} [\partial_{\alpha} \partial_{\beta} \partial_{\gamma} l_{\theta}^{(N)}] &= 2 \text{tr} (\Sigma_{\alpha, \beta, \gamma} + \Sigma_{\gamma, \beta, \alpha}) \\ & \quad - \text{tr} \Sigma_{\alpha \beta, \gamma} - \text{tr} \Sigma_{\beta \gamma, \alpha} - \text{tr} \Sigma_{\gamma \alpha, \beta}, \end{aligned}$$

where E_{θ} denotes the expectation over the distribution $P_{\theta}^{(N)}$ of the observation $z^{(N)} \in \mathbb{C}^N$ at the parameter $\theta \in \Theta$.

Utilizing complex-valued tensorial polynomials, we also have

$$\begin{aligned} \frac{1}{N} E_{\theta} [(\partial_{\alpha} l_{\theta}^{(N)}) (\partial_{\beta} l_{\theta}^{(N)})] &= \text{tr} \Sigma_{\alpha, \beta}, \\ \frac{1}{N} E_{\theta} [(\partial_{\alpha} \partial_{\beta} l_{\theta}^{(N)}) (\partial_{\gamma} l_{\theta}^{(N)})] &= \text{tr} \Sigma_{\alpha \beta, \gamma} \\ & \quad - \text{tr} (\Sigma_{\alpha, \beta, \gamma} + \Sigma_{\gamma, \beta, \alpha}), \end{aligned}$$

and

$$\frac{1}{N} E_{\theta} [(\partial_{\alpha} l_{\theta}^{(N)}) (\partial_{\beta} l_{\theta}^{(N)}) (\partial_{\gamma} l_{\theta}^{(N)})] = \text{tr} (\Sigma_{\alpha, \beta, \gamma} + \Sigma_{\gamma, \beta, \alpha}).$$

To relate the expectation of the derivative of the log likelihood to the quantities $M_{\alpha_1 \cdots \alpha_a, \beta_1 \cdots \beta_b, \gamma_1 \cdots \gamma_c}$ defined as (15), we utilize the theorem proved in [24], which was originally proved for real-valued processes but is still valid for complex-valued processes. We introduce the space \mathcal{D} of power spectral densities of complex-valued processes defined on $[-\pi, \pi]$:

$$\mathcal{D} := \left\{ S \mid S(\omega) = \sum_{h=-\infty}^{\infty} \gamma_h e^{-\sqrt{-1} h \omega}, \right. \\ \left. \overline{\gamma_h} = \gamma_{-h}, \sum_{h=-\infty}^{\infty} |h| |\gamma_h| < \infty \right\}.$$

The space $\mathcal{D}_{\text{ARMA}}$ of power spectral densities for complex-valued stationary ARMA processes, where we have assumed causality and invertibility of the process, is a subspace of \mathcal{D} .

Proposition 4.1 ([24]): For $S_1, \dots, S_a \in \mathcal{D}_{\text{ARMA}}$ and $F_1, \dots, F_a \in \mathcal{D}$,

$$\begin{aligned} & \frac{1}{N} \text{tr} (\Sigma(F_1) \Sigma(S_1)^{-1} \cdots \Sigma(F_a) \Sigma(S_a)^{-1}) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} F_1(\omega) \cdots F_a(\omega) S_1^{-1}(\omega) \cdots S_a^{-1}(\omega) d\omega \\ & \quad + O(N^{-1}), \end{aligned}$$

where the (s, t) -th element $[\Sigma(F)]_{st}$ of the matrix $\Sigma(F)$ of size N is defined as $[\Sigma(F)]_{st} := \int_{-\pi}^{\pi} e^{\sqrt{-1}(s-t)\omega} F(\omega) d\omega$.

For a power spectral density $S = S_{\theta} \in \mathcal{D}_{\text{ARMA}}$ of a complex-valued stationary ARMA process parameterized by $\theta \in \Theta \subset \mathbb{C}^p$ and its variance-covariance matrix $\Sigma = \Sigma_{\theta}^{(N)}$ of size N for $z = z^{(N)} \in \mathbb{C}^N$, the (s, t) -th element $[\partial_{\alpha_1} \cdots \partial_{\alpha_a} \Sigma]_{st}$ of the matrix $\partial_{\alpha_1} \cdots \partial_{\alpha_a} \Sigma$ is calculated as

$$[\partial_{\alpha_1} \cdots \partial_{\alpha_a} \Sigma]_{st} = \int_{-\pi}^{\pi} e^{\sqrt{-1}(s-t)\omega} (\partial_{\alpha_1} \cdots \partial_{\alpha_a} S(\omega)) d\omega.$$

Thus, we have the following proposition, which relates the quantities $M_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c}$ defined as (15) to the quantities $K_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c}$ defined as (35).

Proposition 4.2: For a complex-valued stationary ARMA process parameterized by $\theta \in \Theta \subset \mathbb{C}^p$,

$$\frac{1}{N} E_\theta \left[\partial_\alpha l_\theta^{(N)} \right] = 0, \quad (62)$$

$$\frac{1}{N} E_\theta \left[\partial_\alpha \partial_\beta l_\theta^{(N)} \right] = -g_{\alpha\beta} + O(N^{-1}), \quad (63)$$

$$\begin{aligned} \frac{1}{N} E_\theta \left[\partial_\alpha \partial_\beta \partial_\gamma l_\theta^{(N)} \right] &= -\Gamma_{\alpha\beta\gamma}^{(m)} - \Gamma_{\beta\gamma\alpha}^{(m)} - \Gamma_{\gamma\alpha\beta}^{(m)} \\ &\quad + 2T_{\alpha\beta\gamma} + O(N^{-1}), \end{aligned} \quad (64)$$

