ON THE USE OF KOLMOGOROV STRUCTURE FUNCTION
FOR PERIODOGRAM SMOOTHING

Ciprian Doru Giurcăneanu and Seyed Alireza Razavi

Tampere University of Technology, Department of Signal Processing
P.O. Box 553, FIN-33101 Tampere, Finland
ciprian.giurcaneanu, alireza.razavi@tut.fi

ABSTRACT

In a recent series of papers, it was shown how the periodogram can be smoothed by thresholding the estimated cepstral coefficients either with a carefully designed uniformly most powerful unbiased test (UMPUT), or with the Bayesian information criterion (BIC). In this paper, we devise a fully automatic scheme that selects the threshold by using the Kolmogorov structure function (KSF). For the numerical examples taken from the previous literature, the newly proposed method compares favorably with the existing schemes.

Index Terms— Kolmogorov structure function, periodogram smoothing, total-variance reduction, thresholding, cepstral statistics.

1. INTRODUCTION AND PRELIMINARIES

One of the classical problems in signal processing is to estimate the spectrum $\Phi(\omega)$ of a stationary, discrete time, real-valued signal from the measurements $y_0,\ldots,y_{N-1}$. With the convention that $\omega_p = (2\pi p)/N$, $p \in \{0,\ldots,N-1\}$, are the Fourier frequency grid points, we use the notation Φ_p for $\Phi(\omega_p)$. The estimate of the spectrum at point ω_p is $\hat{\Phi}_p = \frac{1}{N}\left|\sum_{t=0}^{N-1} y_t \exp(-i\omega_p t)\right|^2$, where $i = \sqrt{-1}$ [1]. Because the main drawback is the high variance of $\hat{\Phi}_p$, various methods for smoothing the periodogram have been proposed in the previous literature. In a recent series of papers (see [2, 3] and the references therein), it was introduced a novel approach that is based on thresholding the estimated cepstral coefficients.

We assume that N, the number of samples, is even and we take $M = N/2 + 1$. Under the hypothesis that $\min(\Phi_p, \hat{\Phi}_p) > 0$ for all p, the first M cepstral coefficients and their estimates are given by $c_j = \frac{1}{N}\sum_{p=0}^{N-1} \ln(\Phi_p)\exp(i\omega_j p)$, $\hat{c}_j = \frac{1}{N}\sum_{p=0}^{N-1} \ln(\hat{\Phi}_p)\exp(i\omega_j p) + \gamma\delta_{j,0}$, where $j \in \{0,\ldots,M-1\}$ and $\gamma = 0.577216\ldots$ is the Euler constant [3]. The Kronecker indicator $\delta_{j,0}$ takes value one if $j = 0$, and otherwise takes value zero. The rest of the coefficients can be obtained without difficulties because $c_{N-j} = c_j$ and $\hat{c}_{N-j} = \hat{c}_j$ for $j \in \{1,\ldots,M-2\}$. In [3], it was shown that

$$\frac{1}{N}\sum_{p=0}^{N-1} E\left[(\ln\hat{\Phi}_p - \ln\Phi_p)^2\right] = \sum_{j=0}^{N-1} E\left[(\hat{c}_j - c_j)^2\right], \quad (1)$$

where $E[\cdot]$ denotes the expectation operator. The mirror symmetry of the cepstral coefficients implies for a given $\mathbf{c} = [c_0 \ldots c_{M-1}]^\top$ that

the total variance (TV) term in (1) depends only on the entries of $\hat{\mathbf{c}} = [\hat{c}_0 \ldots \hat{c}_{M-1}]^\top$. To emphasis this aspect, we write

$$\text{TV}(\hat{\mathbf{c}}) = \sum_{j=0}^{N-1} E\left[(\hat{c}_j - c_j)^2\right]. \quad (2)$$

