FEASIBILITY STUDY OF PARAMETER ESTIMATION OF RANDOM SAMPLING JITTER USING THE BISPECTRUM*

Ilan Sharfer[1] and Hagit Messer[2]

Abstract. An actual sampling process can be modeled as a random process, which consists of the regular (uniform) deterministic sampling process plus an error in the sampling times which constitutes a zero-mean noise (the jitter). In this paper we discuss the problem of estimating the jitter process. By assuming that the jitter process is an i.i.d. one, with standard deviation that is small compared to the regular sampling time, we show that the variance of the jitter process can be estimated from the nth order spectrum of the sampled data, $n = 2, 3$, i.e., the jitter variance can be extracted from the 2nd-order spectrum or the 3rd-order spectrum (the bispectrum) of the sampled data, provided the continuous signal spectrum is known. However, when the signal skewness exceeds a certain level, the potential performance of the bispectrum-based estimation is better than that of the spectrum-based estimation. Moreover, the former can also provide jitter variance estimates when the continuous signal spectrum is unknown while the latter cannot. This suggests that the bispectrum of the sampled data is potentially better for estimating any parameter of the sampling jitter process, once the signal skewness is sufficiently large.

1. Introduction and basic theory

The growing demand for high performance signal processing systems requires new ways to improve signal resolution without making use of costly high precision components. One of the popular approaches, which finds increasing use in A/D and D/A converters, is to trade amplitude resolution for timing resolution, by oversampling the signal at a high clock frequency with a low resolution quantizer, and using digital post-filtering. This type of solution places more stringent constraints on the accuracy of the sampling clock, which now becomes one of the key factors in the overall system performance, among other factors such as quantization noise, and various types of electronic noise in the system. Any error in the

precise clock timing is referred to as jitter, and it is often modelled as a discrete wide sense stationary stochastic process. The sampled signal is generally assumed to be a continuous wide sense stationary process, independent of the jitter process.

In most cases, jitter statistics are unknown or only partially known, so a method is sought by which we will be able to detect the existence of jitter, or estimate parameters of its statistics. One such method is based on the spectral estimates of the discrete (sampled) process, which are affected by the jitter. The effect of jitter in sampling on the spectrum and bispectrum of the discrete process was studied previously [1], [2], [8], [9]. Closed-form expressions for the spectrum and bispectrum of the discrete process were given, under the assumption that the jitter process is i.i.d. That is, the actual sampling times are of the form:

$$t_n = nT + \varepsilon_n, \quad n = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

where T is the sampling period and ε_n is the error in the nth sampling time. The error terms, $\{\varepsilon_n\}$, are assumed to be i.i.d. random variables with unknown p.d.f. $f_e(\varepsilon)$. We assume further that the continuous process being sampled is random, ergodic and strictly bandlimited, with the bandwidth smaller than half the sampling rate. Under this model, the spectrum is given by [1], [2]:

$$S_d(\omega) = S_s(\omega)|\Phi(\omega)|^2 + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_s(u)[1 - |\Phi(u)|^2]du, \tag{2}$$

where $S_s(\omega)$ is the discrete spectrum in the case of uniform, jitter-free sampling. The bispectrum is given by [8], [9]:

$$B_d(\omega_1, \omega_2) = \begin{cases} B_s(\omega_1, \omega_2)\Phi(\omega_1)\Phi(\omega_2)\Phi^*(\omega_1 + \omega_2) & (3) \\ \quad + B_1(\omega_1) + B_1(\omega_2) + B_1^*(\omega_1 + \omega_2) + B_0 & (\omega_1, \omega_2) \in IT \\ B_1(\omega_1) + B_1(\omega_2) + B_1(2\pi - \omega_1 - \omega_2) + B_0 & (\omega_1, \omega_2) \in OT, \end{cases}$$

where (ω_1, ω_2) is a bifrequency in the *principal domain* (PD) of the discrete bispectrum [3], which is further divided into the *inner triangle* (IT) and the *outer triangle* (OT) as in [6], and $B_s(\omega_1, \omega_2)$ is the discrete bispectrum in the case of perfect uniform sampling. The jitter by-product terms $B_1(\omega)$ and B_0 are given by:

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

and

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

where

$$\Phi(\omega) = E\{\exp(j\omega\varepsilon_n/T)\} \tag{6}$$

is the characteristic function of the jitter. The results of the present work can be applied also to other models of the sampling process, such as the renewal process sampling, whose effect on the bispectrum was considered in [8].