$$\frac{1}{N} E_\theta \left[(\partial_\alpha l_\theta^{(N)}) (\partial_\beta l_\theta^{(N)}) \right] = g_{\alpha\beta} + O(N^{-1}), \quad (65)$$

$$\frac{1}{N} E_\theta \left[(\partial_\alpha \partial_\beta l_\theta^{(N)}) (\partial_\gamma l_\theta^{(N)}) \right] = \Gamma_{\alpha\beta\gamma}^{(m)} - T_{\alpha\beta\gamma} + O(N^{-1}), \quad (66)$$

$$\frac{1}{N} E_\theta \left[(\partial_\alpha l_\theta^{(N)}) (\partial_\beta l_\theta^{(N)}) (\partial_\gamma l_\theta^{(N)}) \right] = T_{\alpha\beta\gamma} + O(N^{-1}) \quad (67)$$

for $\alpha, \beta, \gamma \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, where $l_\theta^{(N)}$ denotes the log likelihood (56) of the observation $z = z^{(N)} \in \mathbb{C}^N$ of size N and E_θ denotes the expectation over the distribution $P_\theta^{(N)}$ of the observation $z^{(N)} \in \mathbb{C}^N$ at the parameter $\theta \in \Theta$. The quantities $g_{\alpha\beta}$, $T_{\alpha\beta\gamma}$, and $\Gamma_{\alpha\beta\gamma}^{(m)}$ on the right-hand side of the equations are defined as (16), (17), and (18), respectively.

APPENDIX E ASYMPTOTIC EXPANSION OF ESTIMATIVE POWER SPECTRAL DENSITIES

Here, we preset the asymptotic expansion of the risk of the estimative power spectral density with the maximum likelihood estimator. The risk is approximately $\frac{p}{N}$ for most predictive power spectral densities with asymptotically efficient estimators when we use p complex parameters.

Let us fix the true parameter $\theta_0 \in \Theta$ and denote the maximum likelihood estimator by $\hat{\theta}^{(N)}$ for a while; we set $S_0 := S_{\theta_0}$ and $\hat{S}^{(N)} := S_{\hat{\theta}^{(N)}}$. For $\lambda := \sqrt{N}(\hat{\theta}^{(N)} - \theta_0) = O_P(1)$, utilizing (58) and the Taylor expansion of $\hat{S}^{(N)}$ around S_0 , we obtain

$$\begin{aligned} D_{\text{KL}} \left(S_0 \parallel \hat{S}^{(N)} \right) &= \frac{1}{2N} g_{\alpha\beta} \lambda^\alpha \lambda^\beta + \frac{1}{N\sqrt{N}} \left(\frac{1}{2} \Gamma_{\alpha\beta\gamma}^{(m)} - \frac{1}{3} T_{\alpha\beta\gamma} \right) \lambda^\alpha \lambda^\beta \lambda^\gamma \\ &\quad + O_P(N^{-2}), \end{aligned} \quad (68)$$

where the quantities $M_{\alpha_1 \dots \alpha_a, \beta_1 \dots \beta_b, \dots, \gamma_1 \dots \gamma_c}$ appearing on the right-hand side are all evaluated at θ_0 . In particular, the asymptotic expansion of the risk of the maximum likelihood estimator evaluated at the true parameter θ_0 is given by

$$\begin{aligned} R(\hat{S}^{(N)} \mid \theta_0) &= E_{\theta_0} \left[D_{\text{KL}} \left(S_0 \parallel \hat{S}^{(N)} \right) \right] \\ &= \frac{p}{N} + O(N^{-2}) \end{aligned} \quad (69)$$

because $E_{\theta_0} [\lambda^\alpha \lambda^\beta] = g^{\alpha\beta} + O(N^{-1})$, $E_{\theta_0} [\lambda^\alpha \lambda^\beta \lambda^\gamma] = O(N^{-\frac{1}{2}})$, and $g^{\alpha\beta} g_{\alpha\beta} = 2p$.

APPENDIX F ASYMPTOTIC EXPANSION OF BAYESIAN PREDICTIVE POWER SPECTRAL DENSITIES

Here, we present an asymptotic expansion of the Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$ of a complex-valued ARMA process around the maximum likelihood estimator $\hat{\theta}^{(N)}$. This is the first step to obtaining the asymptotic expansion of the risk differences needed in the proof of the Main Theorem.

We follow the original proof [17] for the real-valued ARMA process. However, because we consider complex-valued processes, the definitions of some quantities must be slightly modified. Basically, the proof for the real-valued ARMA process is applied to the proof for the complex-valued ARMA process because a process parameterized by p complex parameters is essentially a process parameterized by $2p$ real parameters. However, we must keep in mind that the evenness $S(\omega) = S(-\omega)$ for $\omega \in [-\pi, \pi]$ of power spectral densities is not valid for complex-valued processes. In what follows, we carefully trace the proof for the real-valued process to ensure that the evenness of power spectral densities is nowhere used in the proof.

