

A suboptimal estimator of the sampling jitter variance using the bispectrum

Ilan Sharfer^a, Hagit Messer^{b,*}

^a*The University of Michigan, Electrical Engineering and Computer Science, 3314 EECS building, North Campus, Ann Arbor, MI 48109-2122, USA*

^b*Department of Electrical Engineering – Systems, Tel-Aviv University, Tel-Aviv 69978, Israel*

Received 12 March 1992; revised 8 January 1993 and 12 November 1993

Abstract

We consider the problem of estimating parameters of an irregular sampling process defined as a uniform sampling process in which the deviations from the nominal sampling times constitute a random IID process (jitter). Emphasis is placed on estimating the variance of the jitter, based on observation of samples taken from a continuous band-limited third-order stationary process. We derive an estimation procedure which uses the bispectrum estimates of a process with a priori known bispectrum. Derivation of the generalized likelihood ratio in the bispectral domain, leads to a statistic with which a bispectrum-based maximum likelihood estimation can be done. We propose a suboptimal estimator, and show that it is asymptotically unbiased and consistent. The dependence of the estimator's performance on the data length and the skewness is studied for a specific example. The estimator's variance is compared to the bispectrum-based Cramer–Rao bound (BCRB), and is shown to approach it for sufficiently large data length or skewness. Computer simulations verify the effectiveness of the proposed estimation method for small jitter.

Zusammenfassung

Wir betrachten die Schätzung von Parametern eines unregelmäßigen Abtastprozesses. Dieser ist definiert als regelmäßiger Abtastprozeß mit statistisch unabhängigen und identisch verteilten zufälligen Abweichungen von den nominellen Abtastzeitpunkten (Jitter). Im Vordergrund steht die Schätzung der Varianz des Jitters basierend auf beobachteten Abtastwerten eines stetigen, bandbegrenzten und bis zur dritten Ordnung stationären Prozesses. Wir leiten eine Schätzmethode ab, die die Bispektrum-Schätzwerte eines Prozesses mit bekanntem Bispektrum verwendet. Die Ableitung des verallgemeinerten Likelihood-Verhältnisses im Bispektrum-Bereich führt zu einer Statistik, die eine auf dem Bispektrum beruhende Maximum-Likelihood-Schätzung ermöglicht. Wir schlagen einen suboptimalen Schätzer vor und zeigen dessen asymptotische Erwartungstreue und Konsistenz. Die Abhängigkeit der Leistungsfähigkeit des Schätzers vom Umfang der Daten und von der Asymmetrie wird anhand eines konkreten Beispiels untersucht. Die Varianz des Schätzers wird mit der auf dem Bispektrum beruhenden Cramer–Rao-Schranke (BCRB) verglichen, und es wird gezeigt, daß sich die Varianz der BCRB bei hinreichend großem Datenumfang oder hinreichend großer Asymmetrie annähert. Computersimulationen bestätigen die Effektivität der vorgeschlagenen Schätzmethode im Fall kleinen Jitters.

*Corresponding author. E-mail: messer@eng.tau.ac.il.

Résumé

Nous considérons dans cet article le problème de l'estimation des paramètres d'un processus à échantillonnage irrégulier défini comme processus à échantillonnage uniforme pour lequel les déviations vis-à-vis des instants d'échantillonnage uniforme constituent un processus aléatoire indépendant à distribution constante (gigue). L'emphase est mise sur l'estimation de la variance de la gigue, sur la base d'observation d'échantillons obtenus à partir d'un processus continu stationnaire d'ordre trois à bande limité. Nous dérivons une procédure d'estimation qui utilise les estimées du bispectre d'un processus dont le bispectre est connu a priori. La dérivation du rapport de vraisemblance généralisé dans le domaine bispectral conduit à une statistique avec laquelle une estimation aux maximum de vraisemblance basée sur le bispectre peut être effectuée. Nous proposons un estimateur sous-optimal, et montrons qu'il est asymptotiquement non biaisé et consistant. La dépendance de la performance de l'estimateur vis-à-vis de la longueur des données et de l'asymétrie de la distribution est étudiée sur un exemple spécifique. La variance de l'estimateur est comparée à la borne de Cramer–Rao sur le bispectre, et il est montré qu'elle s'en approche pour une longueur de données ou une asymétrie de distribution suffisamment grande. Des simulations sur ordinateur permettent de vérifier l'efficacité de la méthode d'estimation proposée dans le cas d'une gigue faible.

Key words: Sampling noise; Jitter; Bispectrum; Higher order spectra

1. Introduction

An increasing number of signal processing applications are done on discrete data. In most cases, the discrete data are obtained by a sampling operation which converts a continuous physical signal to a discrete signal. This operation may be subject to a systematic or random timing error (jitter), which causes the sampling instants to become shifted in an unpredictable manner, sometimes resulting in a significant undesirable effect on the performance of the system. While a considerable amount of research was done on the effects of other disturbances, such as noise and quantization, relatively little was done regarding the problem of jitter in sampling, basically because in most cases sampling clocks are sufficiently accurate, so that the jitter has a negligible effect. However, as sampling rate increases, the jitter may become a factor that should be taken into account.

In previous works, jitter has been discussed mainly in the framework of irregular sampling. Spectral estimation of an irregularly sampled process has been studied in [7], based on observation both of the sampled data and the sampling times. Other works [1,3] treated the data as a discrete sequence and determined the relation between the discrete-time and continuous-time signal spectra in terms of the statistical properties of the jitter, i.e. its characteristic function. This approach seems to be more practical than the former, because it avoids the difficulty of obtaining an accurate observation of the sampling times. In [8,9] this approach is extended further to the bispectrum, which is defined as the two-dimensional Fourier transform of the third-order cumulant sequence. The main result in those papers is that jitter in sampling can be detected without prior assumptions about the statistics of the sampled continuous-time process, beyond that it is a non-Gaussian, third-order stationary stochastic process, and that sampling rate is greater or equal to the Nyquist rate. A test statistic based on bispectrum estimates is developed in [9] and its performance is evaluated. Simulation results of the performance of this test are reported in [11].

A related problem to jitter detection is the problem of estimation of jitter parameters. An efficient estimation procedure can be applied, for example, in ECM systems for classification of jammers and identification of friendly sources. Unlike jitter detection, which under the conditions stated above, can be done only in the bispectral domain, provided that the sampling rate is no smaller than Nyquist rate, estimation of jitter parameters can be done also the spectral domain. The question as to which domain is preferable for carrying out the estimation is considered in [10], where it is shown using Cramer–Rao bounds that the answer depends on the skewness of the continuous process. If the skewness is larger than one, then it

may be possible to achieve better estimation performance by using the bispectrum. However, no specific estimation procedure is proposed there.

In this work we present an algorithm for estimation of the jitter variance using the bispectrum. This procedure can be adapted to the estimation of other jitter parameters, as long as they affect the characteristic function of the jitter. The advantage of estimating jitter variance is that the jitter detection and estimation problems can be solved simultaneously. Specifically, the sampling is assumed to be uniform with additive independent identically distributed (IID) jitter. That is, the normalized sampling times have the form

$$t_n = n + \tau_n, \quad n = 0, \pm 1, \pm 2, \dots, \quad (1)$$

where $\{\tau_n\}$, the normalized jitter, is an IID zero mean random sequence with unknown probability density function (PDF) $f(\tau)$. In this work we will be interested in estimating the parameter θ , which is defined as the variance of the jitter.