We have shown in [9] that the presence of jitter in the sampling process can be detected using the bispectrum of any ergodic strictly band-limited signal which is sampled at a rate greater or equal to the Nyquist rate, even if its second and third order (cumulant) spectra are unknown. By noticing that both 2nd and 3rd order spectra of a signal sampled in the presence of jitter depend on the original spectra and the characteristic function of the jitter process, we propose that parameters of the jitter p.d.f. can be extracted from the actual (estimated) spectra, provided that the original spectra are known.

If $\underline{\theta}$ is a parameter vector that is sufficient for a parametric model of the jitter p.d.f., then the p.d.f. of the actual sampled data depends on the p.d.f. of the continuous process that is sampled, the sampling period T and the parameter vector $\underline{\theta}$. Any optimal estimation of $\underline{\theta}$, such as the maximum likelihood (ML) procedure, should be performed directly on the data vector $\mathbf{X}$ (see Figure 1). However, since $\mathbf{X}$ is generally a non-linear function of $\underline{\theta}$ and highly dependent on the random, continuous signal $x(t)$, we propose to apply an estimation procedure on $\hat{\mathbf{B}}(\mathbf{X})$—the bispectrum estimates, since it was successfully used for the detection of jitter [9]. Such a procedure is essentially suboptimal (unless $\hat{\mathbf{B}}(\mathbf{X})$ is a sufficient statistic of $\mathbf{X}$, which is generally not true), but may be much simpler. However, if we assume that the continuous signal 2nd and 3rd order spectra are known, why not use the estimate of the spectrum of the sampled data $\hat{\mathbf{S}}(\mathbf{X})$ which is usually much simpler to calculate? Both spectrum-based and bispectrum-based estimation procedures are possible and suboptimal, so their relative performance should be evaluated. We do that by comparing the performance achieved by the spectral-domain-jitter-estimation and the bispectral-domain-jitter-estimation. The performance measure we use is the Cramér–Rao lower bound in each domain. First, we calculate the spectral-domain Cramér–Rao bound (SDCRB) which assumes the spectrum estimate $\hat{\mathbf{S}}(\mathbf{X})$ as the input data and provides a lower bound on any unbiased estimate of $\underline{\theta}$ given that data $-\hat{\theta}(\mathbf{S})$. Then, we calculate the bispectral-domain Cramér–Rao bound (BDCRB) which assumes the bispectrum estimate $\hat{\mathbf{B}}(\mathbf{X})$ as the input data and provides a lower bound on any unbiased estimate of $\underline{\theta}$ given that data $-\hat{\theta}(\mathbf{B})$ (see Figure 1). Each of these bounds cannot be smaller than the Cramér–Rao lower bound on any unbiased estimate of $\underline{\theta}$ given the actual data $\mathbf{X}$ as the input data $-\hat{\theta}(\mathbf{X})$. However they indicate the best possible performance of the jitter parameter estimation, when a certain nth-order spectral-domain ($n = 2, 3$) estimate is used as the input data. Therefore, by comparing SDCRB and BDCRB one can tell which domain should be used in order to obtain higher estimation performance.

For derivation of the SDCRB and BDCRB we need two sets of assumptions. In the case of the spectrum we assume that the *process* which is sampled is a stationary, zero-mean, Gaussian process. In the case of the nth order spectrum, $n \geq 3$, we assume that its *estimates* are unbiased and normally distributed with positive definite covariance matrix. This later assumption is asymptotically correct [3], [4] are practically useful, since for any complex Gaussian vector of mean η
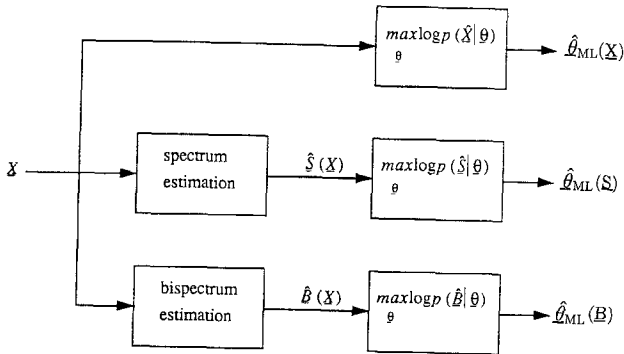
Figure 1. Optimal and suboptimal maximum likelihood estimation of jitter parameters.