More importantly, the identity in (1) can be exploited to smooth the periodogram. For instance, if one transforms the vector of coefficients $\hat{\mathbf{c}}$ to $\check{\mathbf{c}} = [\check{c}_0 \ldots \check{c}_{M-1}]^\top$ such that $\check{c}_j = 0$ whenever c_j is small and $\check{c}_j = \hat{c}_j$ otherwise, then it is likely that the total variance

$$\text{TV}(\check{\mathbf{c}}) = \sum_{j=0}^{N-1} E\left[(\check{c}_j - c_j)^2\right] \quad (3)$$

is smaller than $\text{TV}(\hat{\mathbf{c}})$. This makes also the average variance of $\ln(\check{\Phi}_p) = \sum_{j=0}^{N-1} \check{c}_j \exp(-i\omega_p j)$ to be smaller than that of $\ln(\hat{\Phi}_p)$ for $p \in \{0,\ldots,N-1\}$. The following result is instrumental in deciding which are the coefficients $\hat{c}_j$ that must be turned to zero.

Theorem 1.1. *When $N \gg 1$, the normalized vector of estimation errors $\sqrt{N}(\hat{\mathbf{c}} - \mathbf{c})$ converges in distribution to the normal distribution of mean zero and covariance matrix* $\mathbf{C} = \frac{\pi^2}{6}\begin{bmatrix} 2 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & & & & 2 \end{bmatrix}$.

The interested reader can find in [2] a simplified proof of Theorem 1.1 along with some historical notes and a list of papers containing proofs for the same result. Based on the distributional properties of the estimated $\hat{\mathbf{c}}$, a thresholding procedure was proposed in [3]:

$$\check{c}_j = \begin{cases} 0, & |\hat{c}_j| < \mu\left[\mathbf{C}(j+1, j+1)/N\right]^{1/2}, \\ \hat{c}_j, & \text{otherwise,} \end{cases} \quad (4)$$

where $j \in \{0,\ldots,M-1\}$ and $\mathbf{C}(j+1, j+1)$ is the $(j+1)$-th diagonal element of $\mathbf{C}$. Two different formulas have been used for μ. The first one was derived by combining a uniformly most powerful unbiased test (UMPUT) [4] with some empirical evidence, and has the expression:

$$\mu_{\text{UMPUT}} = (5 - \text{I}_{\text{st}}) + \frac{N - 128}{1920}. \quad (5)$$

Its application is restricted to data sets for which N is an integer power of two and $128 \leq N \leq 2048$. The value of I_{st} depends on the type of the signal and is chosen as follows: $\text{I}_{\text{st}} = 1$ for broadband signal with small dynamic range, $\text{I}_{\text{st}} = 2$ for broadband signal with

medium dynamic range, and $I_{st} = 3$ for narrowband signal with large dynamic range. The second μ-formula was inspired by the Bayesian information criterion (BIC) from [5], and is given by:

$$\mu_{\text{BIC}} = 1 + (\ln M)^{1/2}. \tag{6}$$

In this paper, we show how μ can be selected by using the Kolmogorov structure function (KSF) [6]. To this end, we derive in Section 2 the expression of the KSF for the considered problem by following closely the approach from [7]. After giving in Section 2 the algorithmic steps for finding μ_{KSF}, we compare in Section 3, for simulated data, the performance of various thresholding schemes.

2. KSF AND SELECTION OF THE PARAMETER μ

We note that turning to zero some of the coefficients $\{\hat{c}_j\}_{j=0}^{M-1}$ is similar to a quantization process. This observation makes the connection between the thresholding procedure defined in (4) and the KSF whose expression contains two terms. The first one is given by the code length for the quantized values of $\{\hat{c}_j\}_{j=0}^{M-1}$, while the second one accounts for the distortion produced by the quantization process. We aim to select the threshold value μ_{KSF} which minimizes the KSF.