For the maximum likelihood estimator $\hat{\theta} = \hat{\theta}^{(N)}(z^{(N)}) = \theta_0 + O_P(N^{-\frac{1}{2}})$ for the observation $z^{(N)} \in \mathbb{C}^N$ of size N from the complex-valued Gaussian process having the true parameter $\theta_0 \in \Theta \subset \mathbb{C}^p$, the Bayesian predictive power spectral density is expanded as

$$\begin{aligned} \hat{S}_\pi^{(N)}(\omega) &= S_{\hat{\theta}}(\omega) + \frac{1}{\sqrt{N}} (\partial_\alpha S_{\hat{\theta}}(\omega)) E^\pi [\tilde{\lambda}^\alpha] \\ &\quad + \frac{1}{2N} (\partial_\alpha \partial_\beta S_{\hat{\theta}}(\omega)) E^\pi [\tilde{\lambda}^\alpha \tilde{\lambda}^\beta] + O_P(N^{-\frac{3}{2}}) \end{aligned} \quad (70)$$

around the maximum likelihood estimator $\hat{\theta}$, where

$$E^\pi [\tilde{\lambda}^{\alpha_1} \dots \tilde{\lambda}^{\alpha_a}] := \int_{\Theta} \tilde{\lambda}^{\alpha_1} \dots \tilde{\lambda}^{\alpha_a} \pi(\theta \mid z^{(N)}) \quad (71)$$

for $\tilde{\lambda} = \sqrt{N}(\theta - \hat{\theta})$ and $\alpha_1, \dots, \alpha_a \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. To complete the asymptotic expansion of (70) around the maximum likelihood estimator $\hat{\theta}$, we require the asymptotic expansions of (71) for $a = 1, 2$.

Let us fix, for a while, the observation $z^{(N)} \in \mathbb{C}^N$ and denote the maximum likelihood estimator by $\hat{\theta} = \hat{\theta}^{(N)}(z^{(N)})$. For any $\theta \in \Theta \subset \mathbb{C}^p$ such that $\tilde{\lambda} = \sqrt{N}(\theta - \hat{\theta}) = O(1)$, the asymptotic expansion of $\tilde{l}_\theta^{(N)}(z^{(N)}) := l_\theta^{(N)}(z^{(N)}) + \log \pi(\theta)$ around the maximum likelihood estimator $\hat{\theta}$ is calculated as

$$\tilde{l}_\theta^{(N)} = \tilde{l}_{\hat{\theta}}^{(N)} - \frac{1}{2} J_{\alpha\beta}^{(N)} \tilde{\lambda}^\alpha \tilde{\lambda}^\beta + \frac{A^{(N)}(\tilde{\lambda})}{\sqrt{N}} + O(N^{-1}), \quad (72)$$

where $J_{\alpha\beta}^{(N)} := -\frac{1}{N} \partial_\alpha \partial_\beta l_{\hat{\theta}}^{(N)}$ and

$$\begin{aligned} A^{(N)}(\tilde{\lambda}) &:= \frac{1}{6} \left(\frac{1}{N} \partial_\alpha \partial_\beta \partial_\gamma l_{\hat{\theta}}^{(N)} \right) \tilde{\lambda}^\alpha \tilde{\lambda}^\beta \tilde{\lambda}^\gamma \\ &\quad + \left(\partial_\alpha \log \pi(\hat{\theta}) \right) \tilde{\lambda}^\alpha. \end{aligned} \quad (73)$$

By referring to (72) and the procedure in [9], we can expand (71) as

$$\begin{aligned} E^\pi \left[\tilde{\lambda}^{\alpha_1} \dots \tilde{\lambda}^{\alpha_a} \right] &= \int_{\sqrt{N}(\Theta - \hat{\theta})} \tilde{\lambda}^{\alpha_1} \dots \tilde{\lambda}^{\alpha_a} \frac{1}{G^{(N)}} e^{-\frac{1}{2} J_{\alpha\beta}^{(N)} \tilde{\lambda}^\alpha \tilde{\lambda}^\beta} \\ &\quad \left(1 + \frac{A^{(N)}(\tilde{\lambda})}{\sqrt{N}} + O_P(N^{-1}) \right) d\tilde{\lambda} \end{aligned} \quad (74)$$

for $\tilde{\lambda} = \sqrt{N}(\theta - \hat{\theta})$, where

$$G^{(N)} := \int_{\sqrt{N}(\Theta - \hat{\theta})} e^{-\frac{1}{2} J_{\alpha\beta}^{(N)} \tilde{\lambda}^\alpha \tilde{\lambda}^\beta} d\tilde{\lambda}$$

is a normalization constant.

The terms in (71) for $a = 1, 2$ are expanded as

$$\begin{aligned} E^\pi \left[\tilde{\lambda}^\alpha \right] &= \frac{1}{6\sqrt{N}} L_{\beta\gamma\delta} I^{\alpha\beta\gamma\delta} + \frac{1}{\sqrt{N}} \left(\partial_\beta \log \pi(\hat{\theta}) \right) I^{\alpha\beta} \\ &\quad + O_P(N^{-1}), \end{aligned} \quad (75)$$

$$E^\pi \left[\tilde{\lambda}^\alpha \tilde{\lambda}^\beta \right] = I^{\alpha\beta} + O_P(N^{-1}), \quad (76)$$

where $\alpha, \beta, \gamma, \delta$ run through the indices $\{1, \dots, p, \bar{1}, \dots, \bar{p}\}$, and

$$L_{\alpha_1 \dots \alpha_a} := \frac{1}{N} \partial_{\alpha_1} \dots \partial_{\alpha_a} l_{\hat{\theta}}^{(N)}, \quad (77)$$

$$I^{\alpha_1 \dots \alpha_a} := \int_{\sqrt{N}(\Theta - \hat{\theta})} \tilde{\lambda}^{\alpha_1} \dots \tilde{\lambda}^{\alpha_a} \frac{1}{G^{(N)}} e^{-\frac{1}{2} J_{\alpha\beta}^{(N)} \tilde{\lambda}^\alpha \tilde{\lambda}^\beta} d\tilde{\lambda} \quad (78)$$

for $\alpha_1, \dots, \alpha_a \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. Note that $L_{\alpha_1 \dots \alpha_a}$ and $I^{\alpha_1 \dots \alpha_a}$ are complex-valued random variables because they depend on the realization of the observation $z^{(N)} \in \mathbb{C}^N$ from the process.

