ON TOTAL-VARIANCE REDUCTION VIA THRESHOLDING-BASED SPECTRAL ANALYSIS

*Petre Stoica, Niclas Sandgren**

Systems and Control Division
Department of Information Technology
Uppsala University
P.O. Box 337
SE-751 05 Uppsala Sweden

ABSTRACT

Consider a vector of independent normal random variables with unknown means but known variances. Our problem is to reduce the total variance of these random variables by exploiting the prior information that a significant proportion of them have "small" means. We show that thresholding is an effective means of solving this problem, and propose two schemes for threshold selection: one based on a uniformly most powerful unbiased test, the other on a Bayesian information criterion selection rule. As an example application we consider cepstral analysis and we show via numerical simulation that the simple thresholding scheme proposed herein can achieve significant reductions of total variance.

1. INTRODUCTION

Let $\{\hat{c}_k\}_{k=0}^{M}$ be independent random variables having normal distributions with *unknown* means $\{c_k\}$ and *known* variances $\{s_k^2\}$,

$$\hat{c}_k \sim \mathcal{N}(c_k, s_k^2) \qquad k = 0, \dots, M. \tag{1}$$

The total variance of $\{\hat{c}_k\}$,

$$\text{TV}(\hat{\mathbf{c}}) \triangleq \sum_{k=0}^{M} E(\hat{c}_k - c_k)^2 = \sum_{k=0}^{M} s_k^2 \tag{2}$$

$$\hat{\mathbf{c}} = [\hat{c}_0 \ \cdots \ \hat{c}_M]^T$$

is often an important performance measure in applications. Consequently, the problem of reducing the TV of $\{\hat{c}_k\}$ by exploiting any available information on $\{c_k\}$ is of significant interest. Quite frequently, the only information we have about $\{c_k\}$ is that many of them take on "small" values. For example, this is the case in cepstral analysis (see, e.g., [1]). Our problem, therefore, is to use the a priori information that a large proportion of $\{c_k\}$ are nearly zero, to replace $\{\hat{c}_k\}$ by new estimates $\{\tilde{c}_k\}$ with reduced TV: TV($\tilde{\mathbf{c}}$) $<$ TV($\hat{\mathbf{c}}$) (preferably, TV($\tilde{\mathbf{c}}$) $\ll$ TV($\hat{\mathbf{c}}$)). Note also that typically $M \gg 1$, which is another fact we have to keep in mind when tackling the stated problem.

Several papers in the statistical literature have proposed new estimates $\tilde{c}_k$ of c_k, which have a somewhat smaller variance than the variance s_k^2 of $\hat{c}_k$ when c_k is "small"; see, e.g., [2] and the references therein. However, these estimates have a larger variance than s_k^2 whenever c_k is not "small", which may well nullify the *relatively small* reduction in variance for "small" means. As a consequence, the use of these estimates does not necessarily lead to a satisfactory reduction in the TV.

In this paper we propose a thresholding-based approach for TV reduction. In contrast with the estimation-based approach in [2] and its references, our approach is based on detection. To introduce the main idea behind our approach, we note that the trivial estimate $\tilde{c}_k = 0$ has a variance equal to c_k^2, and therefore it is preferable to $\hat{c}_k$ whenever

$$c_k^2 \leq s_k^2. \tag{3}$$

This observation suggests the following detection-based solution to the TV reduction problem. Let

$$S = \left\{k \in [0, M] \mid c_k^2 \leq s_k^2\right\} \tag{4}$$

and let $\tilde{S}$ denote an estimate of the set S. We use $\tilde{S}$ to obtain new estimates $\tilde{c}_k$ of c_k via the thresholding of $\{\hat{c}_k\}_{k\in\tilde{S}}$:

$$\tilde{c}_k = \begin{cases} 0 & \text{if } k \in \tilde{S} \\ \hat{c}_k & \text{else} \end{cases} \qquad (k = 0, \dots, M). \tag{5}$$

To obtain $\tilde{S}$ we use both a uniformly most powerful unbiased test (UMPUT) and a Bayesian information criterion (BIC) selection rule - see Sections 2 and 3. The so-obtained sets $\tilde{S}$ turn out to be accurate estimates of S in many cases, which leads to a guaranteed TV reduction: TV($\tilde{\mathbf{c}}$) $<$ TV($\hat{\mathbf{c}}$); the larger is S, the more significant is the TV reduction achieved by our approach. In Section 4 we present two numerical examples based on cepstral analysis (see, e.g., [1] and the references therein) to illustrate the TV reductions that can be achieved via the proposed methodology, along with some concluding remarks.

