A Robust Spectral Estimator with Application to a Noise Corrupted Process

Xin Zhou, *Member, IEEE* Steven Kay, *Life Fellow, IEEE*

***Abstract*—When a data set is corrupted by noise, the model for the data generating process is misspecified and can cause parameter estimation problems. As an example, in the case of a Gaussian autoregressive (AR) process corrupted by noise, the data is more accurately modeled as an autoregressive moving average (ARMA) process rather than an AR process. This misspecification leads to bias, and hence, low resolution in AR spectral estimation. However, a new parametric spectral estimator, the realizable information theoretic estimator (RITE) based on a nonhomogeneous Poisson spectral representation, is shown by simulation to be more robust to white noise than the asymptotic maximum likelihood estimator (MLE). We therefore conducted an in-depth investigation and analyzed the statistics of RITE and the asymptotic MLE for the misspecified model. For large data records, RITE and the asymptotic MLE are both asymptotically normally distributed. The asymptotic MLE has a slightly lower variance, but RITE exhibits much less bias. Simulation examples of a white noise corrupted AR process are provided to support the theoretical properties. This advantage of RITE increases as the signal-to-noise-ratio (SNR) decreases.**

***Index Terms*—Non-homogeneous Poisson, spectral analysis, misspecified, distribution, robustness**

I. Introduction

THE spectral representation for a wide sense stationary (WSS) random process relies on the time representation which is a sum of sinusoids with fixed frequencies, random phases and random amplitudes [1]. It forms the basis for spectral estimation. Another less well known representation models the frequencies as random point events distributed according to a nonhomogeneous Poisson process. The likelihood function can be derived for this spectral representation. Since the frequency events are usually not observable, some modifications are applied to the likelihood function. The estimator that maximizes the approximated likelihood function is called the realizable information theoretic estimator (RITE) [2]. It can be used in model-based spectral estimation.

The autoregressive (AR) model is widely used in spectral estimation. The maximum likelihood estimator (MLE) is usually employed for a good estimate of the AR parameters. If we assume a real Gaussian random process, the autocorrelation method, which requires solving the Yule-Walker equations with a suitable autocorrelation function (ACF), can be found efficiently and is equivalent to the approximate MLE [3]. Many other methods for AR parameters estimation that produce the same numerical values for large data records, like the Burg method and the covariance method, are also approximate MLEs [4]. Hence those methods share the desirable properties of the MLE that for large data records they are consistent, asymptotically Gaussian, unbiased and attain the Cramer-Rao lower bound (CRLB) [5] [6]. However, when the observations are corrupted by additive noise, the various methods for AR parameters estimation mentioned above produce severe biases. This sensitivity to the noise addition results in a smoothed AR spectral estimate. Numerous studies indicate that the resolution of estimated AR spectra decreases as the signal-to-noise-ratio (SNR) decreases [7] [8] [9]. This is because the additive noise changes the true model to an autoregressive moving average (ARMA) where AR and moving average (MA) parameters are linked, instead of an AR. Hence the methods above are no longer the true MLEs. To get better resolution, one option is to use an ARMA model estimated by the least squares modified Yule-Walker equations (LSMYWE) [10], but the model order of the MA part depends on the noise type (see supplementary material S.5), and therefore limits its utility in practice.

The MLE for a misspecified model is called a quasi-MLE [11]. A misspecified model has been investigated but to a lesser extent in [11] [12]. Hence more analysis on misspecification is necessary. Compared with the asymptotic Gaussian MLE, RITE shows a robustness to white noise in PSD classification problems [2]. Therefore it would be of interest to investigate how RITE performs in AR spectral estimation.

In this paper, the asymptotic statistical properties of the quasi-MLE and RITE are derived and verified by simulation examples. Both estimators are asymptotically Gaussian distributed but with different means and covariance matrices. An application to spectral estimation using the AR model is provided in the paper. For an AR PSD, we prove that the asymptotic Gaussian likelihood function is more sensitive to white noise than the RITE likelihood function. Our experiments show that in comparison to the quasi-MLE, RITE has smaller bias when white noise is present in AR process.

The paper is organized as follows. Section II gives a brief introduction to RITE. In Section III, the theoretical properties of MLE and RITE are given. Section IV reviews the AR model and gives a simple explanation of the white noise robustness of RITE. In Section V, spectral estimation application using AR model are provided to verify the theory in Section III and to show the robustness of RITE for white noise corrupted data. Section VI summarizes our results and discusses future work.

II. Realizable Information Theoretic Estimator

The background for this section can be found in [2] and [13]. A real discrete-time WSS random process can be represented in the spectral form as a sum of sinusoids with random frequencies, amplitudes and phases:

$$x[n] = \frac{1}{\sqrt{\lambda_0/2}} \sum_{k=1}^{M} A_k \cos(2\pi F_k n + \Phi_k) \qquad -\infty < n < \infty$$

A similar representation that uses two independent Poisson point processes can be found in [14]. The representation herein can be viewed as a marked Poisson process. If the number of events M is fixed, then the model reduces to that in [15]. In this study we take M as the number of events of a nonhomogeneous Poisson random process in frequency with intensity $\lambda(f)$ on the interval $[0, 0.5]$. F_k is the k^{th} point event on the frequency interval $0 \le f \le 0.5$ with "marks" (A_k, Φ_k). $A_1, A_2, \cdots, A_N$ are independent and identically distributed (IID) positive amplitude random variables. $\Phi_1, \Phi_2, \cdots, \Phi_N$ are phase random variables uniformly IID on $[0, 2\pi)$. The amplitude, phase, and frequency random variables are independent of each other.