The first step is to partition the space of the cepstral coefficients. Theorem 1.1 allows us to perform this task by applying the procedure introduced in [7], where Rissanen has considered the following general framework. Let the entries of θ be the true values of the parameters, and $\hat{\theta}$ the vector of their estimates obtained from N observations. It is assumed that the number of parameters is not greater than the number of samples. Additionally, for $N \gg 1$, the random vector $\sqrt{N}(\hat{\theta} - \theta)$ converges in distribution to the normal distribution of mean zero and covariance $\mathbf{J}^{-1}(\theta)$, where $\mathbf{J}(\theta)$ is the Fisher information matrix [1] divided by N. We choose an arbitrary point, $\bar{\theta}$, in the parameter space such that $\mathbf{J}(\bar{\theta})$ is non-singular. A box centered at $\bar{\theta}$ is defined as follows: (i) consider the ellipsoid $(\theta - \bar{\theta})^{\top} \mathbf{J}(\bar{\theta})(\theta - \bar{\theta}) = d/N$, where $d > 0$ is a parameter; (ii) select the largest box within the ellipsoid. The motivation is straightforward: it is possible to partition the parameter space by using boxes, but not by using ellipsoids.

Inspired by the approach from [7], we assign to an arbitrary point, say $\bar{\mathbf{c}}$, from the cepstral coefficients space the ellipsoid given by:

$$(\mathbf{c} - \bar{\mathbf{c}})^{\top} \mathbf{C}^{-1} (\mathbf{c} - \bar{\mathbf{c}}) = d/N. \tag{7}$$

For $j \in \{0, \ldots, M-1\}$, we define $x_j = c_j [N/\mathbf{C}(j+1, j+1)]^{1/2}$ and $\bar{x}_j = \bar{c}_j [N/\mathbf{C}(j+1, j+1)]^{1/2}$, where $\mathbf{C}(j+1, j+1)$ is the $(j+1)$-th diagonal entry of $\mathbf{C}$. Conventionally we take $\mathbf{x} = [x_0 \ldots x_{M-1}]^{\top}$ and $\bar{\mathbf{x}} = [\bar{x}_0 \ldots \bar{x}_{M-1}]^{\top}$ such that to re-write (7) in the simpler form:

$$(\mathbf{x} - \bar{\mathbf{x}})^{\top} (\mathbf{x} - \bar{\mathbf{x}}) = d. \tag{8}$$

Remark that (8) is the equation of a sphere centered at $\bar{\mathbf{x}}$. Hence, the largest box within it is the hyper-cube with side length 2ℓ, where $\ell = \sqrt{d/M}$. A simple comparison of the equations (7) and (8) leads immediately to the conclusion that is more convenient to partition the x-space instead of c-space. Another great simplification comes from the fact that the side length of the hyper-cube centered at $\bar{\mathbf{x}}$ and defined as described above is independent of $\bar{\mathbf{x}}$. Thus, one can construct the partition as follows. Firstly choose a point $\mathbf{x}^0 \in \mathbb{R}^M$ and demarcate the hyper-cube centered at $\mathbf{x}^0$. Secondly fix the centers of its neighbors which are hyper-cubes with side length 2ℓ. Then

the construction continues similarly until the complete set of hyper-cubes is settled.

The only problem which remains to be solved is how to select the point $\mathbf{x}^0$. We briefly explain the selection methodology by considering the j-axis of the cartesian system in the x-space. It was already shown in [3] that the TV of the cepstral coefficients can be reduced by turning to zero the estimate $\hat{c}_j$ whenever $c_j^2 \leq \mathbf{C}(j+1, j+1)/N$. Because $c_j^2 = \mathbf{C}(j+1, j+1)/N$ is equivalent to $x_j \in \{-1, 1\}$, we begin the construction by considering the two intervals of length 2ℓ centered at the points -1 and $+1$, respectively. The smallest zero-centered interval which comprises the two intervals centered at -1 and $+1$ is $(-1 - \ell, 1 + \ell)$. Once the ends of this interval have been fixed, we choose its neighbors to be the intervals $(-1 - 3\ell, -1 - \ell]$ and $[1 + \ell, 1 + 3\ell)$. The procedure continues without difficulties until the entire space is partitioned.