2. UMPUT-BASED ESTIMATION OF S

Because the random variables $\{\hat{c}_k\}$ are independent, the test in the definition of S, see (3), can be performed separately for each $k = 0, \dots, M$. Let the null hypothesis be that (3) holds true,

$$\mathcal{H}_0 : |c_k| \leq s_k \tag{6}$$

and let the alternative hypothesis be that (3) is false,

$$\mathcal{H}_1 : |c_k| > s_k. \tag{7}$$

Also, let the false alarm probability P_{FA} be given (P_{FA} is the probability of inferring that $\mathcal{H}_1$ is true when in fact $\mathcal{H}_0$ is true). Then an

This work was partly supported by the Swedish Science Council (VR).
* Corresponding author. E-mail: niclas.sandgren@it.uu.se.

UMPUT of (6) against (7) is given by ([3]):

$$|\hat{c}_k| \leq \mu s_k \tag{8}$$

where μ is implicitly defined via the equality

$$\text{prob}(|\hat{c}_k| > \mu s_k \text{ under } |c_k| = s_k) = P_{FA}. \tag{9}$$

Below we present some explicit values of μ each corresponding to a given P_{FA} (which will be used later on in the paper):

$$\begin{array}{lcl} \mu = 2 & \leftrightarrow & P_{FA} \approx 0.160 \\ \mu = 3 & \leftrightarrow & P_{FA} \approx 0.023 \\ \mu = 4 & \leftrightarrow & P_{FA} \approx 0.001 \\ \mu = 5 & \leftrightarrow & P_{FA} \approx 0.000. \end{array} \tag{10}$$

As already stated, for a given P_{FA}, the test (8) with μ obtained from P_{FA} via (9) is UMPU. This is clearly an appealing property. However, the performance of the test depends on P_{FA} and therefore the choice of P_{FA} should be considered with some care. Indeed, while simple data-independent choices of P_{FA} - such as $P_{FA} \in [0.001\,,\,0.005]$ - may often work reasonably well, larger TV reductions can be achieved in a given problem by using empirical experience and a modest amount of a priori information on $\{c_k\}$ to choose P_{FA}.

For the cepstral analysis problem considered in Section 4 we introduce some practical thresholding values of μ - see Table 1. Corresponding values of P_{FA}, for some of the values of μ in Table 1, follow from (10). Note that in Table 1 we assumed that $N \in [128\,,\,2048]$: for values of $N < 128$, the distribution of the estimates $\{\hat{c}_k\}$ may differ significantly from the asymptotic distribution in (1), assumed throughout this paper, whereas values of $N > 2048$ are not common in practical applications.

Once μ was selected, the UMPUT-based estimate of S is given by:

$$\tilde{S} = \{k \in [0, M] \mid |\hat{c}_k| \leq \mu s_k\} \tag{11}$$

and the corresponding enhanced estimates $\{\tilde{c}_k\}$ of $\{c_k\}$ are obtained from (5).

The a priori information required to select μ as in Table 1 is quite modest. The corresponding thresholding scheme for TV reduction can be considered to be *essentially automatic*. On the other hand, one might think that a fully automatic scheme could be obtained in the following manner. Let $\tilde{S}_\mu$ denote the estimate in (11) of the set S, corresponding to a specific μ. The $\text{TV}(\tilde{\mathbf{c}}_\mu)$ associated with $\tilde{S}_\mu$ is given by:

$$\text{TV}(\tilde{\mathbf{c}}_\mu) = \sum_{k \in \tilde{S}_\mu} c_k^2 + \sum_{k \notin \tilde{S}_\mu} s_k^2. \tag{12}$$

Ideally, we would like to obtain the value of μ that minimizes the above $\text{TV}(\tilde{\mathbf{c}}_\mu)$. However we cannot determine this minimizing value of μ exactly because the first term in (12) is unknown. To estimate this "optimal" value of μ, first we need to estimate the first term in (12). However, the available estimates of the said term (see, e.g., [4] [5]) are not sufficiently accurate to produce a satisfactory estimate of the μ that minimizes (12). Consequently, such a fully data-dependent scheme for selecting μ will lead typically to larger $\text{TV}(\tilde{\mathbf{c}})$ values than for example the simple, weakly data-dependent scheme in Table 1 (see [1] for more details on this aspect). A satisfactory N-dependent scheme for threshold selection can, though, be derived via a BIC approach - as explained in the section that follows.