and covariance Σ, the Cramér–Rao lower bound matrix on a parameter vector $\underline{\theta}$ is given by J^{-1}, where J, the Fisher information matrix, satisfies:

$$J_{ij} = 2\text{Re}\left\{ \frac{\partial \eta^*}{\partial \theta_i} \Sigma^{-1} \frac{\partial \eta}{\partial \theta_j} \right\} + \text{trace}\left\{ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right\}. \qquad (7)$$

Obviously, it is not possible to assume asymptotic Gaussianity of the 2nd order spectrum estimates.

In this paper we characterize the jitter process by a single parameter—its variance, $\sigma^2 = E\{(\varepsilon_n/T)^2\} = \theta$. This model is valid for any small jitter that has zero mean and is symmetrically distributed [8], [9], since for $\sigma^2 < 1$:

$$\Phi(\omega) \simeq 1 - \frac{\omega^2 \sigma^2}{2} = 1 - \frac{\omega^2 \theta}{2}. \qquad (8)$$

The results, however, can be generalized to any parametric model of the jitter p.d.f.

In Section 2 we derive a closed-form expression for the SDCRB on the variance of the jitter process and we illustrate the results by some examples. In Section 3 we repeat the derivation for the BDCRB, and in Section 4 we compare the performance of the two estimators if indeed the bounds are achievable. In [10] we present a possible bispectral-domain estimator of the jitter variance, and in [14] we compare its performance against the BDCRB by means of Monte-Carlo simulations.

2. Cramér–Rao lower bound in the spectral domain

As we have seen, the existence of timing jitter in the sampling process affects the estimated spectrum. Due to the fact that signal processing in the spectral domain is simpler than in the bispectral domain, it is worthwhile to compare the performance attainable in each domain by means of the Cramér–Rao bound.

2.1. Derivation of the SDCRB

To simplify the analysis, we assume that the sampled continuous process $\{x(t)\}$ is real, stationary Gaussian, with zero mean and finite bandwidth. We assume that we are given N samples of $\{x(t)\}$, taken at time instants corresponding to the i.i.d. jitter model (1):

$$\{x(n)\} = \{x(0), x(1), \ldots, x(N-1)\}, \quad n = 0, \ldots, N-1. \tag{9}$$

Performing the discrete Fourier transform (DFT) on the samples:

$$X(\omega_k) = \sum_{i=0}^{N-1} x(i) e^{-j\omega_k i}; \quad \omega_k = \frac{2\pi k}{N}; \quad k = 1, \ldots, N/2. \tag{10}$$

The DFT can be arranged as a vector:

$$\mathbf{X} = \left[X(\omega_1), \ldots, X(\omega_{N/2}) \right]^T; \quad \dim \mathbf{X} = N/2 \times 1. \tag{11}$$

$\mathbf{X}$ is a zero mean, complex Gaussian vector with positive definite covariance matrix Σ, which tends asymptotically in N to a diagonal matrix (see [5]):

$$E\{\mathbf{X}\} = 0; \quad \mathrm{Cov}\{\mathbf{X}\} = E(\mathbf{X}\mathbf{X}^*) = \Sigma. \tag{12}$$