Let $\mathcal{Q}(\cdot)$ be the operator which gives for and arbitrary $\mathbf{x} \in \mathbb{R}^M$ the centroid of the hyper-cube to which $\mathbf{x}$ belongs. It is easy to prove that $\mathcal{Q}(\mathbf{x}) = [q(x_0) \ldots q(x_{M-1})]^{\top}$, where

$$q(x_j) = \begin{cases} 0, & |x_j| < 1 + \ell \\ \text{sgn}(x_j) \left(1 + 2\ell + 2\ell \left\lfloor \frac{|x_j| - 1 - \ell}{2\ell} \right\rfloor\right), & \text{otherwise} \end{cases}$$

for all $j \in \{0, \ldots, M-1\}$. The signum function is denoted $\text{sgn}(\cdot)$. We also use the notation $\lfloor u \rfloor$ for the largest integer less than or equal to the real-valued number u.

Now we have to evaluate the code length for the quantized values $\{q(x_j)\}_{j=0}^{M-1}$. We define $\eta = \{j : j \in [0, M-1], q(x_j) \neq 0\}$, which is called structure. Furthermore, we denote k the cardinality of η, and we assume that $0 < k < M$. Let $\eta = \{j_0, \ldots, j_{k-1}\}$,

$$\mathbf{z} = [x_{j_0}, \ldots, x_{j_{k-1}}]^{\top}, \tag{9}$$

and $\tilde{\mathbf{z}} = [q(x_{j_0}), \ldots, q(x_{j_{k-1}})]^{\top}$. In other words, we collect in $\tilde{\mathbf{z}}$ all quantized entries of $\mathbf{x}$ which are non-zero. It is evident that the decoder will have full information on the quantized $\mathbf{x}$ if the encoder transmits the structure η and the entries of $\tilde{\mathbf{z}}$. Next we calculate the code length for η and $\tilde{\mathbf{z}}$, and we also compute the distortion produced by quantizing $\mathbf{x}$ to $\mathcal{Q}(\mathbf{x})$. After these preparations, we give the expression of the KSF, and we show how it can be used to choose automatically the value of μ_{KSF} for thresholding-based cepstral analysis.

Kolmogorov structure function

Code length for η According to [7], the description length for η is given by

$$L_{\eta}(\ell) = \min \left\{ M, \left[\ln \binom{M}{k} + \ln k + \log_2(1 + \ln M) \right] \right\}. \tag{10}$$

We assume the decoder has a priori information on the number of samples used in estimation, or equivalently, the decoder knows the value of M. It is obvious how L_{η} depends on the value of ℓ: if one increases ℓ, then is likely that k decreases. To reduce the computational burden, we apply the Stirling approximation (see equation (6.28) in [7]) together with some simple manipulations, and we get

$$\ln \binom{M}{k} \approx (M + \frac{1}{2}) \ln M - (k + \frac{1}{2}) \ln k$$
$$- (M - k + \frac{1}{2}) \ln(M - k) - \frac{1}{2} \ln(2\pi). \tag{11}$$

Code length for $\tilde{\mathbf{z}}$ Based on Theorem 1.1 and the definition (9), we can write $\mathbf{z} = \zeta + \varepsilon$, where the entries of ζ are unknown, and the random vector ε is sampled from a normal distribution with mean zero

and covariance matrix equal to the identity matrix. Hence, we have the likelihood function $f(\mathbf{z};\zeta) = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}||\mathbf{z}-\zeta||^2\right)$, and the maximum likelihood estimator of ζ is $\hat{\zeta}(\mathbf{z}) = \mathbf{z}$. We also have $\hat{\mathbf{z}} = \mathbf{z}$.