$$\theta = E\{\tau_n^2\}. \quad (2)$$

The effect of the jitter, that was defined above, on the bispectrum has been studied previously in [9]. In Appendix A we derive the expression given below of the discrete bispectrum $B_d(\omega_1, \omega_2)$ of the sampled data:

$$B_d(\omega_1, \omega_2) = B_s(\omega_1, \omega_2) \phi(\omega_1) \phi(\omega_2) \phi^*(\omega_1 + \omega_2) + B_1(\omega_1) + B_1(\omega_2) + B_1^*(\omega_1 + \omega_2) + B_0, \quad (3)$$

where (ω_1, ω_2) is a bifrequency in the principal domain of the bispectrum, and $B_s(\omega_1, \omega_2)$ is the discrete bispectrum in the case of uniform, jitter-free sampling. The jitter by-product terms $B_1(\omega)$ and B_0 are given by

$$B_1(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} [|\phi(\omega)|^2 - \phi(\omega) \phi(u) \phi^*(\omega + u)] B_s(\omega, u) du \quad (4)$$

and

$$B_0 = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [1 - |\phi(\omega_1)|^2 - |\phi(\omega_2)|^2 - |\phi(\omega_1 + \omega_2)|^2 + 2\phi(\omega_1) \phi(\omega_2) \phi^*(\omega_1 + \omega_2)] \\ \times B_s(\omega_1, \omega_2) d\omega_1 d\omega_2. \quad (5)$$

The characteristic function of the jitter is the Fourier transform of $f(\tau)$:

$$\phi(\omega) = E\{\exp(j\omega\tau_n)\}. \quad (6)$$

In this work we use the terminology introduced in [6], which separates the *principal domain* of the bispectrum (PD) to two subsets called the *inner triangle* (IT), and the *outer triangle* (OT):

$$\text{IT} \equiv \{(\omega_1, \omega_2) | \omega_1 + \omega_2 \leq \pi, \omega_1 \geq \omega_2, \omega_2 \geq 0\}, \\ \text{OT} \equiv \{(\omega_1, \omega_2) | \omega_1 + \omega_2 \geq \pi, \omega_1 + \omega_2/2 \leq \pi, \omega_1 \geq \omega_2\}. \quad (7)$$

Our underlying assumption is that the sampled process is third-order stationary, strictly band-limited, whose bandwidth is smaller than half the sampling rate. It was shown in [6] that in the case of uniform sampling the bispectrum in the OT is identically zero.

The parameter estimation method is based on using a continuous process with *a priori known* bispectrum as the input to the sampler. This requirement can be justified in several applications. One example is testing of a sampling system – an analog signal with known spectrum/bispectrum is sampled, in order to measure the variance of the jitter introduced by the sampling system. Another example is classification of source

stations. Assuming that the random jitter characteristic of a transmitter is unique, one can identify the signature of a source by estimating the jitter parameters. Notice that estimation of jitter variance can be done also when the bispectrum is unknown, as will be shown later. However, to get a significant estimate one needs extremely large amount of data, which may be unrealistic in practice. The bispectrum of the output sequence is estimated by a standard bispectrum estimation procedure, in which a sample sequence of length N is divided into K records, each of length M , forming a bispectrum estimate for each record, averaging it over a square of L^2 bifrequencies, and averaging each estimate over the records. Refer to Appendix B for details of this estimation algorithm. The bispectrum estimates obtained are asymptotically unbiased and consistent, and in addition have been shown to be asymptotically distributed as a complex Gaussian vector with a positive definite diagonal covariance matrix [4,13]. This property has been used in the formulation of a generalized likelihood ratio test (GLRT) in the bispectral domain, that has been applied in [9] for the problem of detecting the existence of jitter in a proper subset of the bispectral domain. In the present work we extend the likelihood ratio to the whole principal domain, and we propose to use it for estimating θ . We note that any estimation carried out in the bispectral domain is essentially *suboptimal*, because it makes use of a transformation of the sampled data, unless the bispectrum is a sufficient statistic of the data. Therefore, a bispectrum-based maximum likelihood (BML) estimate of θ is suboptimal compared to an ML estimate which uses the data directly. However, an obvious advantage of bispectrum-based estimation is that it is much simpler and robust, because it does not require knowledge of the joint probability distribution of the jitter or the sampled process.

The jitter model considered in this work assumes an IID jitter. In many practical cases this assumption is not valid, e.g. digital magnetic recording where jitter is data dependent, or cases where jitter is correlated. Therefore, the model developed in this work cannot be applied directly to the given problem. However, once the correct relation between jitter variance, or any other parameter and the bispectrum has been established, it is still possible to use the bispectrum for estimation of that parameter. So, the same principles can be used for the more general cases of jitter.

The paper is organized as follows. In Section 2 we develop the likelihood ratio in the principal domain and introduce the suboptimal estimator of θ . In Section 3 we examine the performance of this estimator through a specific example, and discuss the trade-offs between sample size and skewness needed to achieve high performance. In Section 4 we present computer simulation results, and in Section 5 the conclusions.

2. A suboptimal estimator to the jitter variance in the bispectrum

In [9] we have analyzed the performance of jitter detection algorithm done in the outer triangle of the bispectrum. Clearly, no a priori knowledge of the signal statistics is needed, because the bispectrum of a uniformly sampled third-order stationary strictly band-limited signal is identically zero in the outer triangle. The only assumption required is that the sampling rate is sufficiently high, so that aliasing is avoided. In this section we will see that by assuming a priori knowledge of the signal bispectrum, it is possible to use the bispectrum estimates in both the inner and outer triangles for estimating parameters of the jitter. Specifically, we propose a suboptimal estimator of the jitter variance θ , by making use of the dependence of the jitter characteristic function on this parameter. We will analyze the properties of this estimator, its performance and limitations.

The maximum likelihood estimate of the jitter variance, based on the bispectrum estimates (BML) results from maximizing the PDF of the bispectrum estimates (the likelihood function) with respect to the unknown parameter. Instead of maximizing this likelihood function, we choose to maximize the bispectral likelihood ratio (BLR), where the bispectral likelihood is normalized by the PDF of the bispectrum estimates when no jitter is present. Since the latter is a positive quantity, independent of the unknown parameter, maximization of the BLR with respect to the unknown jitter variance results in its BML estimate. We proceed by

formulating the following binary hypothesis problem:

$$H_0: \underline{\xi} \sim N_c(\mu_0, \Sigma_0), \quad H_1: \underline{\xi} \sim N_c(\mu_1, \Sigma_1), \quad (8)$$

where $\underline{\xi}$ is the vector of the bispectrum estimates in the PD, which is asymptotically distributed as a complex Gaussian variate. The expectation of $\underline{\xi}$ under each hypothesis, is just the bispectrum of the sampled signal, depending on the existence of the jitter in the sampling operation. Therefore,

$$\mu_0 = \begin{cases} B_s(\omega_i, \omega_j), & (\omega_i, \omega_j) \in \text{IT}, \\ 0, & (\omega_i, \omega_j) \in \text{OT}, \end{cases} \quad (9)$$

$$\mu_1 = \begin{cases} B_s(\omega_i, \omega_j)\phi(\omega_i; \theta)\phi(\omega_j; \theta)\phi^*(\omega_i + \omega_j; \theta) \\ + B_1(\omega_i; \theta) + B_1(\omega_j; \theta) + B_1^*(\omega_i + \omega_j; \theta) + B_0(\theta), & (\omega_i, \omega_j) \in \text{IT}, \\ B_1(\omega_i; \theta) + B_1(\omega_j; \theta) + B_1(2\pi - \omega_i - \omega_j; \theta) + B_0(\theta), & (\omega_i, \omega_j) \in \text{OT}, \end{cases}$$

where $B_s(\omega_i, \omega_j)$ is the bispectrum in the case of a uniformly sampled signal with no jitter ($\theta = 0$). In (9) we have explicitly denoted the dependence of the bispectrum under the alternative hypothesis on θ . The matrices Σ_0 and Σ_1 in (8) are unknown positive definite covariance matrices under each hypothesis. Define

$$\Sigma = \text{diag} \left\{ \frac{M}{KL^2} S(\omega_i)S(\omega_j)S(\omega_i + \omega_j) \right\}, \quad (\omega_i, \omega_j) \in \text{PD}. \quad (10)$$

Then, $\Sigma_0 = \Sigma$, and $\Sigma_1 = \Sigma$ with $S(\omega) = S(\omega; \theta)$ being the jitter-dependent spectrum of the sampled sequence. However, we neglect the dependency of the spectrum on θ (the justification for this will be discussed later) and assume that $\Sigma_0 = \Sigma_1 = \Sigma$.