We normalize the intensity by $\lambda(f) = \lambda_0 p(f)$. With this normalization, the integral of $p(f)$ over $[0, 0.5]$ is equal to 1. This property allows $p(f)$ to be interpreted as a probability density function (PDF) on $0 \le f \le 0.5$. The power spectral density (PSD) of $x[n]$ can be shown to be $P(f) = \frac{E(A^2)}{2} p(|f|)$ on $-0.5 \le f \le 0.5$ [2]. Here, we are only interested in the case that the total power is 1, i.e. $E(A^2) = 1$ or equivalently $\int_{-\frac{1}{2}}^{\frac{1}{2}} P(f)df = 1$. From the above relations and conditions, we can write the intensity function in terms of the PSD function as $\lambda(f) = 2\lambda_0 P(f)$ on $0 \le f \le 0.5$, and use $P(-f) = P(f)$.

It can be shown that the part of the log-likelihood that depends on $\lambda(f)$ is

$$l = -\int_0^{\frac{1}{2}} \lambda(f)df + \int_0^{\frac{1}{2}} \ln \lambda(f) N(df)$$

with $N(df)$ is the random variable indicating the number of frequency events on the interval $[f, f + df)$. Since we cannot observe the frequency events but only $x[n]$ in general, we proceed by replacing $N(df)$ with its approximate mean:

$$E(N(df)) = \lambda(f)df = 2\lambda_0 P(f)df \approx 2\lambda_0 \bar{I}(f)df$$

where $\bar{I}(f)$ is the normalized periodogram, which is given by

$$\bar{I}(f) = \frac{I(f)}{\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)df}$$

and $I(f)$ is the unnormalized periodogram

$$I(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j2\pi nf) \right|^2$$

In accordance with $\int_{-\frac{1}{2}}^{\frac{1}{2}} P(f)df = 1$, the periodogram is normalized to ensure $\int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{I}(f)df = 1$. We now have the approximated likelihood function:

$$\begin{aligned}
l' &\approx -\int_0^{\frac{1}{2}} \lambda(f)df + \int_0^{\frac{1}{2}} \ln(\lambda(f)) 2\lambda_0 \bar{I}(f)df \\
&= -\lambda_0 + 2\lambda_0 \int_0^{\frac{1}{2}} \ln(2\lambda_0 P(f)) \bar{I}(f)df \\
&= -\lambda_0 + 2\lambda_0 \ln(2\lambda_0) \int_0^{\frac{1}{2}} \bar{I}(f)df + 2\lambda_0 \int_0^{\frac{1}{2}} \bar{I}(f) \ln P(f)df \\
&= -\lambda_0 + \lambda_0 \ln(2\lambda_0) + 2\lambda_0 \int_0^{\frac{1}{2}} \bar{I}(f) \ln P(f)df
\end{aligned}$$

Ignoring the terms that do not depend on the PSD and the scaling λ_0, we have the realizable likelihood function

$$l_R = \int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{I}(f) \ln P(f)df$$

The function $\int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{I}(f) \ln P(f)df$ achieves its maximum when $P(f)$ is identical to $\bar{I}(f)$. If we assume that the PSD depends on a set of parameters, then the estimation of those parameters is chosen to maximize l_R. Note that $\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)df = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$, it follows that

$$\begin{aligned}
l_R &= \frac{1}{\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)df} \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f) \ln P(f)df \\
&= \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]} \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f) \ln P(f)df
\end{aligned}$$

Since the maximization result does not depend on the normalization term $\frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]}$, we finally have the RITE likelihood function as

$$l_R = \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f) \ln P(f)df$$

III. The Statistical Properties of MLE and RITE

The MLE and RITE are both obtained by maximizing their likelihood functions. They are both special cases of M-estimators. Huber introduced M-estimators and analyzed their asymptotic properties [16]. The derivation of the statistical properties for MLE and RITE are based on the theory of M-estimators. More detailed information about the M-estimator can be found in [17].

Let the signal $s[n]$ be a wide sense stationary (WSS) Gaussian random process whose power equals 1. Let $\{x[0], x[1], \cdots, x[N-1]\}$ be a $N \times 1$ observed data set generated from the noise corrupted signal, with PSD function $Q(f; \boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ is the true value of a $q \times 1$ vector parameter. We propose $P(f; \boldsymbol{\theta})$ to be the PSD model of the signal, where $\boldsymbol{\theta}$ is a $p \times 1$ vector parameter. Assume $P(f; \boldsymbol{\theta})$ is suitably smooth on $\boldsymbol{\theta}$. In accordance with the fact that signal power equals 1, we constrain $\int_{-\frac{1}{2}}^{\frac{1}{2}} P(f; \boldsymbol{\theta})df = 1$, or equivalently, the autocorrelation satisfies $r[0] = 1$.

A. The Statistical Properties of Misspecified MLE

For large data records, the asymptotic Gaussian log likelihood function is [18]

$$l_M = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\ln P(f;\boldsymbol{\theta}) + \frac{I(f)}{P(f;\boldsymbol{\theta})}\right) df \quad (1)$$

If $E_{\boldsymbol{\theta^*}}(l_M)$ exists, where $E_{\boldsymbol{\theta^*}}$ represents the expected value with respect to the true model, then we define $\boldsymbol{\theta_0}$ to be the one that maximizes $E_{\boldsymbol{\theta^*}}(l_M)$. The following theorems are valid under the assumption that $\{\boldsymbol{c}\cdot\frac{\partial l_M}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\}$ satisfies the Lyapunov condition for any $p\times 1$ vector $\boldsymbol{c}$ at any frequency.