The partition that we have defined for the x-space induces a partition of the z-space, and we denote $B_k(0), B_k(1), \ldots$ its equivalence classes. Each such equivalence class is given by the interior of a hyper-cube. Like in [7], we conventionally assign to the centroid of the j-th equivalence class ($j \geq 0$) the value of the integral defined by $Q_k(j) = \int_{\hat{\zeta}(\mathbf{z}) \in B_k(j)} f(\mathbf{z};\hat{\zeta}(\mathbf{z}))d\mathbf{z}$. This helps us to define for the equivalence classes a discrete prior whose expression is given by $w_k(j) = \dfrac{Q_k(j)}{\sum_{j' \geq 0} Q_k(j')}$. Because there exists a unique $j^* \geq 0$ such that $\mathbf{z} \in B_k(j^*)$, and $\tilde{\mathbf{z}}$ is the centroid of $B_k(j^*)$, the code length for $\tilde{\mathbf{z}}$ is [7]:

$$L_{\tilde{\mathbf{z}}}(\ell) = -\ln w_k(j^*) = -\ln \frac{Q_k(j^*)}{C_k}, \tag{12}$$

where $C_k = \sum_{j' \geq 0} Q_k(j')$. In general, C_k is not finite, and this poses troubles. In Appendix, we show how such difficulties can be circumvented, and we also compute $Q_k(j^*)$. The following approximate formula is obtained:

$$L_{\tilde{\mathbf{z}}}(\ell) \approx \frac{k}{2} \ln\left[\frac{||\tilde{\mathbf{z}}||^2/k}{\ell^2} \frac{\pi \exp(1)}{2}\right] + \frac{1}{2}\ln k. \tag{13}$$

By definition, the magnitude of each entry of $\mathbf{z}$ is not smaller than $1+2\ell$, which guarantees that $L_{\tilde{\mathbf{z}}}(\ell)$ is strictly positive.

<u>Distortion</u> We compute the distortion with the customary formula,

$$D_{\mathbf{x}}(\ell) = \frac{1}{2} \sum_{j=0}^{M-1} \left[x_j - q(x_j)\right]^2. \tag{14}$$

Furthermore, we use (10)-(11) and (13)-(14) to evaluate the KSF [7]:

$$h_{\mathbf{c}}(\ell) = L_\eta(\ell) + L_{\tilde{\mathbf{z}}}(\ell) + D_{\mathbf{x}}(\ell). \tag{15}$$

Next, KSF is employed as an yardstick to select the optimum ℓ^* from a pre-defined set of possible values of ℓ. More precisely, $\ell^* = \arg\min_\ell h_{\mathbf{c}}(\ell)$. This is equivalent to selecting for the parameter d in (7)-(8) the optimal value $d^* = \left(M^{1/2}\ell^*\right)^2$. Consequently, the threshold to be applied in TV reduction is:

$$\mu_{\text{KSF}} = 1 + \ell^*. \tag{16}$$

3. EXPERIMENTAL RESULTS

The figure of merit for evaluating the performance of the thresholding-based scheme defined in (4) is the ratio $\rho = \text{TV}(\hat{\mathbf{c}})/\text{TV}(\check{\mathbf{c}})$. It is clear that a large value of ρ means a significant reduction of the TV, which will improve the estimate of the log-spectrum as we can see from (1). Like in [3], we calculate $\text{TV}(\hat{\mathbf{c}})$ and $\text{TV}(\check{\mathbf{c}})$ by replacing in (2)-(3) the expectation operator with an average over 1000 Monte Carlo simulations.

For $N \in \{128, 256, 512, 1024, 2048\}$, we generate data according to the same models that have been considered also in [3]: (i) broadband moving average (MA) with a small dynamic range of the log-spectrum $y_t = e_t + 0.55e_{t-1} + 0.15e_{t-2}$; (ii) broadband MA with a medium dynamic range of the log-spectrum $y_t = e_t + 0.4574e_{t-1} + 0.2157e_{t-2} + 0.3951e_{t-3} + 0.1383e_{t-4}$; (iii) narrowband autoregressive moving average (ARMA) with a large dynamic range of the

log-spectrum $y_t = 1.55y_{t-1} - 0.95y_{t-2} + e_t + 0.75e_{t-1} + 0.35e_{t-2}$, where $t \in \{0, \ldots, N-1\}$ and e_t is zero-mean white Gaussian noise with variance one. A plot with the log-spectra of the three models can be found in [3] at page 70.