Since the expectation under the null hypothesis, μ_0 , is known, it can be subtracted from $\underline{\xi}$, leading to the following equivalent decision problem:

$$H_0: \underline{\xi} - \mu_0 \sim N_c(0, \Sigma), \quad H_1: \underline{\xi} - \mu_0 \sim N_c(\mu_1 - \mu_0, \Sigma). \quad (11)$$

The sufficient statistic for the BLR of (11) is known to be the T_c^2 statistic (see [2]), which simplifies due to the asymptotic diagonality of Σ to

$$l = \sum_{(\omega_i, \omega_j) \in \text{PD}} |\hat{\beta}(\omega_i, \omega_j)|^2, \quad (12)$$

where β is the bicoherence function estimate:

$$\hat{\beta}(\omega_i, \omega_j) = \begin{cases} \frac{\hat{B}(\omega_i, \omega_j) - B_s(\omega_i, \omega_j)}{\left[\frac{M}{KL^2} S(\omega_i)S(\omega_j)S(\omega_i + \omega_j) \right]^{1/2}}, & (\omega_i, \omega_j) \in \text{IT}, \\ \frac{\hat{B}(\omega_i, \omega_j)}{\left[\frac{M}{KL^2} S(\omega_i)S(\omega_j)S(\omega_i + \omega_j) \right]^{1/2}}, & (\omega_i, \omega_j) \in \text{OT}. \end{cases} \quad (13)$$

Clearly, subtraction of a constant from the bicoherence estimate $\beta(\omega_i, \omega_j)$ does not affect its variance. Therefore, the test statistic (12) has the following asymptotic distribution:

$$l \sim \chi_{2P}^2(\lambda), \quad \lambda = \sum_{(\omega_i, \omega_j) \in \text{PD}} \lambda(\omega_i, \omega_j), \quad (14)$$

where

$$\lambda(\omega_i, \omega_j) = \begin{cases} \frac{2|B_d(\omega_i, \omega_j) - B_s(\omega_i, \omega_j)|^2}{\frac{M}{KL^2}S(\omega_i)S(\omega_j)S(\omega_i + \omega_j)}, & (\omega_i, \omega_j) \in \text{IT}, \\ \frac{2|B_d(\omega_i, \omega_j)|^2}{\frac{M}{KL^2}S(\omega_i)S(\omega_j)S(\omega_i + \omega_j)}, & (\omega_i, \omega_j) \in \text{OT}, \end{cases} \quad (15)$$

$\chi^2_{2P}(\lambda)$ is the noncentral chi-square distribution with $2P$ degrees of freedom and noncentrality parameter λ , where P is the number of bispectrum estimates. $B_d(\omega_i, \omega_j; \theta)$ is the bispectrum of the sampled signal, which depends on the jitter variance θ . Since we limit our attention to cases where the jitter is small and zero-mean, we can approximate its characteristic function by

$$\begin{aligned} \phi(\omega_i; \theta) &\cong 1 - \frac{\theta\omega_i^2}{2}, \\ \phi(\omega_i; \theta)\phi(\omega_j; \theta)\phi^*(\omega_i + \omega_j; \theta) &\cong 1 - \frac{\theta(\omega_i^2 + \omega_j^2 + (\omega_i + \omega_j)^2)}{2}. \end{aligned} \quad (16)$$

The constant term $B_0(\theta)$ in (9) is usually negligible for small values of θ , because it is proportional to θ^3 (see [8]). The other term $B_1(\omega; \theta)$ has been shown in [9] to be well approximated by

$$B_1(\omega; \theta) \cong \theta\Gamma(\omega), \quad \Gamma(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi-\omega} u(\omega + u)B_s(\omega, u) du. \quad (17)$$

By utilizing the expressions above in (15), we obtain

$$\begin{aligned} \lambda(\omega_i, \omega_j) &= \\ &\begin{cases} \frac{2KL^2\theta^2 \left| -B_s(\omega_i, \omega_j) \frac{\omega_i^2 + \omega_j^2 + (\omega_i + \omega_j)^2}{2} + \Gamma(\omega_i) + \Gamma(\omega_j) + \Gamma(-\omega_i - \omega_j) \right|^2}{MS(\omega_i)S(\omega_j)S(\omega_i + \omega_j)}, & (\omega_i, \omega_j) \in \text{IT}, \\ \frac{2KL^2\theta^2 |\Gamma(\omega_i) + \Gamma(\omega_j) + \Gamma(2\pi - \omega_i - \omega_j)|^2}{MS(\omega_i)S(\omega_j)S(\omega_i + \omega_j)}, & (\omega_i, \omega_j) \in \text{OT}. \end{cases} \end{aligned} \quad (18)$$

We note that in the expression for λ we have neglected the dependence of the spectrum, which appears in the denominator, on θ . This is equivalent to assuming that the covariance matrix Σ does not depend on θ . It will be shown later, by comparison with the BCRB derived in [10], that the effect on the estimator's performance (i.e. its variance) is quite small, particularly when the sampled signal has high skewness.

By combining (18) and (14), we get the following relation between λ and θ :

$$\begin{aligned} \lambda &= \theta^2 \lambda_0, \\ \lambda_0 &= \frac{2KL^2}{M} \left[\sum_{(\omega_i, \omega_j) \in \text{IT}} \frac{\left| -B_s(\omega_i, \omega_j) \frac{(\omega_i^2 + \omega_j^2 + (\omega_i + \omega_j)^2)}{2} + \Gamma(\omega_i) + \Gamma(\omega_j) + \Gamma(-\omega_i - \omega_j) \right|^2}{S(\omega_i)S(\omega_j)S(\omega_i + \omega_j)} \right. \\ &\quad \left. + \sum_{(\omega_i, \omega_j) \in \text{OT}} \frac{|\Gamma(\omega_i) + \Gamma(\omega_j) + \Gamma(2\pi - \omega_i - \omega_j)|^2}{S(\omega_i)S(\omega_j)S(\omega_i + \omega_j)} \right]. \end{aligned} \quad (19)$$

The number of degrees of freedom of the chi-square distribution in (14) is $2P$, where P is the number of points in the PD:

$$P = \frac{M^2}{12L^2}, \tag{20}$$

The exact expression for the chi-square distribution of the statistic (14) is (see [2])

$$p(l; \lambda) = \frac{1}{2^P \sqrt{\pi}} e^{-(\lambda+l)/2} l^{P-1} \sum_{i=0}^{\infty} \frac{\lambda^i l^i}{(2i)!} \frac{\Gamma(1/2 + i)}{\Gamma(P + i)}. \tag{21}$$

In order to maximize the BLR, we have to find the parameter λ which maximizes the likelihood of the statistic (14). Using the relation between λ and θ , given in (19), we get that the ML estimator of θ^2 is approximated by

$$\hat{\theta}_{\text{BML}}^2 = \frac{\hat{\lambda}_{\text{BML}}}{\lambda_0}, \tag{22}$$

where $\hat{\lambda}_{\text{BML}}$ is the ML estimate of λ from the statistic (14), and λ_0 is a constant that is known a priori. Since it is quite difficult to find an analytic solution for λ_{BML} , we propose a *sub-optimal* estimator $\hat{\lambda}$, based on the first moment of the non central chi-square distribution:

$$E\{\chi_{2P}^2(\lambda)\} = \lambda + 2P. \tag{23}$$

Therefore, the suboptimal estimator is defined as

$$\hat{\lambda} = \begin{cases} l - 2P, & l > 2P, \\ 0, & l \leq 2P. \end{cases} \tag{24}$$

We proceed with the analysis of the properties of $\hat{\lambda}$, specifically its bias and its variance. In order to simplify the analysis, we make use of a standard approximation to the noncentral chi-square distribution by a Gaussian distribution with mean $2P + \lambda$ and variance $4(P + \lambda)$. Therefore, the expectation of the estimator $\hat{\lambda}$ is given by

$$\begin{aligned} E\{\hat{\lambda}\} &= \int_{2P}^{\infty} (l - 2P) P(l; \lambda) dl \cong \frac{1}{\sqrt{8\pi(P + \lambda)}} \int_{2P}^{\infty} (l - 2P) e^{-(l - 2P - \lambda)^2/8(P + \lambda)} dl \\ &= \frac{2\sqrt{P + \lambda}}{\sqrt{2\pi}} e^{-\lambda^2/8(P + \lambda)} + \lambda \left[1 - \Phi\left(-\frac{\lambda}{2\sqrt{P + \lambda}}\right) \right], \end{aligned} \tag{25}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du. \tag{26}$$