Theorem 1. *The estimator $\hat{\boldsymbol{\theta}}$ that maximizes* (1) *is asymptotically normally distributed with mean $\boldsymbol{\theta_0}$ and covariance matrix $\boldsymbol{A}^{-1}(\boldsymbol{\theta_0})\boldsymbol{B}(\boldsymbol{\theta_0})\boldsymbol{A}^{-T}(\boldsymbol{\theta_0})$, i.e.*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \overset{a}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}(\boldsymbol{\theta_0})^{-1}\boldsymbol{B}(\boldsymbol{\theta_0})\boldsymbol{A}(\boldsymbol{\theta_0})^{-T})$$

where

$$\begin{aligned}[\boldsymbol{A}(\boldsymbol{\theta_0})]_{ul} =& \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial^2 \ln P(f;\boldsymbol{\theta})}{\partial\theta_u\partial\theta_l}\left(1-\frac{Q(f;\boldsymbol{\theta^*})}{P(f;\boldsymbol{\theta})}\right)\right. \\ &\left. + \frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_u}\frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_l}\frac{Q(f;\boldsymbol{\theta^*})}{P(f;\boldsymbol{\theta})}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} df\end{aligned}$$

$$[\boldsymbol{B}(\boldsymbol{\theta_0})]_{ul} = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_u}\frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_l}\frac{Q^2(f;\boldsymbol{\theta^*})}{P^2(f;\boldsymbol{\theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} df$$

$[\cdot]_{ul}$ denotes the elements at row u column l. The proof is given in supplementary material S.1.

Corollary 1.1. *If the proposed model is the correct one, then $Q(f;\boldsymbol{\theta^*}) = P(f;\boldsymbol{\theta_0})$. It follows that*

$$\boldsymbol{A}(\boldsymbol{\theta^*}) = \boldsymbol{B}(\boldsymbol{\theta^*})$$

and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta^*}) \overset{a}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}(\boldsymbol{\theta^*})^{-1})$$

where

$$[\boldsymbol{A}(\boldsymbol{\theta^*})]_{ul} = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_u}\frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_l}|_{\boldsymbol{\theta}=\boldsymbol{\theta^*}} df$$

This result agrees with the asymptotic CRLB [5] [19] and implies that for large data records the MLE estimator is the one that has the minimum variance among all estimators.

Corollary 1.2. *In the case of a scalar parameter, the quasi-MLE $\hat{\theta}$ is asymptotically normally distributed with mean θ_0 and variance σ^2, or*

$$\sqrt{N}(\hat{\theta} - \theta_0) \overset{a}{\sim} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \frac{2\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{Q(f;\theta^*)}{P(f;\theta)}\frac{\partial \ln P(f;\theta)}{\partial\theta}\right)^2 |_{\theta=\theta_0} df}{\left(\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\frac{\partial^2 \ln P(f;\theta)}{\partial\theta^2}(1-\frac{Q(f;\theta^*)}{P(f;\theta)}) + \left(\frac{\partial \ln P(f;\theta)}{\partial\theta}\right)^2 \frac{Q(f;\theta^*)}{P(f;\theta)}\right)|_{\theta=\theta_0} df\right)^2}$$

Corollary 1.3. *In the case of a scalar parameter, if the proposed model is correct, so that $Q(f;\theta^*) = P(f;\theta_0)$, then the MLE $\hat{\theta}$ is asymptotically normally distributed with mean θ^* and variance σ^2, or*

$$\sqrt{N}(\hat{\theta} - \theta^*) \overset{a}{\sim} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \frac{2}{\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial \ln P(f;\theta)}{\partial\theta}\right)^2 |_{\theta=\theta^*} df}$$

B. The Statistical Properties of RITE

The RITE likelihood function is

$$l_R = \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)\ln P(f;\boldsymbol{\theta}) df \quad (2)$$

Assume $E_{\boldsymbol{\theta^*}}(l_R)$ exists, here we define $\boldsymbol{\theta_0}$ to be the one that maximizes $E_{\boldsymbol{\theta^*}}(l_R)$. The following theorems are valid under the assumption that $\{\boldsymbol{c}\cdot\frac{\partial l_R}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\}$ satisfies the Lyapunov condition for any $p\times 1$ vector $\boldsymbol{c}$ at any frequency.

Theorem 2. *The estimator $\hat{\boldsymbol{\theta}}$ that maximizes* (2) *is asymptotically normally distributed with mean $\boldsymbol{\theta_0}$ and covariance matrix $\boldsymbol{A}^{-1}(\boldsymbol{\theta_0})\boldsymbol{B}(\boldsymbol{\theta_0})\boldsymbol{A}^{-T}(\boldsymbol{\theta_0})$, i.e.*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \overset{a}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}(\boldsymbol{\theta_0})^{-1}\boldsymbol{B}(\boldsymbol{\theta_0})\boldsymbol{A}(\boldsymbol{\theta_0})^{-T})$$

where

$$[\boldsymbol{A}(\boldsymbol{\theta_0})]_{ul} = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial^2 \ln P(f;\boldsymbol{\theta})}{\partial\theta_u\partial\theta_l}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} Q(f;\boldsymbol{\theta^*}) df$$

$$[\boldsymbol{B}(\boldsymbol{\theta_0})]_{ul} = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_u}\frac{\partial \ln P(f;\boldsymbol{\theta})}{\partial\theta_l}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} Q^2(f;\boldsymbol{\theta^*}) df$$

The proof is given in supplementary material S.2. When model is correct, unlike the MLE case, the equality $\mathbf{A}(\boldsymbol{\theta_0}) = \mathbf{B}(\boldsymbol{\theta_0})$ does not hold. Therefore, the expressions for a correct model cannot be simplified.