The thresholds μ_{UMPUT}, μ_{BIC} and μ_{KSF} are computed with formulas from (5), (6) and (16), respectively. The value of ℓ^* in (16) is chosen from the set $\{0.0, 0.1, \ldots, 8.9, 9.0\}$ such that to minimize the KSF in (15). For the sake of comparison, we evaluate ρ also for the case when one knows the values of the true cepstral coefficients and selects $\mu \in \{1.0, 1.1, \ldots, 9.9, 10.0\}$ such that to minimize $\text{TV}(\check{\mathbf{c}})$. The outcome of this procedure is named μ_{genie} because we have assumed knowledge of the ground truth. Remark that, for a given model, μ_{KSF} changes from one realization to another, whereas μ_{UMPUT}, μ_{BIC} and μ_{genie} are the same for all realizations.

The results of the experiments are displayed in Fig. 1. For clarity of plots we do not include the values of ρ obtained with the EbayesThresh method [8] because it was already shown in [3] that thresholding with (4) when $\mu = \mu_{\text{UMPUT}}$ is better than EbayesThresh. In Fig. 1(a) , the performance obtained with μ_{BIC} is modest, whereas both μ_{UMPUT} and μ_{KSF} perform as well as μ_{genie}. We can notice from Fig. 1(b) that, in the case of the second model, all methods for threshold selection lead to very similar results. In Fig. 1(c), the ratio ρ obtained with μ_{BIC} almost coincides in all points with that obtained with μ_{genie}. The ρ-curve corresponding to μ_{UMPUT} is slightly inferior to that of μ_{BIC}. The performance of μ_{KSF} is in between those of μ_{UMPUT} and μ_{BIC}. Remark also how the maximum value of ρ varies from one model to another: is about 500 in Fig. 1(a), becomes about 50 in Fig. 1(b), and is as small as 7 in Fig. 1(c).

4. CONCLUSION

In this paper, we focused on a thresholding-based method for TV-reduction. The main contribution was to show how the KSF can be used for selecting the threshold μ. In the framework of cepstral analysis, we compared the newly proposed method with other two schemes that are considered state-of-the-art: the first one chooses μ with a carefully designed UMPUT, while the second one relies on BIC for the selection of μ. It was shown experimentally that it is better to use KSF than BIC. Even if KSF and UMPUT have similar performance, it is more advantageous to apply the KSF-based scheme because is fully automatic, whereas UMPUT requires a priori information on the type of the observed signal. Let us also note that KSF can be applied to any TV-reduction problem for which exists a theoretical result similar with Theorem 1.1.

APPENDIX

To evaluate C_k, we apply a method that was originally proposed in [9], and further used in [7, 10, 11]. We note that $C_k = \int f(\mathbf{z}; \hat{\zeta}(\mathbf{z}))d\mathbf{z}$, where the domain of integration is the whole region of $\mathbb{R}^k$ in which $\mathbf{z}$ can lie. Because the domain is not upper bounded, the integral is not finite. To solve the problem, we restrict the domain of integration to $\mathcal{Z}(R) = \{\mathbf{z} : ||\hat{\mathbf{z}}||^2 \leq kR\}$, where $R > 0$ is a hyper-parameter, and we evaluate $C_k(R) = \int_{\mathcal{Z}(R)} f(\mathbf{z}; \hat{\zeta}(\mathbf{z}))d\mathbf{z} = \int_{\mathcal{Z}(R)} \frac{1}{(2\pi)^{k/2}}d\mathbf{z} = \frac{[(kR)/2]^{k/2}}{\Gamma(k/2+1)}$, where $\Gamma(\cdot)$ is the Euler integral of the second kind. In the calculations above, we applied the formula for the volume of the sphere in the k-dimensional Euclidean space. More importantly, we have from the obtained