Denoting the estimator's bias by $b(\lambda)$, we get

$$b(\hat{\lambda}) = E\{\hat{\lambda}\} - \lambda \cong \lambda \left[\Phi(-\rho) - \frac{1}{\sqrt{2\pi\rho}} e^{-\rho^2/2} \right], \quad \rho = \frac{\lambda}{2\sqrt{P + \lambda}}. \tag{27}$$

For $\rho \gg 1$ we have that (see [13])

$$\Phi(-\rho) \sim \frac{1}{\sqrt{2\pi\rho}} e^{-\rho^2/2}. \tag{28}$$

From [9], $\lambda = O(N)$ and $\rho = O(\sqrt{N})$, therefore, λ is asymptotically unbiased, because $\sqrt{N}e^{-N}$ goes to 0 as N gets large. In order to calculate the variance of the estimation error, we assume that N is sufficiently large so that the estimator is nearly unbiased, therefore

$$\text{Var}\{\varepsilon_\lambda\} \cong \text{Var}\{\hat{\lambda}\} = E\{\hat{\lambda}^2\} - E^2\{\hat{\lambda}\}, \quad (29)$$

where ε_λ is defined as the estimation error:

$$\varepsilon_\lambda = \hat{\lambda} - \lambda. \quad (30)$$

We calculate $E\{\hat{\lambda}^2\}$ similarly,

$$\begin{aligned} E\{\hat{\lambda}^2\} &= \int_{2P}^{\infty} (l - 2P)^2 P(l; \lambda) dl \cong \frac{1}{\sqrt{8\pi(P + \lambda)}} \int_{2P}^{\infty} (l - 2P)^2 e^{-(l - 2P - \lambda)^2/8(P + \lambda)} dl \\ &= \frac{2\lambda\sqrt{P + \lambda}}{\sqrt{2\pi}} e^{-\lambda^2/8(P + \lambda)} + 4(P + \lambda)[1 - \Phi(-\rho)] + \lambda^2 \left[\frac{1}{\sqrt{2\pi\rho}} e^{-\rho^2/2} - \Phi(-\rho) \right]. \end{aligned} \quad (31)$$

Therefore,

$$\text{Var}\{\varepsilon_\lambda\} \cong 4(P + \lambda)[1 - \Phi(-\rho)] + \lambda^2 \left[\frac{1}{\sqrt{2\pi\rho}} e^{-\rho^2/2} - \Phi(-\rho) \right]. \quad (32)$$

Again, the second term in (32) tends for large N to 0, so

$$\text{Var}\{\varepsilon_\lambda\} \cong 4(P + \lambda). \quad (33)$$

Clearly, the error variance in (33) increases as N gets large, because λ itself is $O(N)$, but the normalized variance, i.e. the standard deviation divided by the mean, tends to 0:

$$\frac{\text{std}\{\varepsilon_\lambda\}}{\lambda} \cong \frac{2\sqrt{(P + \lambda)}}{\lambda} \rightarrow 0, \quad N \rightarrow \infty. \quad (34)$$

In order to illustrate how λ is related to the parameter θ and to the bispectrum, we consider the example which has been discussed previously [9, 10], namely, a special case of a signal with flat (constant) spectrum S and flat bispectrum B . We can replace the sum in (19) with an integral, whose integration domain includes both the inner and the outer triangles. In the inner triangle we obtain

$$\begin{aligned} \lambda_{IT} &= \sum_{(\omega_i, \omega_j) \in IT} \lambda(\omega_i, \omega_j) \\ &= \frac{2KL^2}{MS^3} \sum_{(\omega_i, \omega_j) \in IT} \left| -\frac{B\theta(\omega_i^2 + \omega_j^2 + (\omega_i + \omega_j)^2)}{2} + B_1(\omega_i) + B_1(\omega_j) + B_1(-\omega_i - \omega_j) \right|^2 \\ &= \frac{2KL^2|B|^2\theta^2}{MS^3} \sum_{(\omega_i, \omega_j) \in IT} \left| -\frac{\omega_i^2 + \omega_j^2 + (\omega_i + \omega_j)^2}{2} + \pi^2 + \frac{\omega_i^3 + \omega_j^3 + (\omega_i + \omega_j)^3}{12\pi} - (\omega_i + \omega_j)\pi \right|^2 \\ &\cong \frac{2N|B|^2\theta^2}{4\pi^2S^3} \iint_{IT} \left[-\frac{\omega_1^2 + \omega_2^2 + (\omega_1 + \omega_2)^2}{2} + \pi^2 + \frac{\omega_1^3 + \omega_2^3 + (\omega_1 + \omega_2)^3}{12\pi} - (\omega_1 + \omega_2)\pi \right]^2 d\omega_1 d\omega_2. \end{aligned}$$

The integration domain in (35) is the IT, defined in (7). The result of the integration is

$$\lambda_{IT} \cong 2.358 N \gamma \theta^2, \quad (36)$$

where $\gamma = |B|^2/S^3$ is the average skewness of the signal. Similarly, the integration in the outer triangle yields

$$\lambda_{OT} = \sum_{(\omega_i, \omega_j) \in OT} \lambda(\omega_i, \omega_j) \cong 0.042 N \gamma \theta^2. \quad (37)$$

We realize that λ_{IT} is much larger than λ_{OT} . This means that the information we get from the OT is very small compared to the information we get from the IT. By combining (36) and (37), we get the total λ as

$$\lambda = \lambda_{IT} + \lambda_{OT} \cong \lambda_0 \theta^2, \quad \lambda_0 = 2.4 N \gamma. \quad (38)$$

From the fact that $\lambda_{IT} \gg \lambda_{OT}$ we conclude that the bispectrum estimates in the inner triangle are much more significant for the estimation problem than those in the outer triangle. Additionally, the inner triangle has three times as many points compared to the outer triangle, which also contributes to the relative significance of the IT. Therefore, we may be justified in carrying out the parameter estimation algorithm ignoring the bispectrum estimates in the OT, without any significant loss of information. However, use of the IT bispectrum estimates for this estimation problem is conditioned on an a priori knowledge of the spectrum and bispectrum of the continuous signal that is sampled. In cases where we do not have such a priori knowledge of the bispectrum, we may still be able to make use of the OT bispectrum estimates for the purpose of jitter detection as proposed in [9], or estimation of its variance by using a recursive algorithm.

3. Estimator performance evaluation

In this section we will examine the performance of the estimator θ^2 , using the specific example introduced above, of a signal with flat spectrum and bispectrum. This example which is analytically simple, will also lead us to a relation between the performance of the estimator and the skewness of the signal. When considering the bias of the estimator, $b(\theta^2)$, it is appropriate to normalize it with respect to θ^2 , in order to have a relative measure of the estimator's bias. From the relation between λ and θ^2 , and using Eq. (27) for $b(\lambda)$, we get

$$\frac{b(\hat{\theta}^2)}{\theta^2} = - \left[\Phi(-\rho) - \frac{1}{\sqrt{2\pi\rho}} e^{-\rho^2/2} \right], \quad (39)$$

ρ depends on λ and P , both are functions of N . The number of bispectrum estimates in the principal domain, P , depends on the choice of M and L (see (20) and Appendix B). It is maximized for the choice $M = N$ and $L = \sqrt{N}$,¹ and given by

$$P = \frac{N}{12}, \quad (40)$$

Using (39) and (40) it is now possible to explore the dependency of the normalized bias on N and on the average skewness γ . Fig. 1 shows the normalized bias as a function of θ , for $N = 10^4$ and three values of γ : 1, 10, and 100. The normalized bias can be decreased by increasing ρ which implies increasing N or γ . Also for larger values of θ , the normalized bias gets smaller. However, it should be pointed out that as θ becomes larger, the approximate expression relating λ and θ is less accurate. Therefore as θ gets large (typically for $\theta > 0.1$) the bias is determined by the model accuracy, rather than the bias implied by the estimator. The effect of the sample size on the normalized bias is illustrated in Fig. 2, for three values of N : 10^4 , 3×10^4 , and 10^5 , with constant skewness, $\gamma = 10$. In both cases it is clear that for small values of θ , the normalized bias is

¹Theoretically, for consistency of the bispectrum estimates one needs $L > \sqrt{N}$. Practically, however, $L = \sqrt{N}$ is acceptable.