The expansion of (70) now becomes

$$\hat{S}_\pi^{(N)}(\omega) = S(\omega | \hat{\theta}) + \frac{1}{N} B_\pi^{(N)}(\omega | \hat{\theta}) + O_P(N^{-\frac{3}{2}}), \quad (79)$$

where $B_\pi^{(N)}$ is the $O_P(1)$ term defined as

$$\begin{aligned} B_\pi^{(N)}(\omega | \theta) &:= \frac{1}{6} L_{\beta\gamma\delta} I^{\alpha\beta\gamma\delta} (\partial_\alpha S(\omega | \theta)) \\ &\quad + I^{\alpha\beta} (\partial_\beta \log \pi(\theta)) (\partial_\alpha S(\omega | \theta)) \\ &\quad + \frac{1}{2} I^{\alpha\beta} (\partial_\alpha \partial_\beta S(\omega | \theta)). \end{aligned} \quad (80)$$

Utilizing the complex-valued tensorial Hermite polynomials defined in Appendix C and by Proposition 4.2 in Appendix D, we obtain an asymptotic expansion $B_\pi^{(N)}(\omega | \theta) = G_\pi^{(N)}(\omega | \theta) + H^{(N)}(\omega | \theta) + O_P(N^{-\frac{1}{2}})$, which yields the asymptotic expansion in (20).

Functions $G_\pi^{(N)}$ and $H^{(N)}$ represent the parallel and orthogonal parts of the quantity $N(\hat{S}_\pi^{(N)} - S_{\hat{\theta}^{(N)}})$, respectively; see also [12], [17].

APPENDIX G EXISTENCE OF BAYESIAN PREDICTIVE POWER SPECTRAL DENSITIES FOR $AR(p; \mathbb{C})$

First, we prove that the κ -prior (28) for $AR(p; \mathbb{C})$ is proper on $\tilde{\Theta}_1 = U \times \dots \times U$ if $\kappa < 1$ and is improper if $\kappa \geq 1$. If $\kappa \leq 0$, then the κ -prior $\pi^{(\kappa)}$ is certainly integrable on $\tilde{\Theta}_1$; therefore, we may assume $\kappa > 0$. Because

$$|1 - \xi^i \bar{\xi}^j|^2 - |\xi^i - \xi^j|^2 = (1 - |\xi^i|^2)(1 - |\xi^j|^2) \geq 0,$$

we have $|\xi^i - \xi^j|^2 / |1 - \xi^i \bar{\xi}^j|^2 \leq 1$. Thus, for $0 < \kappa < 1$,

$$\begin{aligned} \int_{\tilde{\Theta}_1} \pi^{(\kappa)}(\xi) d\xi &= \int_{\tilde{\Theta}_1} \frac{\prod_{1 \leq i < j \leq p} |\xi^i - \xi^j|^2}{\prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j)^\kappa} d\xi \\ &= \int_{\tilde{\Theta}_1} \frac{1}{\prod_{i=1}^p (1 - |\xi^i|^2)^\kappa} \\ &\quad \left(\prod_{1 \leq i < j \leq p} \frac{|\xi^i - \xi^j|^2}{|1 - \xi^i \bar{\xi}^j|^2} \right)^\kappa \\ &\quad \prod_{1 \leq i < j \leq p} |\xi^i - \xi^j|^{2(1-\kappa)} d\xi \\ &\leq 2^{p(p-1)(1-\kappa)} \prod_{i=1}^p \left(\int_U \frac{1}{(1 - |\xi|^2)^\kappa} d\xi^i \right) \\ &= 2^{p(p-1)(1-\kappa)} \left(\frac{\pi}{1 - \kappa} \right)^p \end{aligned}$$

because $\int_U (1 - |\xi|^2)^{-\kappa} d\xi^i = \int_0^1 \int_{-\pi}^\pi (1 - r^2)^{-\kappa} r dr d\theta = \frac{\pi}{1-\kappa}$. Therefore, the κ -prior $\pi^{(\kappa)}$ is integrable on $\tilde{\Theta}_1$ if $\kappa < 1$.
Set

$$m := \min_{\xi \in \Xi} \prod_{1 \leq i < j \leq p} \frac{|\xi^i - \xi^j|^2}{|1 - \xi^i \bar{\xi}^j|^{2\kappa}} > 0,$$

where $\Xi := V_1 \times \dots \times V_p \subset \tilde{\Theta}_1$ and

$$V_i := \left\{ \xi \in U \mid \frac{1}{2} < |\xi| < 1, \frac{2\pi}{N}(i-1/2) < \arg \xi < \frac{2\pi}{N}i \right\}.$$

We see that $\pi^{(\kappa)}$ is not integrable on $\tilde{\Theta}_1$ if $\kappa \geq 1$, because

$$\begin{aligned} \int_{\tilde{\Theta}_1} \pi^{(\kappa)}(\xi) d\xi &\geq \int_{\Xi} \pi^{(\kappa)}(\xi) d\xi \\ &\geq m \prod_{i=1}^p \int_{V_i} (1 - |\xi|^2)^{-\kappa} d\xi = +\infty. \end{aligned}$$

Next, we prove that a function $p_\xi^{(N)}(z^{(N)}) \pi^{(\kappa)}(\xi)$ of ξ is integrable on the parameter space $\tilde{\Theta}_1$ if $\kappa < 2$. The explicit form of the determinant $|\Sigma^{(N)}|$ of the variance-covariance matrix $\Sigma^{(N)}$ of $AR(p; \mathbb{C})$ of the form in (23) is $\gamma_0 = |\Sigma^{(1)}| \leq \dots \leq |\Sigma^{(p)}| = |\Sigma^{(p+1)}| = \dots = \prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j)^{-1}$; see Section 5.5 (b), (c), and (d) in [25] or Theorem 3.1 in