identity that $\lim_{R\to\infty}C_k(R) = \infty$, which makes necessary to consider a second step in the computation of C_k, as it was suggested by Rissanen. Remark also that the code length given by (12) increases when R becomes bigger. Hence, we choose R to be the smallest possible value allowed by the definition of $\mathscr{Z}(R)$, which leads to $\hat{R} = ||\hat{\mathbf{z}}||^2/k$ and $C_k(\hat{R}) = \frac{(||\hat{\mathbf{z}}||^2/2)^{k/2}}{\Gamma(k/2+1)}$. Then we define $\mathscr{Z}(R_1, R_2) = \{\mathbf{z} : kR_1 \leq ||\hat{\mathbf{z}}||^2 \leq kR_2\}$, where $R_1 < R_2$ are strictly positive hyper-parameters, and we integrate again:

$$C_k(R_1,R_2)$$

$$= \int_{\mathscr{Z}(R_1,R_2)} \frac{f(\mathbf{z};\hat{\zeta}(\mathbf{z}))}{C_k(\hat{R})}\mathrm{d}\mathbf{z} = \int_{\mathscr{Z}(R_1,R_2)} \frac{1/(2\pi)^{k/2}}{\frac{(||\mathbf{z}||^2/2)^{k/2}}{\Gamma(k/2+1)}}\mathrm{d}\mathbf{z} \quad (17)$$

$$= \frac{\Gamma(k/2+1)}{\pi^{k/2}}\int_{kR_1}^{kR_2}\frac{1}{r^{k/2}}\frac{\pi^{k/2}}{\Gamma(k/2+1)}\frac{k}{2}r^{k/2-1}\mathrm{d}r \quad (18)$$

$$= \frac{k}{2}\int_{kR_1}^{kR_2}\frac{1}{r}\mathrm{d}r = \frac{k}{2}\ln\frac{R_2}{R_1}. \quad (19)$$

All the calculations above are straightforward except (17), where we take $||\mathbf{z}||^2 = r$ such that the volume is $V = \frac{\pi^{k/2}}{\Gamma(k/2+1)}r^{k/2}$ and the volume element is $\mathrm{d}V = \frac{\pi^{k/2}}{\Gamma(k/2+1)}\frac{k}{2}r^{k/2-1}\mathrm{d}r$. Then we get (18) without difficulties. By collecting the previous results, we have $C_k = C_k(R)C_k(R_1,R_2) = \frac{(||\hat{\mathbf{z}}||^2/2)^{k/2}}{\Gamma(k/2)}\ln\frac{R_2}{R_1}$, which implies

$$\ln C_k = \frac{k}{2}\ln||\hat{\mathbf{z}}||^2 - \frac{k}{2}\ln 2 - \ln\Gamma\left(\frac{k}{2}\right) + \ln\ln\frac{R_2}{R_1} \quad (20)$$

$$\approx \frac{k}{2}\ln\frac{||\hat{\mathbf{z}}||^2}{k} + \frac{1}{2}\ln k + \frac{k}{2} + ct, \quad (21)$$

where $ct = \ln\ln(R_2/R_1) - 1/2\ln(4\pi)$. Note that (21) was derived from (20) by applying the Stirling approximation for $\ln\Gamma(k/2)$. Additionally, R_1 and R_2 are nuisance parameters, but for our problem it might be natural to take $R_1 = (1+\ell)^2$. Unfortunately this choice leads to numerical difficulties when one wants to compute $\ln\ln(1/R_1)$ because $R_1 > 1$. On contrary, $\ln\ln(R_2/R_1)$ is well defined because we assume $0 < R_1 < R_2$. Moreover, for ℓ small, the value of $||\mathbf{z}||^2/k$ is about the same for all $\mathbf{z} \in B_k(j^*)$, which allows us to replace $||\hat{\mathbf{z}}||^2/k$ by $||\tilde{\mathbf{z}}||^2/k$ in (21). So,

$$\ln C_k \approx \frac{k}{2}\ln\frac{||\tilde{\mathbf{z}}||^2}{k} + \frac{1}{2}\ln k + \frac{k}{2}. \quad (22)$$

Because in the k-dimensional space $B_k(j^*)$ is a hyper-cube with side length 2ℓ, we readily obtain: $Q_k(j^*) = \int_{\hat{\zeta}(\mathbf{z})\in B_k(j^*)}f(\mathbf{z};\hat{\zeta}(\mathbf{z}))\mathrm{d}\mathbf{z} = \frac{1}{(2\pi)^{k/2}}\int_{\mathbf{z}\in B_k(j^*)}\mathrm{d}\mathbf{z} = \left(\frac{2\ell^2}{\pi}\right)^{k/2}$. This identity together with (12) and (22) lead to (13).