The diagonal elements of Σ are related to the spectrum in each frequency:

$$\Sigma_{ii} = E\{|X(\omega_i)|^2\} = 2\pi N S_d(\omega_i) + O(1). \tag{13}$$

Since we are assuming that the sampling instants are affected by the jitter process, the spectrum $S_d(\omega)$ in (13) depends on the jitter characteristic function (see (2)). Denoting explicitly the dependence of the characteristic function on θ (the jitter variance), we obtain:

$$S_d(\omega; \theta) = S_s(\omega)|\Phi(\omega; \theta)|^2 + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_s(u)\left[1 - |\Phi(u; \theta)|^2\right] du, \tag{14}$$

where $S_s(\omega)$ is the signal discrete spectrum for $\theta = 0$, i.e., perfect sampling at a rate greater or equal to the Nyquist rate. In this case, θ affects only the covariance matrix $\Sigma(\theta)$. Using (12) and (13), we can write down the likelihood function of θ given $\mathbf{X}$, i.e., the likelihood function in the spectral domain:

$$p_1(\mathbf{X}; \theta) = \pi^{-N/2}|\Sigma(\theta)|^{-1} \exp[-\mathbf{X}^* \Sigma^{-1}(\theta)\mathbf{X}] \tag{15}$$

or by taking the logarithm and using (13):

$$\log p_1(\mathbf{X}; \theta) = -\frac{N}{2} \log \pi - \sum_{i=1}^{N/2} \log[2\pi N S_d(\omega_i; \theta)] - \sum_{i=1}^{N/2} \frac{|X(\omega_i)|^2}{2\pi N S_d(\omega_i; \theta)}. \tag{16}$$

$\frac{|X(\omega_i)|^2}{2\pi N}$ is the periodogram at frequency ω_i, which can be defined, for convenience, as the spectrum estimate $\hat{S}(\omega_i)$. Differentiating (16), the maximum likelihood estimate $\hat{\theta}_{\mathrm{ML}}$ is obtained:

$$\frac{\partial \log p_1(\mathbf{X}; \theta)}{\partial \theta} = \sum_{i=1}^{N/2} \frac{1}{S_d^2(\omega_i; \theta)} \frac{\partial S_d(\omega_i; \theta)}{\partial \theta} \left(\hat{S}(\omega_i) - S_d(\omega_i; \theta)\right)\big|_{\hat{\theta}_{\mathrm{ML}}} = 0. \tag{17}$$

Finding an analytic solution for $\hat{\theta}_{ML}$ is quite difficult. However, the estimate can be found iteratively, with the Newton–Raphson or Scoring algorithms. The Cramér–Rao bound for an unbiased estimate of θ can be found easily:

$$J(\theta) = E\left\{-\frac{\partial^2 \log p_1(\mathbf{X};\theta)}{\partial \theta^2}\right\} = \sum_{i=1}^{N/2} \frac{1}{S_d^2(\omega_i;\theta)}\left[\frac{\partial S_d(\omega_i;\theta)}{\partial \theta}\right]^2; \quad w_i = \frac{2\pi i}{N}.$$

(18)

In (18) we used $E\{\hat{S}(\omega_i)\} = S_d(\omega_i;\theta)$. For large N, the sum (18) can be interchanged with an integral, therefore,

$$J(\theta) = \frac{N}{2\pi}\int_0^\pi \left[\frac{\partial \log S_d(\omega;\theta)}{\partial \theta}\right]^2 d\omega = \frac{N}{2\pi}\int_0^\pi \frac{1}{S_d^2(\omega;\theta)}\left[\frac{\partial S_d(\omega;\theta)}{\partial \theta}\right]^2 d\omega.$$

(19)

The same result can be obtained by using Bangs formula (see [13]). Calculation of (19) is done by substitution of $S_d(\omega;\theta)$ from (14), and is usually computed numerically.

2.2. Numerical examples

In order to illustrate the results developed above, we assume a signal with flat spectrum S and a few common distributions of the jitter process. We consider four kinds of zero-mean, symmetrical jitters with variance $\sigma^2 - E\{(\varepsilon_n/T)^2\}$:

a. Gaussian jitter — $\Phi_g(\omega) = \exp\left(-\frac{\omega^2\sigma^2}{2}\right)$

b. Exponential jitter — $\Phi_e(\omega) = \dfrac{1}{1 + \frac{\omega^2\sigma^2}{2}}$

c. Uniform jitter — $\Phi_u(\omega) = \dfrac{\sin\omega\sqrt{3\sigma^2}}{\omega\sqrt{3\sigma^2}}$

d. Discrete jitter — $\Phi_d(\omega) = \cos(\omega\sigma)$.

First, we calculate the discrete spectrum $S_d(\omega;\theta)$ from (14) and its derivative with respect to θ in each case.