[26]. Thus, if $N \geq p$, we have

$$\begin{aligned} p^{(N)}(z^{(N)} | \xi) \pi^{(\kappa)}(\xi) &= \pi^{-N} |\Sigma^{(N)}|^{-1} e^{-z^{(N)*} (\Sigma^{(N)})^{-1} z^{(N)}} \\ &\quad \frac{\prod_{1 \leq i < j \leq q} |\xi^i - \xi^j|^2}{\prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j)^\kappa} \\ &\leq \pi^{-N} \frac{\prod_{1 \leq i < j \leq q} |\xi^i - \xi^j|^2}{\prod_{i=1}^p \prod_{j=1}^p (1 - \xi^i \bar{\xi}^j)^{\kappa-1}} \\ &= \pi^{-N} (\pi^{(\kappa-1)}(\xi)) \end{aligned}$$

for $z^{(N)} \in \mathbb{C}^N$. If $\kappa < 2$ and $N \geq p$, then $\int_{\bar{\Theta}_1} p_\xi^{(N)}(z^{(N)}) \pi^{(\kappa)}(\xi) d\xi$ is bounded, regardless of the sample $z^{(N)} \in \mathbb{C}^N$. Therefore, the Bayesian predictive power spectral densities $\hat{S}_{\pi^{(\kappa)}}^{(N)}$ for $AR(p; \mathbb{C})$ based on the κ -prior $\pi^{(\kappa)}$ exists if $\kappa < 2$ and $N \geq p$.

The posterior $\pi^{(\kappa)}(\xi | z^{(N)})$ given an observation $z^{(N)}$ based on the prior $\pi^{(\kappa)}(\xi)$ is calculated as

$$\begin{aligned} \pi^{(\kappa)}(\xi | z^{(N)}) &\propto p^{(N)}(z^{(N)} | \xi) \pi^{(\kappa)}(\xi) \\ &\propto \pi^{(\kappa-1)}(\xi) \times e^{-z^{(N)*} (\Sigma^{(N)})^{-1} z^{(N)}} \end{aligned} \quad (81)$$

if $N \geq p$, which shows that the family of κ -priors is not closed under sampling, because of the exponential term. The conjugation property of the prior family is a topic for future research.

APPENDIX H PROOF OF THE MAIN THEOREM

Here, we prove the Main Theorem. The generalization, including the i.i.d. case, for this theorem is discussed in Section VI. Except Proposition 8.3, most of the propositions presented in this section are merely complexified versions of previous works; see [12], [17]. In particular, the proofs for Proposition 8.1 and Proposition 8.2 only rely on the form (20) of the asymptotic expansion of Bayesian predictive power spectral densities (or Bayesian predictive distributions for the i.i.d. case). Therefore, we present the minimal outlines for these proofs because the procedures in previous works can be directly applied to them.

We first present the asymptotic expansion of the risk $R(\hat{S}_\pi^{(N)} | \theta)$ for a Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$. The original proof of Proposition 8.1 for the i.i.d. case was introduced in [12], and the proof for real-valued processes is reported in [17].

The covariant derivative $\overset{(e)}{\nabla}_\alpha$ of the vector field V^β is defined as $\overset{(e)}{\nabla}_\alpha V^\beta := \partial_\alpha V^\beta + \Gamma_{\alpha\gamma}^\beta V^\gamma$, where $\Gamma_{\alpha\gamma}^\beta := \Gamma_{\alpha\gamma}^\beta - T_{\alpha\gamma\delta} g^{\delta\beta}$.

Proposition 8.1: For a complex parameter space Θ and a possibly improper prior π , we have

$$\begin{aligned} R(\hat{S}_\pi^{(N)} | \theta) &= \frac{1}{2N^2} g^{\alpha\beta} \left(\partial_\alpha \log \frac{\pi}{\pi_J} + \frac{1}{2} T_\alpha \right) \left(\partial_\beta \log \frac{\pi}{\pi_J} + \frac{1}{2} T_\beta \right) \\ &\quad + \frac{1}{N^2} \overset{(e)}{\nabla}_\alpha \left(g^{\alpha\beta} \left(\partial_\beta \log \frac{\pi}{\pi_J} + T_\beta \right) \right) + C + O(N^{-\frac{5}{2}}), \end{aligned} \quad (82)$$

where C is a term independent of the prior π .

Proof: Recall that the asymptotic expansion of a Bayesian predictive power spectral density $\hat{S}_\pi^{(N)}$ is given by (20). We follow the procedure in [17] but replace the summation rule $i, j, k = 1, \dots, p$ with $\alpha, \beta, \gamma = 1, \dots, p, \bar{1}, \dots, \bar{p}$. ■

The asymptotic expansion of the risk difference $R(\hat{S}_{\pi_1}^{(N)} | \theta) - R(\hat{S}_{\pi_2}^{(N)} | \theta)$ is calculated as follows; see also [12], [17].

Proposition 8.2: For a complex parameter space Θ and a possibly improper prior π , we have

$$\begin{aligned} N^2 \left(R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_\pi^{(N)} | \theta) \right) &= g^{\alpha\beta} \left(\partial_\alpha \log \frac{\pi}{\pi_J} \right) \left(\partial_\beta \log \frac{\pi}{\pi_J} \right) \\ &\quad - \left(\frac{\pi}{\pi_J} \right)^{-1} \Delta \left(\frac{\pi}{\pi_J} \right) + O(N^{-\frac{1}{2}}) \\ &= -2 \left(\frac{\pi}{\pi_J} \right)^{-\frac{1}{2}} \Delta \left(\frac{\pi}{\pi_J} \right)^{\frac{1}{2}} + O(N^{-\frac{1}{2}}) \end{aligned}$$

for $\theta \in \Theta$.