3. BIC-BASED ESTIMATION OF S

Let S be defined as in (4), and let $f_S(\hat{\mathbf{c}}\,;\mathbf{c})$ denote the likelihood function of $\{\hat{c}_k\}_{k=0}^M$ that corresponds to a generic set S (which may possibly be empty). The BIC estimate of S is obtained as follows (see, e.g., [6] [7] and references in the latter paper):

$$\min_S \ \text{BIC}(S) \tag{13}$$

where

$$\text{BIC}(S) = \min_{\mathbf{c}} \left[-2 \ln f_S(\hat{\mathbf{c}}\,;\mathbf{c})\right] + n_S \ln(M+1). \tag{14}$$

In (14), n_S is the number of "free" parameters in the vector $\mathbf{c}$ (this number will depend on S, see below). Note that BIC is an asymptotic selection rule that holds only if $n_S \ll M$ - this condition will be verified later on, when an expression for n_S will be available.

It follows from distributional properties of $\{\hat{c}_k\}$ (see (1) and the related discussion) that, to within an additive constant,

$$-2 \ln f_S(\hat{\mathbf{c}}\,;\mathbf{c}) = \sum_{k \in S} \frac{(\hat{c}_k - c_k)^2}{s_k^2} + \sum_{k \in \bar{S}} \frac{(\hat{c}_k - c_k)^2}{s_k^2} \tag{15}$$

where $\bar{S}$ is the difference set $[0, M] - S$. Let $\hat{S}$ be defined similarly to S, but for $\{\hat{c}_k\}$ in lieu of $\{c_k\}$:

$$\hat{S} = \left\{k \in [0, M] \mid \hat{c}_k^2 \leq s_k^2\right\}. \tag{16}$$

To simplify the remaining derivation, we take a short cut based on the observation that $\hat{S}$ can be expected to belong to the estimate $\tilde{S}$ of S (which is yet to be derived). Therefore, we can assume that S contains $\hat{S}$: $\hat{S} \subset S$. Under this natural assumption we obtain from (15):

$$\min_{\mathbf{c}} \left[-2 \ln f_S(\hat{\mathbf{c}}\,;\mathbf{c})\right] = \sum_{k \in (S - \hat{S})} \frac{(|\hat{c}_k| - s_k)^2}{s_k^2} \tag{17}$$

where the minimum value is obtained at

$$\begin{cases} c_k = \hat{c}_k & k \in \hat{S} \text{ and } k \in \bar{S} \\ c_k = s_k \,\text{sign}(\hat{c}_k) & k \in (S - \hat{S}). \end{cases} \tag{18}$$

Note that the number of "free" parameters in the previous minimization problem is

$$n_S = M + 1 - |S - \hat{S}| \tag{19}$$

where $|S - \hat{S}|$ denotes the number of elements of the difference set $(S - \hat{S})$ - indeed, in the minimization of the negative log-likelihood function in (17), $|S - \hat{S}|$ parameters $\{c_k\}$ have been constrained to be $\pm s_k$. Also note that in the application example considered in this paper, viz. cepstral analysis, and presumably in other applications as well, $|S - \hat{S}|$ is quite close to M; therefore the condition $n_S \ll M$ required for the validity of the BIC rule is satisfied.

It follows from (13), (14), (17), and (19) that the BIC estimate $\tilde{S}$ of S is obtained via the minimization, with respect to S, of the function:

$$\begin{aligned} &\sum_{k \in (S - \hat{S})} \frac{(|\hat{c}_k| - s_k)^2}{s_k^2} - |S - \hat{S}| \ln(M+1) \\ &= \sum_{k \in (S - \hat{S})} \frac{(|\hat{c}_k| - s_k)^2 - s_k^2 \ln(M+1)}{s_k^2}. \end{aligned} \tag{20}$$

Table 1. Values of μ recommended for thresholding-based cepstral analysis.