For Gaussian jitter distribution we obtain:

$$S_d(\omega;\theta) = Se^{-\theta\omega^2} + \frac{1}{2\pi}\int_{-\pi}^{\pi} S[1 - e^{-\theta u^2}]du$$

$$= \frac{S}{2\pi}\left[2\pi(e^{-\theta\omega^2} + 1) - \sqrt{\frac{\pi}{\theta}}\operatorname{erf}(\pi\sqrt{\theta})\right].$$

(20)

The derivative with respect to θ is:

$$\frac{\partial S_d(\omega;\theta)}{\partial \theta} = \frac{S}{2\pi}\left[-2\pi\omega^2 e^{-\theta\omega^2} + \operatorname{erf}(\pi\sqrt{\theta})\sqrt{\frac{\pi}{\theta}}\frac{1}{2\theta} - \frac{\pi}{\theta}e^{-\pi^2\theta}\right].$$

(21)

For symmetric exponential distribution we obtain:

$$S_d(\omega; \theta) = S\left[\frac{1}{\left(1 + \frac{\omega^2\theta}{2}\right)^2} + 1 - \frac{1}{2\pi}\left(\frac{\pi}{1 + \frac{\theta\pi^2}{2}} + \sqrt{\frac{2}{\theta}}\, \mathrm{tg}^{-1}\left(\pi\sqrt{\frac{\theta}{2}}\right)\right)\right].$$

(22)

$$\frac{\partial S_d(\omega; \theta)}{\partial \theta} = S\left[-\frac{\omega^2}{\left(1 + \frac{\omega^2\theta}{2}\right)^3} - \frac{1}{2\pi}\left(-\frac{\pi^3}{2\left(1 + \frac{\theta\pi^2}{2}\right)^2}\right.\right.$$

$$\left.\left. - \mathrm{tg}^{-1}\left(\pi\sqrt{\frac{\theta}{2}}\right)\sqrt{\frac{2}{\theta}}\frac{1}{2\theta} + \frac{\pi}{2\theta\left(1 + \frac{\theta\pi^2}{2}\right)}\right)\right].$$

(23)

For discrete jitter distribution we have:

$$S_d(\omega; \theta) = S\left[\cos^2 \omega\sqrt{\theta} + \frac{1}{2} - \frac{\sin 2\pi\sqrt{\theta}}{4\pi\sqrt{\theta}}\right].$$

(24)

$$\frac{\partial S_d(\omega; \theta)}{\partial \theta} = S\left[-\frac{\omega \sin 2\omega\sqrt{\theta}}{2\sqrt{\theta}} - \frac{1}{4\theta}\left(\cos 2\pi\sqrt{\theta} - \frac{\sin 2\pi\sqrt{\theta}}{2\pi\sqrt{\theta}}\right)\right].$$

(25)

The case of uniform distribution was checked as well, but calculations of the sampled spectrum and its derivative were done numerically. In all four cases, the calculation of the bound in (19) was carried out by numeric integration, in which the integration interval $[0, \pi]$ was divided to 50 subintervals.

The following figures present the normalized SDCRB vs. the parameter θ. The normalized SDCRB is defined by:

$$\mathrm{SDCRB}_{\mathrm{norm}} = \frac{\sqrt{(J^{-1}(\theta))}}{\theta}$$

(26)

namely, the ratio between the minimal standard deviation in θ estimate and θ. The number of samples is $N = 1\mathrm{E}4$, assuming that it is sufficiently large for asymptotic behavior of the spectral estimates. We note that the normalized bound in (26) is equivalent to the square root of SDCRB for $\log \theta$ estimate.