Proof: Use (82) and follow [17]. ■

Proposition 8.3: Let ϕ be a positive continuous function on a Kähler parameter space Θ , and define a family $\pi^{(\kappa)} := \phi^{-\kappa+1} \pi_J$ of priors for $\kappa \in \mathbb{R}$. Then, we have

$$\begin{aligned} N^2 \left(R(\hat{S}_{\pi_1}^{(N)} | \theta) - R(\hat{S}_{\pi_2}^{(N)} | \theta) \right) &= -(\kappa_1 - \kappa_2) \frac{\Delta \phi}{\phi} \\ &\quad + (\kappa_1 - \kappa_2)(\kappa_1 + \kappa_2) g^{i\bar{j}} (\partial_i \log \phi) (\partial_{\bar{j}} \log \phi) \\ &\quad + O(N^{-\frac{1}{2}}) \end{aligned}$$

for $\theta \in \Theta$, where $\pi_1 := \pi^{(\kappa_1)}$ and $\pi_2 := \pi^{(\kappa_2)}$.

Proof: Let $\kappa_1 = 1$ and $\kappa_2 = \kappa$. Using (54) for $a = \frac{1-\kappa}{2}$, we have

$$\begin{aligned} N^2 \left(R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi^{(\kappa)}}^{(N)} | \theta) \right) &= -2 \frac{\Delta \phi^{\frac{1-\kappa}{2}}}{\phi^{\frac{1-\kappa}{2}}} + O(N^{-\frac{1}{2}}) \\ &= (1 - \kappa) \frac{\Delta \phi}{\phi} + (\kappa^2 - 1) g^{i\bar{j}} (\partial_i \log \phi) (\partial_{\bar{j}} \log \phi) \\ &\quad + O(N^{-\frac{1}{2}}). \end{aligned}$$

The formula

$$\begin{aligned} N^2 \left(R(\hat{S}_{\pi_1}^{(N)} | \theta) - R(\hat{S}_{\pi_2}^{(N)} | \theta) \right) &= -N^2 \left(R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi_1}^{(N)} | \theta) \right) \\ &\quad - N^2 \left(R(\hat{S}_{\pi_J}^{(N)} | \theta) - R(\hat{S}_{\pi_2}^{(N)} | \theta) \right) \end{aligned}$$

yields the statement. ■

If there exists a positive continuous eigenfunction $\phi > 0$ of the Laplacian Δ with a negative eigenvalue $-K < 0$, then $\frac{\Delta \phi}{\phi} = -K$, which yields the proof of Theorem 4.1.

APPENDIX I RELATION WITH α -PARALLEL PRIORS

In this appendix, we show the relation between κ -priors and α -parallel priors.

First, we define the α -connection. Let Θ be a complex parameter space in \mathbb{C}^p . For $\alpha \in \mathbb{R}$, set

$$\Gamma_{\beta\gamma\delta}^{(\alpha)} := \Gamma_{\beta\gamma\delta}^{(m)} - \frac{1+\alpha}{2} T_{\beta\gamma\delta}, \quad (83)$$

$$\Gamma_{\beta\gamma}^{(\alpha)\delta} := \Gamma_{\beta\gamma\epsilon}^{(\alpha)} g^{\epsilon\delta} \quad (84)$$

for $\beta, \gamma, \delta \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$. For complex-valued stationary ARMA processes, the quantities $g_{\alpha\beta}$, $T_{\alpha\beta\gamma}$, and $\Gamma_{\alpha\beta\gamma}^{(m)}$ on the right-hand side of the equations are defined as (16), (17), and (18), respectively. For the i.i.d. case, these quantities are defined as (36), (37), and (38), respectively. Note that $\Gamma_{\beta\gamma\delta}^{(-1)}$ corresponds to $\Gamma_{\beta\gamma\delta}^{(m)}$. On the other hand, $\Gamma_{\beta\gamma\delta}^{(+1)}$ corresponds to $\Gamma_{\beta\gamma\delta}^{(e)}$. The quantity $\Gamma_{\beta\gamma}^{(\alpha)\delta} - T_{\beta\gamma\delta}$ is often denoted by $\Gamma_{\beta\gamma\delta}^{(e)}$. The quantity $\Gamma_{\beta\gamma}^{(\alpha)\delta}$ is called the coefficients of the α -connection $\nabla^{(\alpha)}$ on the parameter space. The α -connection with $\alpha = -1$ is called the mixture connection (m -connection) $\nabla^{(m)}$, and the α -connection with $\alpha = +1$ is called the exponential connection (e -connection) $\nabla^{(e)}$. The α -connection with $\alpha = 0$ is called the Levi-Civita connection $\nabla^{(0)}$. From (83), we have

$$\Gamma_{\beta\gamma\delta}^{(\alpha)} = \Gamma_{\beta\gamma\delta}^{(0)} - \frac{\alpha}{2} T_{\beta\gamma\delta}. \quad (85)$$

For the geometrical interpretation of the α -connection $\nabla^{(\alpha)}$ on the statistical manifold, see [23], [27].

If the complex parameter space Θ is Kähler, the nontrivial elements of the coefficients $\Gamma_{\beta\gamma}^{(\alpha)\delta}$ of the 0-connection $\nabla^{(0)}$ are only $\Gamma_{ij}^{(0)k}$ and $\Gamma_{\bar{i}\bar{j}}^{(0)\bar{k}}$; other coefficients, such as $\Gamma_{ij}^{(0)\bar{k}}$, $\Gamma_{\bar{i}\bar{j}}^{(0)k}$, and $\Gamma_{\bar{i}\bar{j}}^{(0)\bar{k}}$, are all equal to 0; see Section 8.5 in [20].