Table 1
Threshold versus sample size and skewness

$N@ \gamma = 10$	$\theta_{th}(\text{dB})$	$\gamma@ N = 10^4$	$\theta_{th}(\text{dB})$
10^4	- 18.1	1	- 13.1
3×10^4	- 19.3	10	- 18.1
10^5	- 20.6	100	- 23.1

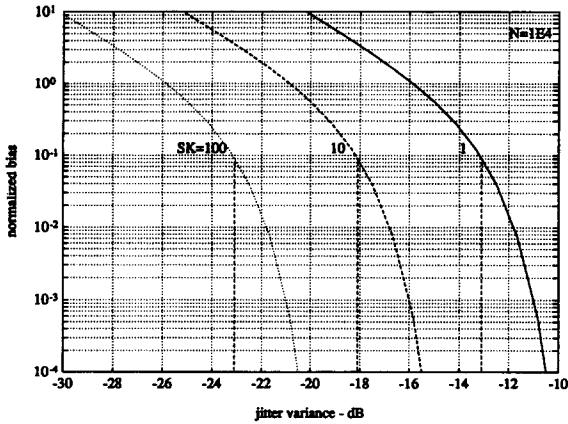


Fig. 1. Normalized estimation bias variation with skewness.

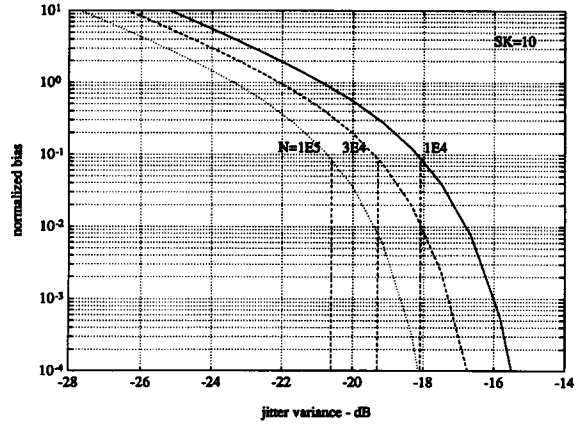


Fig. 2. Normalized estimation bias variation with sample size.

large, but it decreases rapidly as θ exceeds some threshold. We defined the threshold as the value of θ for which $\rho = 1$:

$$\rho = \frac{\lambda}{2\sqrt{P + \lambda}} \Big|_{\theta_{th}} = 1. \tag{41}$$

Substitution of (38) and (40) in (41) yields

$$\frac{2.4N\gamma\theta_{th}^2}{2\sqrt{N/12 + 2.4N\gamma\theta_{th}^2}} = 1. \tag{42}$$

Solving for θ_{th}^2 we get

$$\theta_{th}^2 = \frac{9.6\gamma + \sqrt{\gamma^2 2.4^2 (16 + 4N/3)}}{2 \cdot 2.4^2 N \gamma^2}, \tag{43}$$

which can be simplified for $N \gg 1$:

$$\theta_{th}^2 \cong \frac{1}{2.4\sqrt{3}\sqrt{N\gamma}} \tag{44}$$

or

$$\theta_{th} \cong \frac{1}{\sqrt{2.4^4 \sqrt{3}}} N^{-1/4} \gamma^{-1/2}. \tag{45}$$

Table 1 gives the values of θ_{th} according to (45), corresponding to the values assigned to N and γ in the figures above.

The normalized bias at the threshold is approximately 8.3%. Both figures clearly indicate that as θ exceeds the threshold given in Table 1, the slope of the curves becomes steeper. The table also shows that the performance depends on the fourth root of N and on the square root of γ . That means that increasing N by 10 dB lowers the threshold by 2.5 dB, while increasing γ by 10 dB lowers the threshold by 5 dB. We note that the threshold criterion (41) is similar to the condition for high probability of detection used in [9].

We now consider the variance of the estimator of θ^2 . Using relation (19) between λ and θ , the estimator's variance is

$$\text{Var}\{\hat{\theta}^2\} = \frac{\text{Var}\{\hat{\lambda}\}}{\lambda_0^2}. \tag{46}$$

Since the estimator is asymptotically unbiased, its variance equals the mean square estimation error. In addition it is possible to calculate the bispectral Cramer–Rao bound (BCRB) for θ^2 , by using the asymptotically Gaussian bispectrum estimates as the input data. The bound has already been derived in [10], both for the general case and for the particular case considered here. The bound is given by

$$E\{(\hat{\theta} - \theta)^2\} \geq \frac{1}{J(\theta)}, \quad J(\theta) = J_1(\theta) + J_2(\theta),$$

$$J_1(\theta) = \frac{N}{2\pi^2} \iint_{PD} \left| \frac{\partial}{\partial \theta} B_d(\omega_1, \omega_2; \theta) \right|^2 \frac{1}{S_d(\omega_1; \theta) S_d(\omega_2; \theta) S_d(\omega_1 + \omega_2; \theta)} d\omega_1 d\omega_2, \tag{47}$$

$$J_2(\theta) = \frac{N}{4\pi^2} \iint_{PD} \left[\frac{\partial}{\partial \theta} \log S_d(\omega_1; \theta) + \frac{\partial}{\partial \theta} \log S_d(\omega_2; \theta) + \frac{\partial}{\partial \theta} \log S_d(\omega_1 + \omega_2; \theta) \right]^2 d\omega_1 d\omega_2.$$

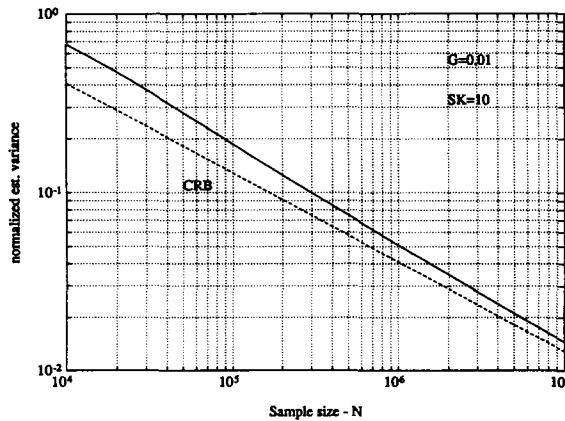


Fig. 3. Normalized estimation variance versus N .

The second term $J_2(\theta)$ reflects the dependence of Σ , the bispectrum covariance matrix, on θ . In the derivation of the suboptimal estimator it was assumed that Σ does not depend on θ . This is equivalent to $J_2(\theta) = 0$, and results in a more relaxed bound. Since $J_2(\theta)$ depends only on the spectrum of the signal, it will have a small effect on the bound, when the signal has high skewness. For example, at $\gamma = 1$ the difference between the bounds is

$$10 \log \frac{J_1(\theta) + J_2(\theta)}{J_1(\theta)} = 10 \log \frac{2.4 + \pi^4/45}{2.4} = 2.79 \text{ dB}, \tag{48}$$

while at $\gamma = 10$ it becomes

$$10 \log \frac{24 + \pi^4/45}{24} = 0.37 \text{ dB}. \tag{49}$$

This example shows clearly that there is no significant loss of performance as a result of ignoring the dependency of the spectrum on θ , even at moderate skewness values. It can be easily verified that $J_1(\theta)$ is equal to λ_0 given by (38) for θ tending to 0. The bound for θ^2 is obtained as

$$\text{Var}\{\hat{\theta}^2\} \geq \frac{4\theta^2}{J(\theta)} = \frac{4\lambda}{\lambda_0^2}. \tag{50}$$