Corollary 2.1. *In the case of a scalar parameter, RITE $\hat{\theta}$ is asymptotically normally distributed with mean θ_0 and variance σ^2, or*

$$\sqrt{N}(\hat{\theta} - \theta_0) \overset{a}{\sim} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \frac{2\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(Q(f;\theta^*)\frac{\partial \ln P(f;\theta)}{\partial\theta}\right)^2 |_{\theta=\theta_0} df}{\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} Q(f;\theta^*)\frac{\partial^2 \ln P(f;\theta)}{\partial\theta^2}|_{\theta=\theta_0} df\right)^2}$$

IV. Spectral Estimation Application with AR Model

We assume an AR Gaussian process $s[n]$

$$s[n] = -\sum_{k=1}^{p} a[k]s[n-k] + u[n]$$

where $u[n]$ is the driving noise of the model with variance σ_u^2, p is the order of the AR process, and $a[k]$ is the k^{th} AR coefficient. The AR PSD $P(f)$ is [6]:

$$\frac{\sigma_u^2}{|1+a[1]\exp(-j2\pi f)+\cdots+a[p]\exp(-j2\pi fp)|^2} \tag{3}$$

As we stated in Section III, the PSD is restricted to yield a power of 1, so σ_u^2 is not actually a parameter, but a function of $a[1], a[2], \cdots, a[p]$.

Alternatively, an AR process can be expressed by $r[0]$, which in our case equals 1, and the reflection coefficients $k_1, k_2, \cdots, k_p$, which are restricted in $(-1, 1)$ to guarantee a stable process. The Levinson algorithm transfers the reflection coefficients to the AR parameters [6]. In our case, since $r[0] = 1$, we modify the Levinson algorithm as follows:
For $j = 1$

$$a_1[1] = k_1$$

For $j = 2, 3, \cdots, p$

$$a_j[i] = \begin{cases} a_{j-1}[i] + k_j a_{j-1}[j-i] & \text{for } i = 1, 2, \cdots, j-1 \\ k_j & \text{for } i = j \end{cases}$$

$$\sigma_u^2 = \prod_{i=1}^{p}(1-k_i^2)$$

where $a_p[i]$'s are the AR parameter $a[i]$'s. Since the general Levinson algorithm has $\sigma_u^2 = r[0]\prod_{i=1}^{p}(1-k_i^2)$ while here it is $\sigma_u^2 = \prod_{i=1}^{p}(1-k_i^2)$, we denote the above approach as the modified Levinson algorithm.

A. White Noise Sensitivity of Likelihood Function

Many existing AR spectral estimators (Burg method, covariance method, etc.) are approximate MLEs. They are unbiased and have minimum variances *if there is no modeling error*. An important problem is their sensitivity to observation noise. The effect of noise flattens the estimated PSD and reduces the resolution. This is due to the severe bias of the misspecified, i.e., quasi-MLE. However, RITE is shown to have less bias than the quasi-MLE when white noise is present, and this results in improved resolution.

If white noise $w[n]$ is present in addition to the signal $s[n]$, then the data is $x[n] = s[n] + w[n]$. We assume that $w[n]$ is independent of $s[n]$. The true PSD $Q(f)$ of the observed data becomes

$$Q(f) = P(f) + \sigma_w^2$$

where σ_w^2 is the variance of the observation noise. To analyze how robust the estimator is, we could try analyzing how the white noise affects the likelihood function. By taking the expected value of the likelihood function of RITE, we get from (2)

$$\begin{aligned} E_{\boldsymbol{\theta}^*}(l_R) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} (P(f;\boldsymbol{\theta}^*) + \sigma_w^2)\ln P(f;\boldsymbol{\theta})df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P(f;\boldsymbol{\theta}^*)\ln P(f;\boldsymbol{\theta})df + \int_{-\frac{1}{2}}^{\frac{1}{2}} \sigma_w^2 \ln\frac{\sigma_u^2}{|A(f)|^2}df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P(f;\boldsymbol{\theta}^*)\ln P(f;\boldsymbol{\theta})df + \int_{-\frac{1}{2}}^{\frac{1}{2}} \sigma_w^2 \ln \sigma_u^2 df \\ &\quad - \int_{-\frac{1}{2}}^{\frac{1}{2}} \sigma_w^2 \ln |A(f)|^2 df \end{aligned}$$

where $\boldsymbol{\theta} = [a[1],\ a[2], \cdots, a[p]]^T$ and $A(f)$ is the Fourier transform of $[1,\ a[1], \cdots,\ a[p]]$. Since a stable AR process has all its poles inside the unit circle, $A(f)$ is minimum phase [6] and

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln|A(f)|^2 df = 0$$

which leads to

$$\begin{aligned} E_{\boldsymbol{\theta}^*}(l_R) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P(f;\boldsymbol{\theta}^*)\ln P(f;\boldsymbol{\theta})df + \int_{-\frac{1}{2}}^{\frac{1}{2}} \sigma_w^2 \ln \sigma_u^2 df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P(f;\boldsymbol{\theta}^*)\ln P(f;\boldsymbol{\theta})df \\ &\quad + \sigma_w^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln(\prod_{i=1}^{p}(1-k_i^2))df \end{aligned}$$

Here k_i is the true reflection coefficient. Taking the derivative with respect to σ_w^2, we get

$$\frac{\partial E_{\boldsymbol{\theta}^*}(l_R)}{\partial \sigma_w^2} = \ln(\prod_{i=1}^{p}(1-k_i^2))$$

If we do the same operations to l_M, we have from (1)

$$\begin{aligned} E_{\boldsymbol{\theta}^*}(l_M) &= -\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P(f;\boldsymbol{\theta}) + \frac{P(f;\boldsymbol{\theta}^*) + \sigma_w^2}{P(f;\boldsymbol{\theta})}df \\ &= -\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P(f;\boldsymbol{\theta}) + \frac{P(f;\boldsymbol{\theta}^*)}{P(f;\boldsymbol{\theta})}df \\ &\quad - \sigma_w^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{P(f;\boldsymbol{\theta})}df \\ &= -\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P(f;\boldsymbol{\theta}) + \frac{P(f;\boldsymbol{\theta}^*)}{P(f;\boldsymbol{\theta})}df \\ &\quad - \frac{\sigma_w^2}{\sigma_u^2}\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 df \end{aligned}$$

Hence,

$$\frac{\partial E_{\boldsymbol{\theta}^*}(l_M)}{\partial \sigma_w^2} = -\frac{1}{\sigma_u^2}\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 df$$