N \ Signal type	Broadband with small dynamic range	Broadband with medium dynamic range	Narrowband with large dynamic range
$N = 128$	4	3	2
$N \in (128, 2048)$	$4+\frac{N-128}{1920}$	$3+\frac{N-128}{1920}$	$2+\frac{N-128}{1920}$
$N = 2048$	5	4	3

Interestingly, the minimizing set $\tilde{S}$ can be obtained explicitly:

$$\tilde{S} = \left\{ k \in [0, M] \mid (|\hat{c}_k| - s_k)^2 \leq s_k^2 \ln(M+1) \right\}. \quad (21)$$

To see this, observe that for $S = \tilde{S}$ all terms in (20) are negative, and that if $S \neq \tilde{S}$ then (20) increases (compared with its value at $S = \tilde{S}$) either because it gets positive terms, or it loses negative terms, or both.

To re-write the definition (21) of $\tilde{S}_{BIC}$ in a form similar to that of $\tilde{S}_{UMPUT}$ in (11), we note the following facts:

(i) Clearly $\hat{S} \subset \tilde{S}$ (for $M \geq 2$);
(ii) For $k \notin \hat{S}$, the inequality in (21) can be re-written as

$$|\hat{c}_k| \leq \mu_N s_k \quad ; \quad \mu_N = 1 + [\ln(M+1)]^{1/2} \quad (22)$$

(iii) The above inequality, (22), holds trivially for $k \in \hat{S}$.

It follows from these observations that the BIC estimate of S in (21) can be re-written in a form perfectly analogous to that of the UMPUT estimate in (11):

$$\tilde{S} = \{ k \in [0, M] \mid |\hat{c}_k| \leq \mu_N s_k \} \quad (23)$$

where μ_N is as defined in (22).

The difference between the two set estimates in (11) and, respectively, (23) consists in the values selected for μ. By simple numerical computations it can be readily seen that the BIC choice of μ agrees reasonably well with the values of μ recommended in Table 1 for the UMPUT in the case of signals whose spectra have a medium dynamic range.

4. NUMERICAL EXAMPLES AND CONCLUDING REMARKS

To illustrate the TV reductions that can be achieved by using the proposed thresholding-based scheme, we consider the application of this scheme to the cepstral analysis problem (see [1]). Let $\{y(t)\}_{t=0}^{N-1}$ be an observed sample of a stationary, discrete-time, real-valued signal with spectrum $\Phi(\omega)$. For notational convenience, we let $\{\Phi_p\}$ denote the values taken by the spectrum at the Fourier frequency grid points:

$$\omega_p = \frac{2\pi}{N} p \quad ; \quad p = 0, \ldots, N-1. \quad (24)$$

The periodogram estimate of Φ_p is given by (see, e.g., [1]):

$$\hat{\Phi}_p = \frac{1}{N} \left| \sum_{t=0}^{N-1} y(t) e^{-i\omega_p t} \right|^2 \quad ; \quad p = 0, \ldots, N-1. \quad (25)$$

Assuming that $\Phi_p, \hat{\Phi}_p > 0$ ($p = 0, \ldots, N-1$) and that N is even (for simplicity), let

$$c_k = \frac{1}{N} \sum_{p=0}^{N-1} \ln(\Phi_p) e^{i\omega_k p} \quad ; \quad k = 0, \ldots, M \quad (26)$$

$$\hat{c}_k = \frac{1}{N} \sum_{p=0}^{N-1} \ln(\hat{\Phi}_p) e^{i\omega_k p} + \gamma \delta_{k,0} \quad ; \quad k = 0, \ldots, M \quad (27)$$

$$\delta_{k,0} = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{else} \end{cases}$$

where $M = N/2$ and where $\gamma = 0.577216...$ is the so-called Euler's constant. Consequently, in this section the random variables $\{\hat{c}_k\}$ are the estimates given by (27), and their means $\{c_k\}$ are given by (26). Two cases of (log)spectra are considered: with small and, respectively, medium range, see below for details. In each case we will show the ratio TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$) for $N = 128, 256, 512, 1024$, and 2048. The enhanced estimates $\{\tilde{c}_k\}$ are obtained via (5) where $\tilde{S}$ is given either by (11) and Table 1 [for UMPUT] or by (23) and (22) [for BIC]. We remind the reader that in the case of cepstral analysis, TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$) is a measure of the performance of the corresponding estimated log-spectra (see [1]).