5. REFERENCES

[1] P. Stoica and R. Moses, *Spectral analysis of signals*, Prentice Hall, 2005.

[2] P. Stoica and N. Sandgren, "Smoothed nonparametric spectral estimation via cepstrum thresholding - introduction of a method for smoothed nonparametric spectral estimation," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 34–45, 2006.

[3] P. Stoica and N. Sandgren, "Total-variance reduction via thresholding: Application to cepstral analysis," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 66–72, 2007.

[4] E. L. Lehmann, *Testing statistical hypotheses*, John Wiley & Sons, 1970.

[5] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
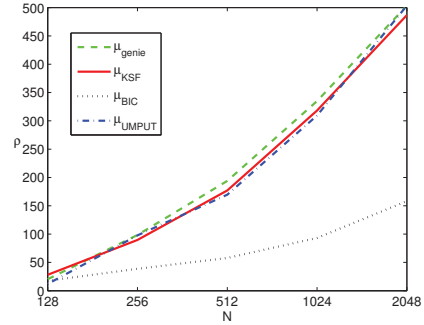
(a) Broadband MA signal with a medium dynamic range of the log-spectrum.
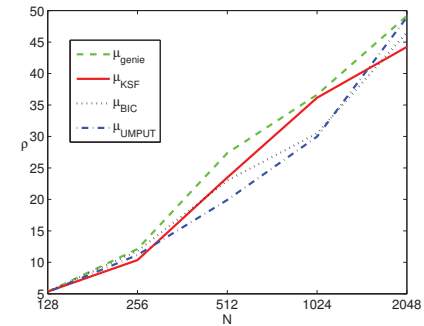
(b) Broadband MA signal with a medium dynamic range of the log-spectrum.
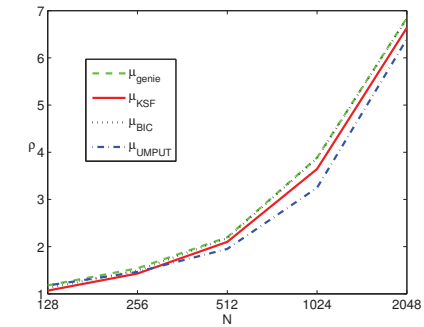
(c) Narrowband ARMA signal with a large dynamic range of the log-spectrum.

Fig. 1. The ratio $\rho = \mathrm{TV}(\hat{\mathbf{c}})/\mathrm{TV}(\check{\mathbf{c}})$ versus the sample size N for various selections of the threshold μ.

[6] M. Li and P.M.B. Vitanyi, *An introduction to Kolmogorov Complexity and its applications*, Springer Verlag, 1997.

[7] J. Rissanen, *Information and Complexity in Statistical Modeling*, Springer Verlag, 2007.

[8] A. Antoniadis, M. Jansen, I. M. Johnstone, and B. W. Silverman, "EbayesThresh: Matlab software for empirical Bayes thresholding," *Available: http://www-lmc.imag.fr/lmc-sms/Anestis.Antoniadis/EBayesThresh/*, 2004.

[9] J. Rissanen, "MDL denoising," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2537–2543, 2000.

[10] E.P. Liski, "Normalized ML and the MDL principle for variable selection in linear regression," in *Festschrift for Tarmo Pukkila on his 60th birthday*, E.P. Liski, J. Isotalo, J. Niemelä, S. Puntanen, and G.P.H. Styan, Eds., pp. 159–172. Univ. of Tampere, 2006.

[11] T. Roos, P. Myllymaki, and J. Rissanen, "MDL denoising revisited," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3347–3360, 2009.