Figure 2 shows, that for small values of θ, the normalized bound decreases when θ increases. In addition, the bound exhibits a similar behavior for the distributions that were checked. When θ is smaller than $-10dB$ ($= 0.1$), the bound decreases linearly on a log scale, because $J^{-1}(\theta)$ is nearly constant for small values of θ. It indicates that large values of θ can be estimated with higher precision than small values. However, when θ increases, $J^{-1}(\theta)$ increases as well. The result is that for θ larger than $-10dB$ the normalized bound goes through a transition region and then starts to increase monotonically. The explanation for this effect may be that when θ is large reversal of samples becomes more frequent. For example, assuming uniform jitter, we will have samples reversal when there is an overlap between two adjacent sampling intervals. This occurs at $\theta = 1/12$,
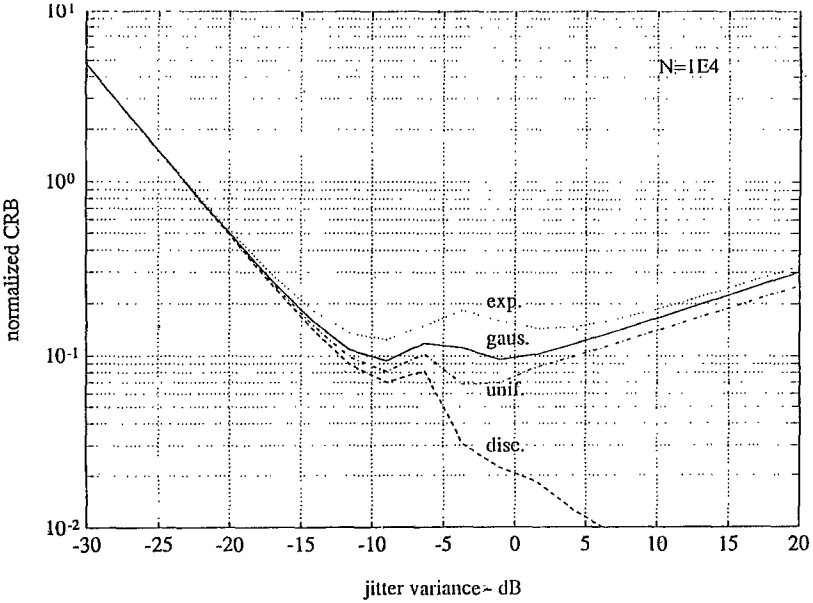
Figure 2. Normalized CRB for jitter estimation in the spectral domain (SDCRB).

which corresponds to a maximal deviation of half the sampling period. We also note the difference between the continuous and discrete jitter distributions. In the latter case, the bound continues to decrease for large values of θ. The reason for that may be understood from (14), which describes the sampled spectrum and contains two terms. The first is simply the original spectrum attenuated by the jitter characteristic function $|\Phi(\omega; \theta)|^2$. The second term is a "white noise" contribution to the spectrum. In the case of a continuous jitter distribution, $|\Phi(\omega; \theta)|$ goes to 0 when θ increases. Therefore, the original contents of the signal disappear, while the white noise term becomes the significant spectral term. For this reason, it is not possible to get a precise estimation of θ, because most of the information in the signal is lost in the sampling process. This loss of information is indicated by the increase in the bound. On the other hand, in the case of discrete jitter distribution, the characteristic function is periodic in $\sigma = \sqrt{\theta}$. As can be seen from (24), only half of the spectrum is "whitened" when θ increases, while the other half is retained. This means, that part of the information still remains, therefore, precise estimation of θ is theoretically still possible, as is evident from the figure.

Clearly, values of θ larger than $-10dB$ are not common in practical situations. For all values of θ, the bound is lowered when the sample size N is increased (the

normalized bound is inversely proportional to the square root of N).

We have noted previously that $J(\theta)$ is nearly constant for small values of θ. This asymptotic value of $J(\theta)$ can be found easily by approximating $\Phi(\omega; \theta)$ by the second moment (assuming the jitter has zero mean):

$$\Phi(\omega; \theta) \simeq 1 - \frac{\omega^2 \theta}{2}, \quad \theta << 1. \tag{27}$$

Using approximation (27) in (14), we obtain the sampled spectrum as:

$$S_d(\omega; \theta) \simeq S_s(\omega)[1 - \omega^2 \theta] + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_s(u)\theta u^2 du \tag{28}$$

and its derivative

$$\frac{\partial S_d(\omega; \theta)}{\partial \theta} \simeq -\omega^2 S_s(\omega) + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_s(u)u^2 du. \tag{29}$$