It is easy to verify that the estimator $\hat{\theta}^2$ asymptotically achieves the bound (50), since from (32) and (46) we have

$$\text{Var}\{\hat{\theta}^2\} = \frac{4(P + \lambda)}{\lambda_0^2}. \tag{51}$$

Therefore, to achieve the bound it is necessary that $\lambda \gg P$. Since λ increases linearly with N , it is sufficient to choose the estimation parameters L and M in such a way that P increases as a power of N that is smaller than 1. For choice $M = N$ and $L = N^c$ we get

$$P = \frac{N^{2(1-c)}}{12}, \quad \frac{1}{2} < c < 1. \tag{52}$$

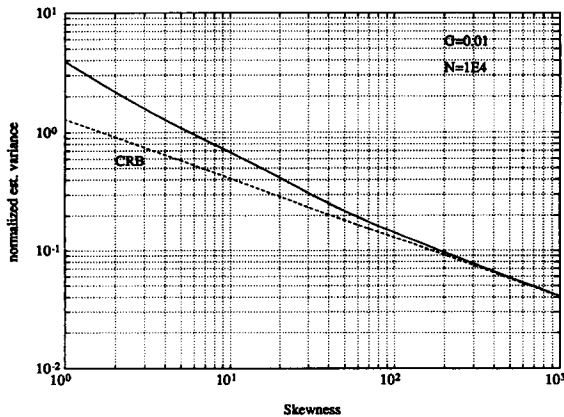


Fig. 4. Normalized estimation variance versus skewness.

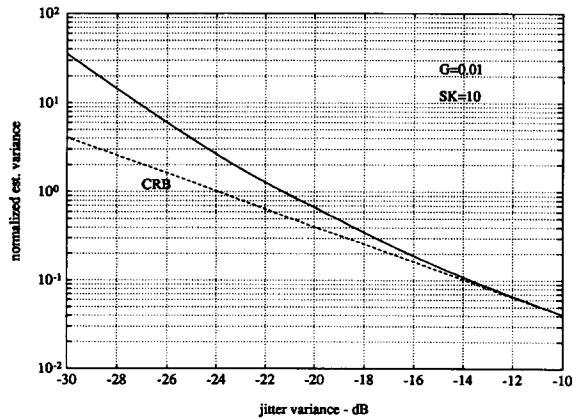


Fig. 5. Normalized estimation variance versus jitter.

The choice of c determines how fast the estimator’s variance converges to the bound. In order to avoid biased bispectrum estimates, it is necessary to choose c close to 1/2, but that comes at the expense of slow convergence to the bound. Additionally, due to the fact that λ is proportional to the average skewness γ , the bound can be approached by increasing γ , independent of N . Fig. 3 illustrates the dependence of the normalized estimator’s variance on the sample size N , and compares it with the normalized BCRB:

$$\frac{\sqrt{\text{Var}\{\hat{\theta}^2\}}}{\theta^2} \geq \sqrt{\frac{4}{\lambda}} \tag{53}$$

The graph is drawn for $\theta = 0.01$ and $\gamma = 10$. It shows that the estimator’s variance indeed approaches the BCRB, although rather slowly, because N is required to be larger than 10^6 in order to be within 1 dB from the bound. The rate of convergence can be increased by a proper choice of c , so that the condition $\lambda \gg P$ is satisfied for smaller N . For example, if the requirement is that the condition $\lambda = 10P$ (0.4 dB from the bound), is satisfied at $N = 10^6$, then c has to be set as follows:

$$c = \frac{1}{2} \left(1 - \frac{\log(1.2 \cdot 2.4\theta^2\gamma)}{\log N} \right) = 0.71. \tag{54}$$

Clearly, the higher the value of c the larger the bias of the bispectrum estimates, because there are fewer estimation points in the principal domain. On the other hand, when c is larger the variance of each bispectrum estimate is smaller. The figures shown were all drawn for $c = 0.65$.

Fig. 4 shows the normalized variance of $\hat{\theta}^2$ as a function of the skewness, and in comparison with the normalized BCRB. The fixed parameters are $\theta = 0.01$ and $N = 10^4$. Obviously, the variance is reduced as γ increases, while approaching further towards the bound. At any skewness value it is possible to approach the bound by increasing N . In Fig. 5 we examine the estimator’s performance for θ in the range between 10^{-3} and 10^{-1} , for $N = 10^4$ and $\gamma = 10$. The normalized BCRB in that range is shown as well. It is seen that as θ increases the normalized variance decreases and gets closer to the bound, because λ increases along with θ while P remains unchanged, which results in a decrease in the variance. This can be verified by (51) as well:

$$\frac{\sqrt{\text{Var}\{\hat{\theta}^2\}}}{\theta^2} = \sqrt{\frac{4}{\lambda}} \sqrt{\frac{P}{\lambda} + 1}. \tag{55}$$

4. Computer simulations

Computer simulations of the proposed jitter estimation algorithm were done to verify that the assumptions made in the analysis of the estimator are reasonable, and to determine the performance of the proposed jitter estimation method.

In each simulation run, a random sequence is generated, which corresponds to samples taken at uniformly spaced time instants from a continuous band-limited third-order stationary process. This process has a known nonzero bispectrum in the IT, and zero bispectrum in the OT. A random IID jitter sequence is generated as well, whose elements are each a realization of a random variable with a given variance and distribution. The actual samples are then reconstructed from the uniformly spaced samples, using the sampling times given by the realization of the jitter process. The reconstruction is carried out by an ideal low-pass filter according to the sampling theorem. The bispectrum of the jittered samples is estimated by a bispectrum estimation program [5], and the resulting bispectrum estimates are used to calculate the statistic (14), and the estimator (24). The bispectrum estimation program uses a slightly different estimation algorithm than the one used here. In this algorithm the N samples are divided to blocks of length L each, bispectrum estimation is done for each block, and then averaged over all the blocks. Since the test statistic does not depend on the bispectrum estimation algorithm, all analytical results remain valid.

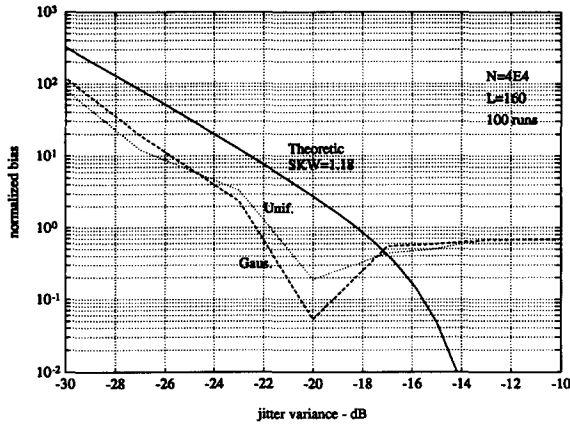


Fig. 6. Normalized estimation bias versus jitter - simulation.

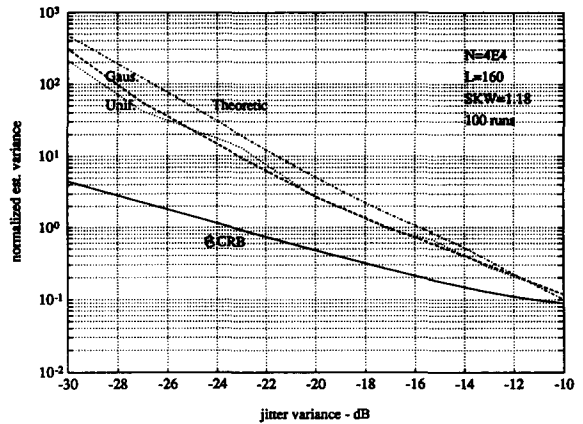


Fig. 7. Normalized estimation variance versus jitter - simulation.

The stochastic sampled process is generated in the frequency domain by random uncorrelated zero-mean spectral components which are multiplicatively coupled. In order to obtain a band-limited signal, no coupling is made between frequencies lying in the OT (sum of frequencies equals to 2π). The random process generated in this way has a nonzero bispectrum in the IT, and zero bispectrum in the OT. The actual signal samples are then generated by means of an inverse DFT, which transforms the signal from its frequency domain representation to the time domain. By using proper scaling of the spectral components, the resulting signal has flat spectrum, but does not have a flat bispectrum. Therefore, we will use its average skewness, in order to compare our simulations with the analytical results. As a consequence of this, the estimator’s bias was slightly different than the analytical results, but the estimator’s variance agreed well generally. Each simulation consists of 100 Monte-Carlo runs. The simulations were carried out for jitter with Gaussian and uniform distributions, and a sampled signal with average skewness $\bar{\gamma} = 1.185$, which was the maximal skewness that could be achieved with the method described above (see [11] for full description of the computer simulations and performance results).