The non-negative function $\rho^{(\alpha)}$ with $\alpha \in \mathbb{R}$ defined on the complex parameter space Θ is called an α -parallel prior, if

$$\partial_\beta \rho^{(\alpha)} = \Gamma_{\beta\gamma}^{(\alpha)\gamma} \rho^{(\alpha)} \quad (86)$$

for $\beta, \gamma \in \{1, \dots, p, \bar{1}, \dots, \bar{p}\}$ on the complex parameter space Θ ; see [14]. The quantity $\left(\partial_\beta - \Gamma_{\beta\gamma}^{(\alpha)\gamma}\right) \rho^{(\alpha)}$ is referred

to as a contravariant derivative of $\rho^{(\alpha)}$ with respect to the α -connection $\nabla^{(\alpha)}$. The α -parallel prior $\rho^{(\alpha)}$ corresponds to the volume element that is parallel with respect to the α -connection $\nabla^{(\alpha)}$. In general, the existence of the α -parallel prior for each $\alpha \in \mathbb{R}$ is not guaranteed. However, we can provide a sufficient condition for the existence of a family of α -parallel priors on a Kähler parameter space.

Let us compute the contravariant derivatives of priors with respect to the α -connection $\nabla^{(\alpha)}$. Let ϕ be a non-negative function globally defined on a Kähler parameter space $\Theta \subset \mathbb{C}^p$. We define a family $\{\pi^{(\kappa)}\}_{\kappa \in \mathbb{R}}$ of priors by $\pi^{(\kappa)} := \phi^{-\kappa+1} \pi_J$, where π_J denotes the Jeffreys prior. Note that $\pi^{(1)}$ corresponds to the Jeffreys prior π_J . As Θ is Kähler, we have

$$\begin{aligned} \Gamma_i^{(0)} &:= \Gamma_{ij}^{(0)j} + \Gamma_{i\bar{k}}^{(0)\bar{k}} = \Gamma_{ij}^{(0)j} = \Gamma_{i\bar{j}\bar{k}}^{(0)} g^{j\bar{k}} \\ &= (\partial_i g_{j\bar{k}}) g^{j\bar{k}} = \partial_i \log \pi_J, \end{aligned} \quad (87)$$

where we used the Jacobi formula (51). As

$$\Gamma_i^{(0)} \pi^{(\kappa)} = (\partial_i \log \pi_J) \phi^{-\kappa+1} \pi_J = \phi^{-\kappa+1} \partial_i \pi_J, \quad (88)$$

we have

$$\begin{aligned} \Gamma_i^{(\alpha)} \pi^{(\kappa)} &= \left(\Gamma_i^{(0)} - \frac{\alpha}{2} T_i \right) \pi^{(\kappa)} \\ &= \phi^{-\kappa+1} \partial_i \pi_J - \frac{\alpha}{2} T_i \phi^{-\kappa+1} \pi_J, \end{aligned} \quad (89)$$

where $T_i := T_{ik\bar{j}} g^{k\bar{j}} + T_{i\bar{j}k} g^{k\bar{j}}$. On the other hand, we have

$$\partial_i \pi^{(\kappa)} = \phi^{-\kappa+1} \partial_i \pi_J - (\kappa - 1) \phi^{-\kappa+1} (\partial_i \log \phi) \pi_J. \quad (90)$$

Thus, the contravariant derivative of $\pi^{(\kappa)}$ with respect to the α -connection $\nabla^{(\alpha)}$ is

$$\left(\partial_i - \Gamma_i^{(\alpha)} \right) \pi^{(\kappa)} = \left(\frac{\alpha}{2} T_i - (\kappa - 1) (\partial_i \log \phi) \right) \phi^{-\kappa+1} \pi_J. \quad (91)$$

We see that if $\partial_i \log \phi$ is proportional to T_i , the proposed prior $\pi^{(\kappa)}$ is an α -parallel prior for some α . Thus, we have the following proposition.

Proposition 9.1: If $T_i = -4c \partial_i \log \phi$ for some $c \neq 0$, then the prior $\pi^{(\kappa)}$ is an α -parallel prior with $\alpha = (1 - \kappa)/2c$.

Moreover, the term $\left(\partial_\alpha \log \frac{\pi_J}{\pi} + \frac{1}{2} T_\alpha\right)$, which appears in the definition (21) of the parallel part $G_\pi^{(N)}$ of the risk difference between the Bayesian predictive distribution and the estimative distribution with the maximum likelihood estimator, now becomes

$$\partial_i \log \frac{\pi^{(\kappa)}}{\pi_J} + \frac{1}{2} T_i = -(\kappa - (1 - 2c)) \partial_i \log \phi \quad (92)$$

for $i = 1, \dots, p$. Thus, we have the following proposition.

Proposition 9.2: If $T_i = -4c \partial_i \log \phi$ for some $c \neq 0$, then the risk of the Bayesian predictive distribution asymptotically dominates the risk of the estimative distribution with the maximum likelihood estimator when $\kappa = 1 - 2c$, i.e., $\alpha = (1 - \kappa)/2c = 1$.

For the non-negative function ϕ defined as (27) for $\text{AR}(p; \mathbb{C})$, we have $T_i = -4\partial_i \log \phi = 4\bar{\xi}^j g_{i\bar{j}}$, i.e., $c = 1$. Thus, $\pi^{(\kappa)}$ is an α -parallel prior with $\alpha = (1 - \kappa)/2$ for $\text{AR}(p; \mathbb{C})$. In particular, the proposed prior $\pi^{(-1)}$, which asymptotically achieves the constant risk improvement, is a (+1)-parallel prior. Moreover, the risk of the Bayesian predictive power spectral density based on the proposed prior $\pi^{(-1)}$ asymptotically dominates the risk of the estimative predictive power spectral density with the maximum likelihood estimator.