In the computation of TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$) the expectation operation in the definition of TV (see, e.g, (2)) is approximated by an average over 1000 Monte-Carlo simulations. We also use Monte-Carlo simulations to estimate TV($\tilde{\mathbf{c}}$) for $\tilde{S}$ given by (11) with $\mu \in [0, 10]$; then we pick the value of μ that minimizes the so-estimated TV($\tilde{\mathbf{c}}$), which we call μ_{genie} because its determination requires knowledge of the true sequence $\{c_k\}$. In the figures that follow we also show the *ultimate* ratio TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$) for $\mu = \mu_{genie}$.

The signals used in our simulation study have been generated as follows:
(i) Broadband MA with a small dynamic range of the log-spectrum

$$y(t) = e(t) + 0.55\, e(t-1) + 0.15\, e(t-2)$$
$$t = 0, \ldots, N-1$$

(ii) Broadband MA with a medium dynamic range of the log-spectrum

$$y(t) = e(t) + 0.4574\, e(t-1) + 0.2157\, e(t-2) + 0.3951\, e(t-3) + 0.1383\, e(t-4) \qquad t = 0, \ldots, N-1$$

where $e(t)$ is a normal white noise with mean zero and unit variance.

The true log-spectra of these signals are shown in Fig. 1 to lend support to the assertions made above about their dynamic spectral range. The corresponding ratios TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$) for $\mu = \mu_{UMPUT}$, μ_{BIC}, and μ_{genie} are displayed in Figs 2 - 3. The following remarks on the simulation results shown in these figures are in order:

- The TV reduction obtained using $\mu = \mu_{UMPUT}$ in case (i) is nearly optimal (see Fig. 2, where the curve corresponding to μ_{UMPUT} is very close to the curve corresponding to μ_{genie}

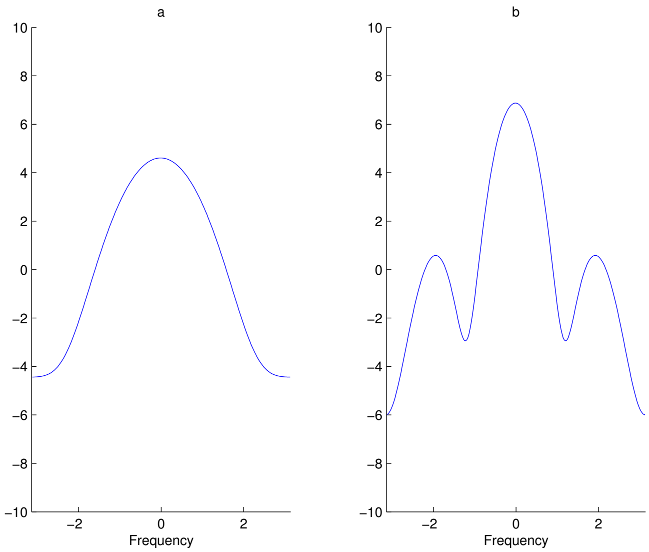

Fig. 1. The true log-spectra (in dB) of the two signals: (a) broadband MA signal with small dynamic range, and (b) broadband MA signal with medium dynamic range.