The computer simulation results are presented in the following figures. Fig. 6 shows the normalized estimation bias $b(\theta^2)/\theta^2$, as a function of θ . The bias is determined by the difference between the mean of the estimator and θ^2 . The normalized bias predicted by (55) is shown as well. The number of bispectrum estimates in the PD is $P = L^2/12$, where L is the block length of the simulation. The graph shows that for $\theta < 0.01$ the simulation and analytical results are quite similar, except for a gap which is due to the assumption that bispectrum is flat. As θ increases the normalized bias does not decrease further, because the approximate relation between λ and θ^2 becomes less accurate. For sufficiently small θ , the bias can be reduced further by increasing N or $\bar{\gamma}$.

In Fig. 7, a comparison is made between the normalized variance of the estimator (sample variance) and the analytical results. The normalized bispectral Cramer–Rao bound is shown as a reference. The figure shows a plot of $\text{std}(\theta^2)/\theta^2$ as a function of θ . The simulation points are in close agreement with the analytical results, and both curves approach the BCRB as θ increases. The difference between the simulation and analytical results becomes quite small (less than 1 dB) when both are close to the bound.

Fig. 8 shows the normalized bias of the estimator as a function of the sample size N , for constant jitter variance $\theta = 0.05$. The normalized bias is reduced initially as N increases, but for larger N it remains unchanged because it is limited by the accuracy of the first-order approximation of the characteristic function of the jitter in (16).

Finally, the normalized variance of the estimator as a function of N is presented in Fig. 9, and compared to the analytical results and to the bispectrum-based BCRB. A good agreement between simulation and

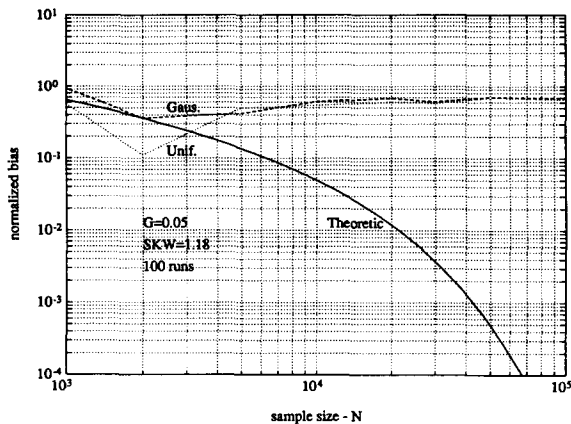


Fig. 8. Normalized estimation bias versus sample size simulation.

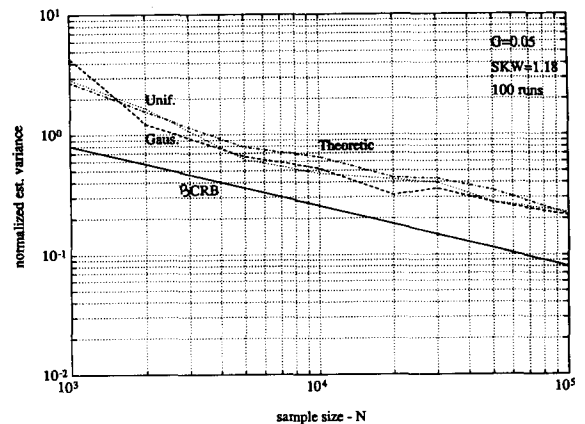


Fig. 9. Normalized estimation variance versus sample size simulation.

analysis is seen, with a difference of less than 1 dB on most of the simulation points. The estimator's variance converges rather slowly to the BCRB, because the simulation block length was chosen as $L \cong \sqrt{N}$. The convergence rate can be increased by using a smaller block length, but this will result in an increased bias of the bispectrum estimates, because there will be less estimation points in the PD.

5. Conclusions

The problem of estimating parameters of a jitter process is important in view of the increasing demand for high performance signal processing applications that use sampled data. The present and previous works [8–10] can be regarded as a feasibility study, in which we have shown that on certain conditions the bispectrum is a domain which can be highly sensitive to the existence of jitter in sampling, and provides a useful signal processing tool for either detection or estimation of this effect. The asymptotic Gaussianity of the bispectrum estimates is a useful property and has simplified the derivation of the likelihood ratio on which the estimation is based. A disadvantage in using the bispectrum in particular and higher-order spectra (HOS) in general is that a relatively large sample size is required in order to satisfy the asymptoticity assumption, and the need for high computation power, although using the FFT algorithm can help considerably. The proposed estimator has been applied on a simulated model of jittered sampling, and was shown to be an effective method for estimating jitter. We have pointed out that in order to improve the performance we can trade-off the sample size with the skewness of the process. Therefore, finding a method which generates a stationary band-limited process with high skewness is a useful research direction. Additionally, it is intuitively clear that the high frequencies of the signal are more affected by the jitter than the low frequencies. This observation suggests that a high-pass process, in which the bispectrum is supported near the outer edge of the inner triangle, has higher potential to detect or estimate the jitter, compared to a low-pass process. This claim, however, needs to be substantiated by further research.

Acknowledgments

We are most grateful to Prof. M.J. Hinich of the university of Texas at Austin for his active support throughout the research, by providing us with his bispectrum estimation routines, and for his useful comments and suggestions.

Appendix A. Derivation of Eq. (3)

In this appendix we outline the derivation leading to Eq. (3) of the bispectrum of the sampled data. We assume a continuous, zero-mean, third-order stationary, band-limited process $\{x(t)\}$. It can be written using the Cramer representation as

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(j\Omega t) dZ(\Omega), \quad (\text{A.1})$$

where

$$\begin{aligned} E\{dZ(\Omega)\} &= 0, & E\{dZ(\Omega_1) dZ(\Omega_2)\} &= 2\pi\delta(\Omega_1 + \Omega_2)S_c(\Omega_1) d\Omega_1, \\ E\{dZ(\Omega_1) dZ(\Omega_2) dZ(\Omega_3)\} &= 2\pi\delta(\Omega_1 + \Omega_2 + \Omega_3)B_c(\Omega_1, \Omega_2) d\Omega_1 d\Omega_2, \end{aligned} \quad (\text{A.2})$$

$S_c(\Omega)$ and $B_c(\Omega_1, \Omega_2)$ are the spectrum and bispectrum of the continuous process $\{x(t)\}$, respectively, and $\delta(\cdot)$ is the Dirac delta function (see [4]). The bispectrum of the discrete (sampled) sequence is defined as the two-dimensional DTFT of the third-order cumulant sequence:

$$C(k, m) = E\{x(t_n)x(t_{n+k})x(t_{n+m})\}. \quad (\text{A.3})$$

The sampling times $\{t_n\}$ are obtained by multiplying the normalized sampling times in (1) by T , the nominal sampling period. Assuming that the jitter process $\{\tau_n\}$ and $\{x(t)\}$ are independent, and using (A.1) and (A.2) we can express (A.3) in terms of the jitter and the bispectrum as

$$\begin{aligned} C(k, m) &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} E\{\exp(-j(\omega_1 + \omega_2)\tau_n + j\omega_1\tau_{n+k} + j\omega_2\tau_{n+m})\} \\ &\quad \cdot \exp(-j(\omega_1 + \omega_2)n + j\omega_1(n+k) + j\omega_2(n+m)) B_s(\omega_1, \omega_2) d\omega_1 d\omega_2, \end{aligned} \quad (\text{A.4})$$

where

$$\omega \equiv \Omega T, \quad B_s(\omega_1, \omega_2) \equiv \frac{1}{T^2} B_c\left(\frac{\omega_1}{T}, \frac{\omega_2}{T}\right). \quad (\text{A.5})$$