By Parseval's theorem, $\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 df = 1+\sum_{i=1}^{p} a^2[i]$. Thus,

$$\begin{aligned}\frac{\partial E_{\boldsymbol{\theta}^*}(l_M)}{\partial \sigma_w^2} &= -\frac{1}{\sigma_u^2}\left(1+\sum_{i=1}^{p} a^2[i]\right)\\ &< -\frac{1}{\sigma_u^2} = -\prod_{i=1}^{p}\frac{1}{1-k_i^2}\end{aligned}$$

Since $\prod_{i=1}^{p}\frac{1}{1-k_i^2} > 1$, we have

$$\begin{aligned}\left|\frac{\partial E_{\boldsymbol{\theta}^*}(l_M)}{\partial \sigma_w^2}\right| &> \prod_{i=1}^{p}\frac{1}{1-k_i^2}\\ &> \ln\left(\prod_{i=1}^{p}\frac{1}{1-k_i^2}\right) = \left|\ln\prod_{i=1}^{p}(1-k_i^2)\right|\\ &= \left|\frac{\partial E_{\boldsymbol{\theta}^*}(l_R)}{\partial \sigma_w^2}\right|\end{aligned}$$

Consider a narrow-band process, for which the k_i's may be close to 1. The closer they are to 1, the larger are $\left|\frac{\partial E_{\boldsymbol{\theta}^*}(l_R)}{\partial \sigma_w^2}\right|$ and $\left|\frac{\partial E_{\boldsymbol{\theta}^*}(l_M)}{\partial \sigma_w^2}\right|$, and the more effect the noise has. The MLE likelihood function is more severely affected by noise when the AR random process is narrow-band. As an example, take the AR(4) process that we employed in the next section, for

$$[k_1, k_2, k_3, k_4] = [-0.71, 0.98, -0.70, 0.93]$$

we have that

$$\begin{aligned}\left|\frac{\partial E_{\boldsymbol{\theta}^*}(l_R)}{\partial \sigma_w^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} &= 6.6\\ \left|\frac{\partial E_{\boldsymbol{\theta}^*}(l_M)}{\partial \sigma_w^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} &= 2.2\times 10^4\end{aligned}$$

This is a sensitivity difference of several orders of magnitude.

V. Simulation Examples

We consider spectral estimation of an AR process in noise to test our estimator. White Gaussian noise (WGN), white mixture Gaussian noise and white Laplacian noise give similar results for both RITE and asymptotic MLE. In the case of IID impulsive noise modeled by α-Stable noise, RITE does not perform as well as in white noise case, but still outperforms the asymptotic MLE. In this section therefore, we present only the results for WGN. The other simulations are included in supplemental material S.4. In the simulation, we assume the AR model order is known. Thus $a[1],\ a[2],\cdots,a[p]$ are the p parameters to be estimated. Alternatively, we can estimate the reflection coefficients $k_1, k_2, \cdots, k_p$.

No analytical solution is available for RITE with AR model. To find the global maximum of l_R, one option is a grid search. Since the reflection coefficients are guaranteed to give a valid AR process, we do the search over the reflection coefficients. The estimation procedure is:
a) Create a p-dimensional grid with each dimension in the range $(-1, 1)$.
b) Assign a value from the gridded domain to the reflection coefficients.
c) Transform the reflection coefficients to the AR parameters by using the modified Levision algorithm.
d) Plug the AR parameters into (3) to get the PSD.
e) Plug the PSD into (2) and get the value of l_R.
f) Repeat b) to e) over the valid grid and find the one that maximizes l_R.
g) The final estimated PSD is obtained by repeating c) to e) with the solution of the reflection coefficients found in f).

If the grid is fine enough, then the solution of a grid search should be very close to the global maximum. This method is recommended when p is small. However, as p increases, a grid search suffers the curse of dimensionality. Hence we recommend an alternative option, using the Matlab optimization toolbox function *fmincon*, which is a gradient-based method which finds the local minimum of an objective function with constraints. In our procedure, we
a) Create a function that transfers the reflection coefficients to the objective function $-l_R$.
b) Use a standard AR spectral estimator (Burg method, co-variance method, etc.), to get an estimate of the reflection coefficients and assign it as the initial value.
c) Given the above function, the constraints $-1 < k_i < 1$, and the proper initial value, *fmincon* outputs the solution of a *local minimum* near the initial value.
d) Transfer the solution of the reflection coefficients found in c) to the estimated PSD (like the step g in grid search procedure).

For a higher order AR process, this method is more efficient, but it only gives the local minimum. Hopefully, with a proper initial value, this local solution will also yield the true global solution.

Next we use a grid search for an AR(1) example and *fmincon* for an AR(4) example.

A. AR(1) Process Example

The PSD of an AR(1) process is

$$P(f) = \frac{\sigma_u^2}{1+2a[1]\cos(2\pi f)+a^2[1]}$$

Note that our restriction $r[0] = 1$ leads to $\sigma_u^2 = 1 - a^2[1]$. Hence

$$P(f) = \frac{1-a^2[1]}{1+2a[1]\cos(2\pi f)+a^2[1]}$$

We generated a Gaussian AR(1) process with $a[1] = -0.9$. If the observed process is not corrupted by noise, then the MLE (stands for the asymptotic MLE in this section) and RITE are unbiased estimators but with different variances. In this case, MLE is the optimal estimator since it has a smaller variance and attains the CRLB. It is proved that the RITE variance is larger than the MLE variance as shown in supplementary material S.3. By Corollary 1.3 and Corollary 2.1, the MLE has variance:

$$\frac{1}{\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\frac{\partial \ln P(f;\theta)}{\partial \theta}\right)^2\Big|_{\theta=\theta_0} df}$$

and the RITE variance is:

$$\frac{\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(P(f;\theta^*) \frac{\partial \ln P(f;\theta)}{\partial \theta} \right)^2 |_{\theta=\theta_0} df}{\frac{N}{2} \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} P(f;\theta^*) \frac{\partial^2 \ln P(f;\theta)}{\partial \theta^2} |_{\theta=\theta_0} df \right)^2}$$

where

$$\theta = a[1]$$

$$\theta_0 = \theta^* = -0.9$$

$$\frac{\partial \ln P(f;\theta)}{\partial \theta} = -\frac{2\theta}{1-\theta^2} - \frac{2\theta + 2\cos(2\pi f)}{1+\theta^2+2\theta\cos(2\pi f)} \quad (4)$$

$$\begin{aligned}\frac{\partial^2 \ln P(f;\theta)}{\partial \theta^2} &= -\frac{1}{(1-\theta)^2} - \frac{1}{(1+\theta)^2} \\ &- \frac{2}{1+\theta^2+2\theta\cos(2\pi f)} \\ &+ \frac{4\cos(2\pi f)+4\theta}{(1+\theta^2+2\theta\cos(2\pi f))^2}\end{aligned} \quad (5)$$

To be fair, RITE and MLE are both calculated using a fine grid search. The theoretical $N\times$variance vs N values are plotted in solid lines as Fig. 1(a). Simulated results are shown as circles.

When the observed data is embedded in WGN, MLE and RITE converge to means other than the true value. Thus to compare the performance of the two estimators, we need to evaluate the mean square error (MSE), which equals the variance plus the squared bias. The theoretical variance is computed by using Corollary 1.2 and Corollary 2.1, where $\theta^* = -0.9$. By definition, θ_0 is the value that maximizes the expected value of the likelihood function. $\frac{\partial \ln P(f;\theta)}{\partial \theta}$, $\frac{\partial^2 \ln P(f;\theta)}{\partial \theta^2}$ are listed in (4), (5), and $Q(f;\theta^*) = P(f;\theta^*) + \sigma_w^2$. As illustrated in Fig. 1(b), the MSE of RITE grows slower than the one of the quasi-MLE and this advantage increases as the SNR decreases. As shown in Fig. 1(c), although the quasi-MLE has a smaller variance, the bias weakens its performance for low SNR range. Therefore, when the squared bias exceeds the variance, RITE exhibits more noise robustness than the quasi-MLE. This example verifies the Theorems 1 and 2, at least for an AR(1) process in WGN.

B. AR(4) Process Example

1) Burg and RITE AR Spectral Estimator: The Burg method is approximately an MLE and therefore is asymptotically unbiased with variance that attains the CRLB. It has been shown to have good resolution for a narrow-band PSD when the data is not noise corrupted [6]. Similar to our approach of computing RITE, the Burg method first estimates reflection coefficients, and then calculates the AR parameters by the Levinson algorithm. For a fair comparison between Burg method and RITE, instead of using the Levinson algorithm, we use the modified Levinson algorithm, described in Section IV, to obtain the Burg estimation. For RITE estimation, instead of performing a grid search, which requires a high computational cost, we use the Matlab *fmincon* function with the reflection coefficients estimated by the Burg method as our initial point to hopefully find the global maximum solution.

2) Simulations: The AR(4) parameters are set to be

$$[a[1], a[2], a[3], a[4]] = [-2.7428, 3.7906, -2.6454, 0.93]$$

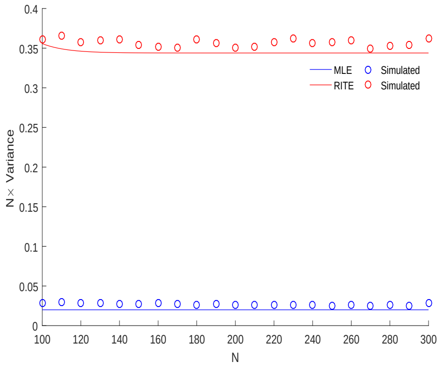
or equivalently,

$$[k_1, k_2, k_3, k_4] = [-0.71, 0.98, -0.70, 0.93]$$

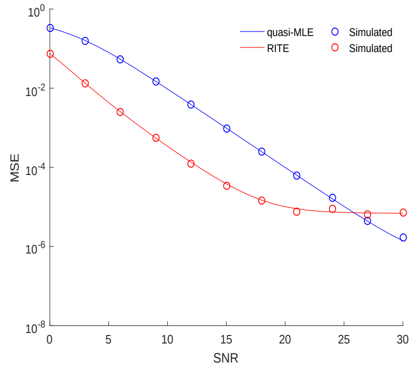
The data length N is 350. Results are presented in Figs. 2 to 5. The studies illustrated in the figures demonstrate that the Burg method has less variance than RITE, but when WGN is present, there is a large bias. For SNR$= 40dB$, the resolution of the Burg method (Figs. 4(a), 5(a)) is degraded due to noise, while RITE (Figs. 2(a), 3(a)) has very good resolution. When SNR$= 35, 15dB$, the Burg method (Figs. 4(b), 4(c), 5(b), 5(c)) is unable to resolve the two peaks. As shown in the Figs. 2(b) and 3(b), RITE is not clearly affected as SNR decreases to $35dB$. Even if the SNR is reduced to $15dB$ (Figs. 2(c), 3(c)), RITE still produces some estimates with good resolution. The overlaid plots show that RITE has more variance than Burg. However, the average plots verify that RITE has less bias and higher resolution, demonstrating that RITE is indeed more robust to WGN as compared to the Burg method. However, a potential problem of using *fmincon* to find RITE is that, the iterative optimization may only produce a local maximum and not the true RITE. Actually, most of the poor estimates for RITE (flattened PSDs) are due to local minima, as evidenced by results of a fine grid search which yield larger values of the likelihood. As an example, there is a single outlier, which is only a local maximum, as shown in Fig. 6(a) when $SNR = 30dB$ and $N = 300$. This outlier $\mathbf{k}_o = [-0.713, 0.908, -0.221, 0.186]$ produces $l_R = 0.78$, while another possible solution, found by *fmincon* with $[-0.7, 0.7, -0.7, 0.7]$ as the initial value, $\mathbf{k}_g = [-0.724, 0.968, -0.689, 0.698]$ has $l_R = 0.82$. Therefore the outlier is not the true RITE. The estimation results generated by $\mathbf{k}_o$ and $\mathbf{k}_g$ are shown in Fig. 6(b).