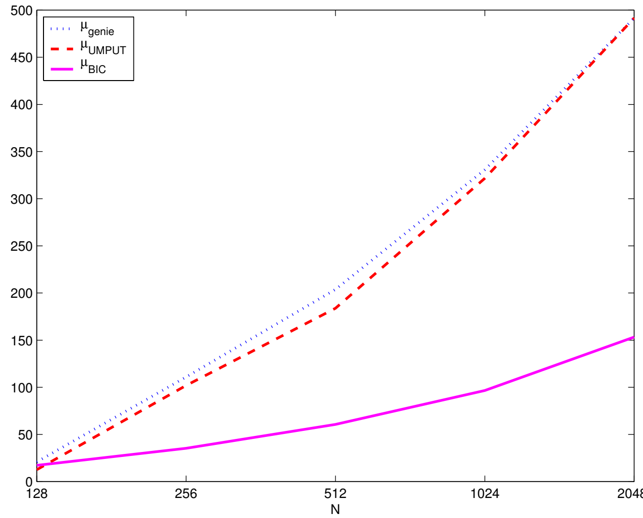

Fig. 2. The ratio TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$), versus N, in the case of the broadband MA signal with a small dynamic range of the log-spectrum.

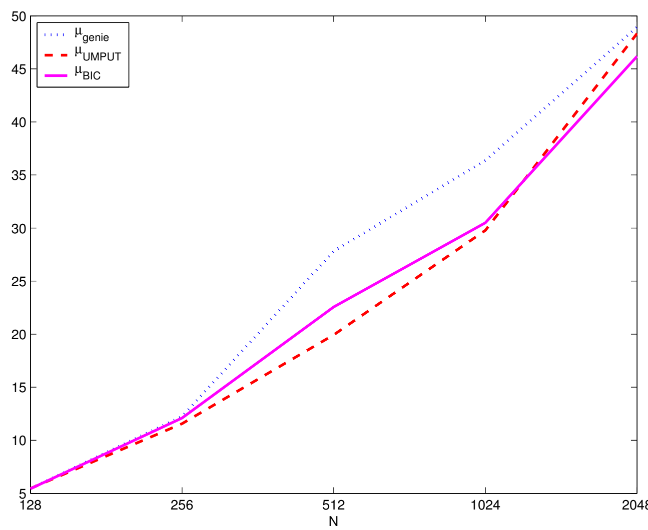

Fig. 3. The ratio TV($\hat{\mathbf{c}}$)/TV($\tilde{\mathbf{c}}$), versus N, in the case of the broadband MA signal with a medium dynamic range of the log-spectrum.

for all values of N), and it ranges from 13 at $N = 128$ to 492 at $N = 2048$.

- From Fig. 2 we also note that the TV reduction obtained by using $\mu = \mu_{BIC}$ is less impressive (it varies from 17 at $N = 128$ to 153 at $N = 2048$) than that obtained using $\mu = \mu_{UMPUT}$.
- Finally, the TV reductions obtained in case (ii) using $\mu = \mu_{UMPUT}$ and $\mu = \mu_{BIC}$ are very similar to one another and quite close to the TV reduction obtained for $\mu = \mu_{genie}$ (see Fig. 3). The obtained TV reduction varies from about 5 at $N = 128$ to about 47 at $N = 2048$ for both $\mu = \mu_{UMPUT}$ and $\mu = \mu_{BIC}$ as well as $\mu = \mu_{genie}$.

5. REFERENCES

[1] P. Stoica and N. Sandgren. Smoothed Nonparametric Spectral Estimation via Cepstrum Thresholding. *IEEE Sign. Proc. Mag.*, Submitted.

[2] J.R. Magnus. Estimation of the Mean of a Univariate Normal Distribution with Known Variance. *The Econometrics Journal*, vol. 5, pp.225-236, 2002.

[3] E.L. Lehmann. *Testing Statistical Hypotheses.* Chap. 4.2. Wiley, NY, 1986.

[4] K.M. Lal Saxena and K. Alam. Estimation of the Non-Centrality Parameter of a Chi Squared Distribution. *The Annals of Statist.*, vol. 10, pp.1012-1016, 1982.

[5] I.M. Johnstone. Chi-Square Oracle Inequalities. In *State of the Art in Prob. and Statist.*, IMS Lecture Notes - Monographs, vol. 36, pp.399-418, 2001.

[6] G. Schwarz. Estimating the Dimension of a Model. *Ann. Statist.*, vol. 6, pp.461-464, 1978.

[7] P. Stoica and Y. Selén. Model-Order Selection: a Review of Information Criterion Rules. *IEEE Sign. Proc. Mag.*, vol. 21, pp.36-47, 2004.