Further evaluation of (A.4) requires to distinguish between different regions in the cumulant plane, which correspond to the following cases:

$$C(k, m) = \begin{cases} \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi(\omega_1)\phi(\omega_2)\phi^*(\omega_1 + \omega_2) B_s(\omega_1, \omega_2) \exp(j\omega_1 k + j\omega_2 m) d\omega_1 d\omega_2, & k \neq m \neq 0, \\ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\phi(\omega_2)|^2 B_s(\omega_1, \omega_2) \exp(j\omega_2 m) d\omega_1 d\omega_2, & k = 0, \quad m \neq 0, \\ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\phi(\omega_1)|^2 B_s(\omega_1, \omega_2) \exp(j\omega_1 k) d\omega_1 d\omega_2, & k \neq 0, \quad m = 0, \\ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\phi(\omega_1 + \omega_2)|^2 B_s(\omega_1, \omega_2) \exp(j(\omega_1 + \omega_2)m) d\omega_1 d\omega_2, & k = m \neq 0, \\ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} B_s(\omega_1, \omega_2) d\omega_1 d\omega_2, & k = m = 0, \end{cases}$$

where

$$\phi(\omega) \equiv E\{\exp(j\omega\tau_n)\}. \quad (\text{A.6})$$

The double integration is defined on a hexagonal in the (ω_1, ω_2) plane, such that $|\omega_1| \leq \pi$, $|\omega_2| \leq \pi$, and $|\omega_1 + \omega_2| \leq \pi$, because the sampled process $\{x(t)\}$ is assumed to be stationary and band-limited (see [6]). However, due to jitter in sampling, the bispectrum corresponding to $C(k, m)$ in (A.6), will be nonzero outside the stationary support set. It can be easily verified that the DTFT of $C(k, m)$ inside the stationary support set, denoted by $B_d(\omega_1, \omega_2)$, is given by

$$B_d(\omega_1, \omega_2) = B_2(\omega_1, \omega_2) + B_1(\omega_1) + B_1(\omega_2) + B_1^*(\omega_1 + \omega_2) + B_0, \tag{A.7}$$

where

$$\begin{aligned} B_2(\omega_1, \omega_2) &= B_s(\omega_1, \omega_2)\phi(\omega_1)\phi(\omega_2)\phi^*(\omega_1 + \omega_2), \\ B_1(\omega) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [|\phi(\omega)|^2 - \phi(\omega)\phi(u)\phi^*(\omega + u)] B_s(\omega, u) du, \\ B_0 &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [1 - |\phi(\omega_1)|^2 - |\phi(\omega_2)|^2 - |\phi(\omega_1 + \omega_2)|^2 + 2\phi(\omega_1)\phi(\omega_2)\phi^*(\omega_1 + \omega_2)] \\ &\quad \times B_s(\omega_1, \omega_2) d\omega_1 d\omega_2. \end{aligned} \tag{A.8}$$

In the nonstationary support set, the bispectrum is given by

$$B_d(\omega_1, \omega_2) = B_1(\omega_1) + B_1(\omega_2) + B_1(2\pi - \omega_1 - \omega_2) + B_0. \tag{A.9}$$

The stationary and nonstationary support sets of the bispectrum, are also called the inner triangle (IT) and the outer triangle (OT), respectively (see [6]). Clearly, in the absence of jitter ($\phi(\omega) = 1$), the bispectrum in the inner triangle is just a replica of the bispectrum of the continuous signal, and the bispectrum in the outer triangle is identically zero.

Appendix B. Bispectrum estimation

In this appendix, the bispectrum estimation procedure used in this work is summarized. Given N samples of a zero mean process $\{x(t)\}$:

$$\{x(n)\} = \{x(0), \dots, x(N - 1)\}, \quad n = 0, \dots, N - 1. \tag{B.1}$$

This sequence is divided to K records, each of length M :

$$N = MK. \tag{B.2}$$

A DFT is performed on each of the K records:

$$\begin{aligned} X^{(k)}(\omega_m) &= \sum_{i=0}^{M-1} x^{(k)}(i)\exp(-j\omega_m i), \quad k = 0, \dots, K - 1, \\ \omega_m &= \frac{2\pi m}{M}, \quad m = 0, \dots, M - 1. \end{aligned} \tag{B.3}$$

The bispectrum estimate for the k th record is given by

$$\hat{B}^{(k)}(\omega_m, \omega_n) = X^{(k)}(\omega_m)X^{(k)}(\omega_n)X^{*(k)}(\omega_m + \omega_n). \tag{B.4}$$

In order to get a consistent estimate of the bispectrum the variance of (B.4) need to be reduced. This is done by averaging these estimates over a square of L^2 bifrequencies:

$$\hat{B}_L^{(k)}(\omega_m, \omega_n) = \frac{1}{L^2} \sum_{r=-L/2}^{L/2} \sum_{s=-L/2}^{L/2} \hat{B}^{(k)}(\omega_{m+r}, \omega_{n+s}). \quad (\text{B.5})$$

Finally, the estimates are averaged over the K records:

$$\hat{B}(\omega_m, \omega_n) = \frac{1}{K} \sum_{i=0}^{K-1} \hat{B}_L^{(i)}(\omega_m, \omega_n). \quad (\text{B.6})$$

The estimate (B6) is biased:

$$E\{\hat{B}(\omega_m, \omega_n)\} = B_d(\omega_m, \omega_n) + O\left(\frac{L}{M}\right), \quad (\text{B.7})$$

where $B_d(\omega_m, \omega_n)$ is the bispectrum of the sampled signal. The variance of the estimate is given by

$$\text{Var}\{\text{Re}(\hat{B}(\omega_m, \omega_n))\} = \text{Var}\{\text{Im}(\hat{B}(\omega_m, \omega_n))\} = \frac{M}{2KL^2} S(\omega_m)S(\omega_n)S(\omega_m + \omega_n) + O\left(\frac{L}{N}\right), \quad (\text{B.8})$$

$S(\omega)$ is the spectrum of the sampled process. By appropriate choice of K , L , and M the estimate (B.6) will be asymptotically unbiased and consistent [4, 12].

References

- [1] H. Akaike, "Effect of timing-error on the power spectrum of sampled data", *Ann. Inst. Statist. Math.*, Vol. 11, No. 3, 1960, pp. 145–165.
- [2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958, Chapter 5.
- [3] A.V. Balakrishnan, "On the problem of time jitter in sampling", *IRE Trans. Inform. Theory*, Vol. 8, 1962, pp. 226–236.
- [4] D.R. Brillinger and M. Rosenblatt, "Asymptotic theory of k th order spectra", in: B. Harris, ed., *Spectral Analysis of Time Series*, Wiley, New York, 1967, pp. 135–188.
- [5] M.J. Hinich and J. Dalle Molle, Private communication.
- [6] M.J. Hinich and M.A. Wolinsky, "A test for aliasing using bispectral analysis", *J. Amer. Statist. Assoc.*, Vol. 83, No. 402, 1988, pp. 499–502.
- [7] M.I. Moore, P.J. Thomson and T.G.L. Shirlcliffe, "Spectral analysis of ocean profiles from unequally spaced data", *J. Geophys. Res.*, Vol. 93, No. C1, 1988, pp. 655–664.
- [8] I. Sharfer and H. Messer, "The effect of jittered time-samples on the discrete bispectrum", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process'91*, Toronto, Canada, May 1991.
- [9] I. Sharfer and H. Messer, "The bispectrum of sampled-data. Part I – Detection of sampling jitter", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 41, No. 1, January 1993, 296–312.
- [10] I. Sharfer and H. Messer, "Feasibility study of parameter estimation of random sampling jitter using the bispectrum", *Circuits, Systems, Signal Process.*, Vol. 13, No. 4, 1994, pp. 1–19.
- [11] I. Sharfer and H. Messer, "The bispectrum of sampled-data: Part II – Monte–Carlo simulations of detection and estimation of the sampling jitter", *IEE Trans. Signal Process.*, October 1994, to appear.
- [12] J.W. Van Ness, "Asymptotic normality of bispectral estimates", *Ann. Math. Statist.*, Vol. 37, 1966, pp. 1257–1275.
- [13] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley, New York, 1968.