VI. Conclusion

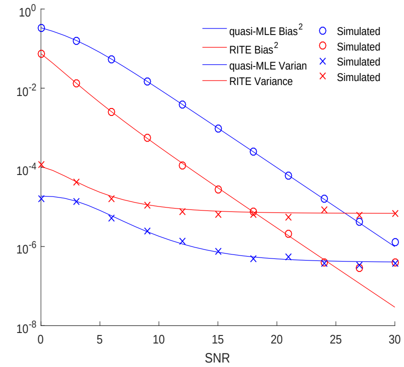
We have introduced RITE as a new method for PSD estimation. RITE and the quasi-MLE are compared analytically when the data is misspecified. In particular, the misspecification example of additive noise is studied in detail. Theoretical results have been verified via simulation for a Gaussian AR(1) process. Examples employed in this study demonstrate that RITE is indeed more robust than the quasi-MLE when WGN is present, resulting in a smaller bias. This improvement has been demonstrated for AR spectral estimation when observation noise is present. In this study, higher order AR examples had to rely on an iterative search algorithm to find the global maximum. It is not clear if this was attained. These studies have established a solid foundation to further our goal of searching for the global maximum which is the true RITE, and which may have an even better performance. Therefore, an efficient method of computing RITE will be explored in future works. It should be emphasized that RITE is a general approach to model-based spectral estimation in the presence of model inaccuracies. Its robustness properties need to be explored for other scenarios in which data models are inaccurate, which is the ”rule rather than the exception”.



(a) $N\times$ Variance

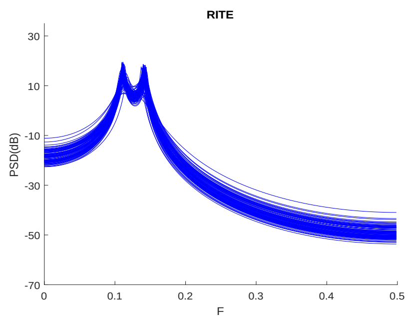


(b) Mean Square Error, N=50000

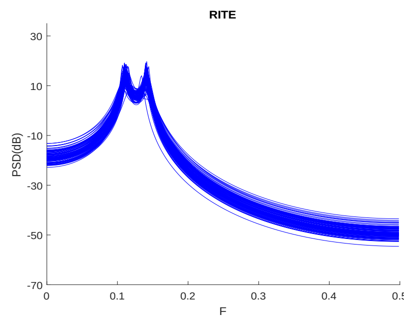


(c) Variance and Squared Bias, N=50000

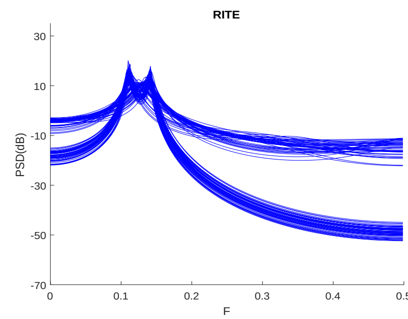
Fig. 1: AR(1) Example



(a) SNR=40dB

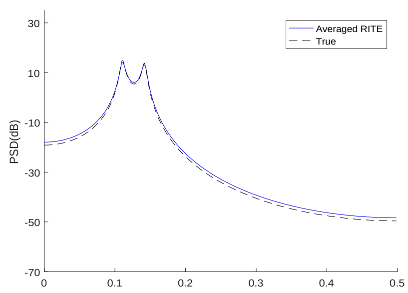


(b) SNR=35dB

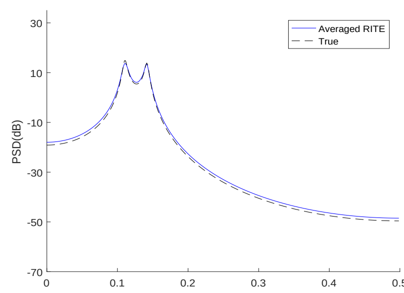


(c) SNR=15dB

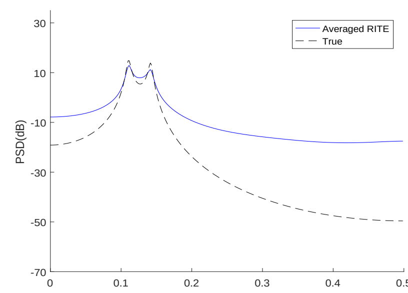
Fig. 2: 100 Overlaid RITE Realizations



(a) SNR=40dB



(b) SNR=35dB



(c) SNR=15dB

Fig. 3: Average of RITE Realizations

References

[1] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*. Springer Science & Business Media, 2013.

[2] S. Kay, "Poisson maximum likelihood spectral inference." unpublished http://www.ele.uri.edu/faculty/kay/New%20web/Books.htm, 2017.

[3] G. M. Jenkins and D. G. Watts, "Spectral analysis," 1968.

[4] S. Kay and J. Makhoul, "On the statistics of the estimated reflection coefficients of an autoregressive process," *IEEE transactions on acoustics, speech, and signal processing*, vol. 31, no. 6, pp. 1447–1455, 1983.

[5] H. Cramir, "Mathematical methods of statistics," *Princeton U. Press, Princeton*, 1946.

[6] S. M. Kay, *Modern spectral estimation*. Pearson Education India, 1988.

[7] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, no. 4, pp. 661–675, 1971.