ACKNOWLEDGMENT

The authors express their gratitude to Keisuke Yano for helpful comments.

REFERENCES

- [1] K. S. Miller, *Complex Stochastic Processes: An Introduction to Theory and Application*. Reading, MA, USA: Addison-Wesley, 1974.
- [2] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, vol. 59. Hoboken, NJ, USA: Wiley, 2009.
- [3] S. Svetunkov, *Complex-Valued Modeling in Economics and Finance*. New York, NY, USA: Springer-Verlag, 2012.
- [4] A. Hirose, *Complex-Valued Neural Networks: Theories and Applications*, vol. 5. Singapore: World Scientific, 2003.
- [5] F. Komaki, "An estimating method for parametric spectral densities of Gaussian time series," *J. Time Ser. Anal.*, vol. 20, no. 1, pp. 31–50, Jan. 1999.
- [6] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York, NY, USA: Springer-Verlag, 1991.
- [7] J. O. Berger and R.-Y. Yang, "Noninformative priors and Bayesian testing for the AR(1) model," *Econ. Theory*, vol. 10, nos. 3–4, pp. 461–482, Aug. 1994.
- [8] A. Zellner, "Maximal data information prior distributions," in *New Developments in the Applications of Bayesian Methods*. Amsterdam, The Netherlands: Elsevier, 1977, pp. 211–232.
- [9] A. Philippe and J. Rousseau, "Non-informative priors in the case of Gaussian long-memory processes," *Bernoulli*, vol. 8, no. 4, pp. 451–473, 2002.
- [10] L. D. Brown, "Admissible estimators, recurrent diffusions, and insoluble boundary value problems," *Ann. Math. Statist.*, vol. 42, no. 3, pp. 855–903, Jun. 1971.
- [11] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.
- [12] F. Komaki, "Shrinkage priors for Bayesian prediction," *Ann. Statist.*, vol. 34, no. 2, pp. 808–819, Apr. 2006.
- [13] J. Choi and A. Mullhaupt, "Kählerian information geometry for signal processing," *Entropy*, vol. 17, no. 4, pp. 1581–1605, 2015.
- [14] J. Takeuchi and S. Amari, " α -parallel prior and its properties," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1011–1023, Feb. 2005.
- [15] P. Whittle, "Estimation and information in stationary time series," *Arkiv Matematik*, vol. 2, no. 5, pp. 423–434, Aug. 1953.
- [16] T. W. Anderson, "Estimation for autoregressive moving average models in the time and frequency domains," *Ann. Statist.*, vol. 5, no. 5, pp. 842–865, Sep. 1977.
- [17] F. Tanaka and F. Komaki, "Asymptotic expansion of the risk difference of the Bayesian spectral density in the autoregressive moving average model," *Sankhya A*, vol. 73, no. 1, pp. 162–184, Feb. 2011.
- [18] S.-I. Amari, "Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections, and divergence," *Math. Syst. Theory*, vol. 20, no. 1, pp. 53–82, Dec. 1987.
- [19] F. Tanaka, "Superharmonic priors for autoregressive models," *Inf. Geometry*, vol. 1, no. 2, pp. 215–235, Dec. 2018.
- [20] M. Nakahara, *Geometry, Topology and Physics*. Boca Raton, FL, USA: CRC Press, 2003.
- [21] F. Komaki, "On asymptotic properties of predictive distributions," *Biometrika*, vol. 83, no. 2, pp. 299–313, Jun. 1996.
- [22] J. Aitchison, "Goodness of prediction fit," *Biometrika*, vol. 62, no. 3, pp. 547–554, 1975.
- [23] S.-I. Amari and M. Kumon, "Differential geometry of edgeworth expansions in curved exponential family," *Ann. Inst. Stat. Math.*, vol. 35, no. 1, pp. 1–24, Dec. 1983.
- [24] M. Taniguchi, "On the second order asymptotic efficiency of estimators of Gaussian ARMA processes," *Ann. Statist.*, vol. 11, no. 1, pp. 157–169, Mar. 1983.
- [25] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*. Berkeley, CA, USA: Univ. California Press, 1958.
- [26] P. Shaman, "Approximations for stationary covariance matrices and their inverses with application to ARMA models," *Ann. Statist.*, vol. 4, no. 2, pp. 292–301, Mar. 1976.
- [27] S. Amari, *Differential-Geometrical Methods in Statistics* (Lecture Notes on Statistics), vol. 28. New York, NY, USA: Springer-Verlag, 1985, p. 1.

Hidemasa Oda received the B.Sc. and M.Sc. degrees in mathematical science from The University of Tokyo, Japan, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree with the Department of Mathematical Informatics. From 2013 to 2014, he worked at Nippon Life Insurance Company, Tokyo, as a Pension Actuary. From 2014 to 2019, he worked at SmartNews Inc., Tokyo, as a Software Engineer. His research interests include Bayesian theory, information geometry, and mathematical physics.

Fumiyasu Komaki (Member, IEEE) received the B.Eng. and M.Eng. degrees in mathematical engineering from The University of Tokyo, Japan, in 1987 and 1989, respectively, and the Ph.D. degree in statistical science from the Graduate University for Advanced Studies, Japan, in 1992. He is currently a Professor with the Department of Mathematical Informatics, The University of Tokyo, and a Unit Leader at the RIKEN Center for Brain Science. His research interests include prediction theory, Bayesian theory, and information geometry.