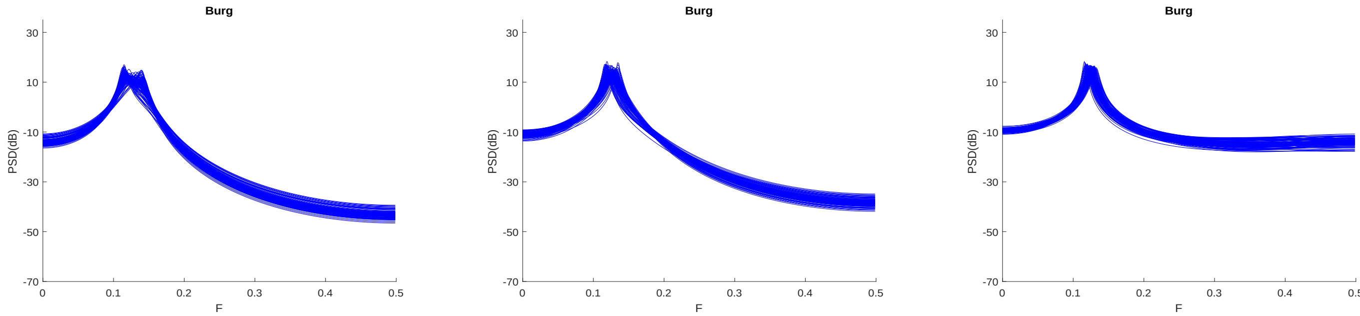
[8] W. Chen and G. Stegen, "Experiments with maximum entropy power spectra of sinusoids," *Journal of Geophysical Research*, vol. 79, no. 20, pp. 3019–3022, 1974.

[9] L. Marple, "Resolution of conventional fourier, autoregressive, and special arma methods of spectrum analysis," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'77.*, vol. 2, pp. 74–77, IEEE, 1977.

[10] J. Cadzow, "High performance spectral estimation–a new arma method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 5, pp. 524–529, 1980.

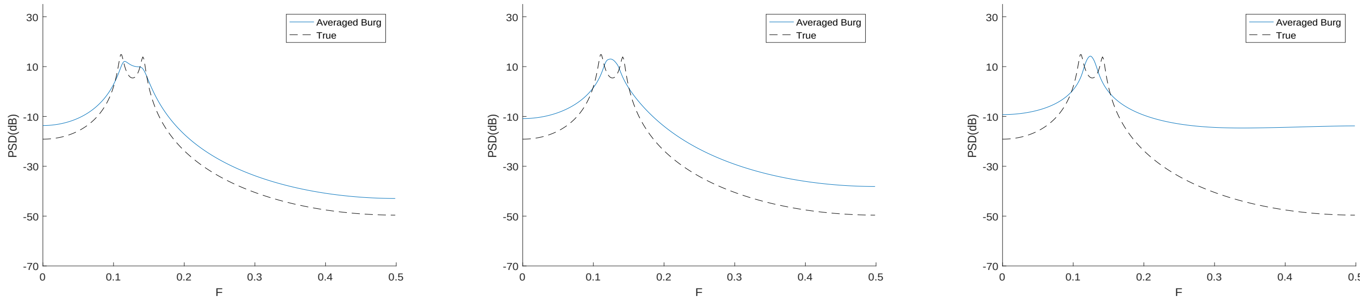
[11] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society*, pp. 1–25, 1982.

[12] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Processing Magazine*,



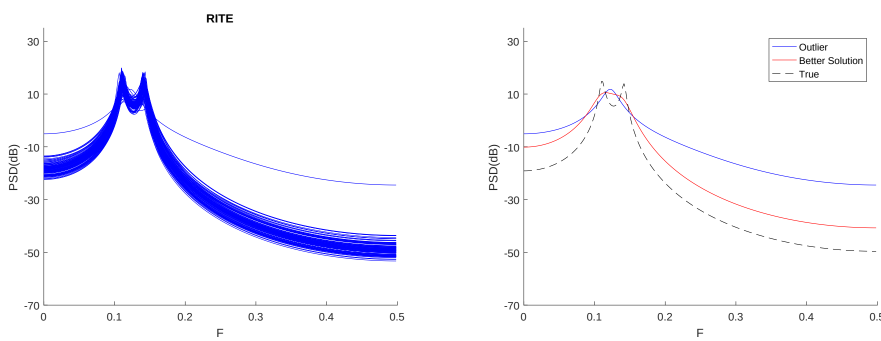
(a) SNR=40dB (b) SNR=35dB (c) SNR=15dB

Fig. 4: 100 Overlaid Burg Realizations



(a) SNR=40dB (b) SNR=35dB (c) SNR=15dB

Fig. 5: Average of Burg Realizations



(a) 100 Overlaid Realizations. (b) Outlier and Better Solution

Fig. 6: SNR=30dB, N=300

vol. 34, no. 6, pp. 142–157, 2017.

[13] J. F. C. Kingman, *Poisson processes*. Wiley Online Library, 1993.

[14] M. Grigoriu, “A spectral representation based model for monte carlo simulation,” *Probabilistic Engineering Mechanics*, vol. 15, no. 4, pp. 365–370, 2000.

[15] S. Kay, “Representation and generation of non-gaussian wide-sense stationary random processes with arbitrary psds and a class of pdfs,” *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3448–3458, 2010.

[16] P. J. Huber *et al.*, “Robust estimation of a location parameter,” *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[17] L. A. Stefanski and D. D. Boos, “The calculus of m-estimation,” *The American Statistician*, vol. 56, no. 1, pp. 29–38, 2002.

[18] S. M. Kay, “Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1),” *PTR Prentice-Hall, Englewood Cliffs*, 1993.

[19] C. R. Rao, “Information and the accuracy attainable in the estimation of statistical parameters,” in *Breakthroughs in statistics*, pp. 235–247, Springer, 1992.