Substituting (29) in (19) and neglecting the dependence of $S_d(\omega; \theta)$ on θ in the denominator, we get:

$$J(\theta) \simeq \frac{N}{2\pi} \int_0^{\pi} \frac{1}{S_s^2(\omega)} \left(-\omega^2 S_s(\omega) + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_s(u)u^2 du \right)^2 d\omega, \quad \theta \to 0. \tag{30}$$

In the case of a flat spectrum, (30) becomes:

$$J(\theta) \simeq \frac{N}{2\pi} \int_0^{\pi} \left(-\omega^2 + \frac{\pi^2}{3} \right)^2 d\omega = \frac{2\pi^4 N}{45}, \quad \theta \to 0. \tag{31}$$

Figure 3 compares the asymptotic bound (31) to the exact bounds, which were calculated numerically in Figure 2. The figure demonstrates that the exact bounds indeed approach the asymptotic bound when θ is small. At $\theta = -20dB$, the difference between the bounds is not larger than $1dB$.

In this section we have developed the expression for the Cramér–Rao bound in the spectral domain, we have graphically analyzed its dependence on θ, and we have related its behavior to the effect of spectrum whitening. Next, we will repeat the analysis in the bispectral domain and compare the results.

3. Cramér–Rao lower bound in the bispectral domain

3.1. Derivation of the BDCRB

In the previous section we have calculated the CRB in the spectral domain, assuming that the sampled process is Gaussian. In this section we will find the CRB in the bispectral domain, by invoking the asymptotic Gaussianity of the bispectral estimates, as done in [9]. We have seen that the bispectral estimates vector, η, is asymptotically complex Gaussian (see [11]):

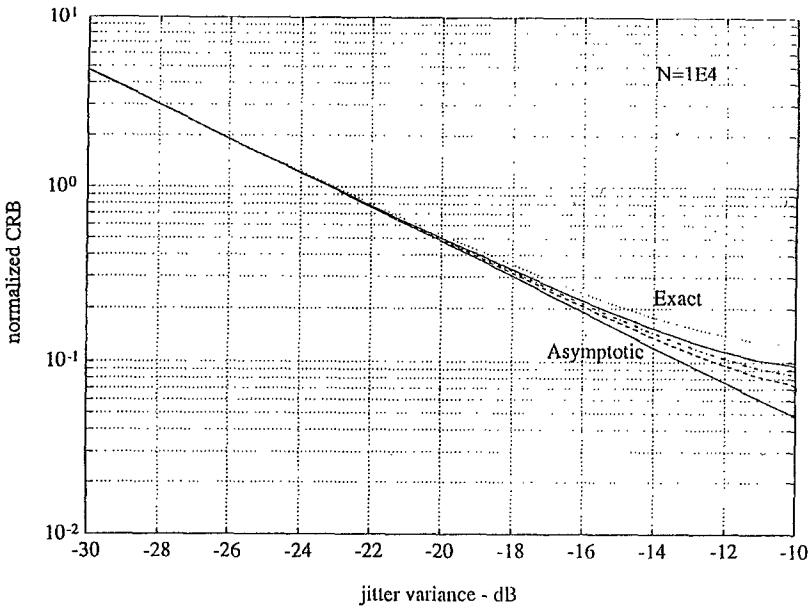$$\eta \sim N_C(\mu(\theta), \Sigma(\theta)), \tag{32}$$

Figure 3. Exact vs. asymptotic normalized SDCRB.

where $\eta(\theta)$ is the expectation of the estimator, and $\Sigma(\theta)$ is the covariance matrix. The expectation $\mu(\theta)$ is related to the original bispectrum, $B_s(\omega_1, \omega_2)$, and to θ by the characteristic function of the jitter:

$$\mu_l(\theta) = B_d(\omega_i, \omega_j, \theta)$$
$$= B_s(\omega_i, \omega_j)\Phi(\omega_i; \theta)\Phi(\omega_j; \theta)\Phi^*(\omega_i + \omega_j; \theta) + B_1(\omega_i; \theta)$$
$$+ B_1(\omega_j; \theta) + B_1^*(\omega_i + \omega_j; \theta) + B_0(\theta), \quad \omega_i, \omega_j \in \text{IT} \qquad (33)$$
$$= B_1(\omega_i; \theta) + B_1(\omega_j; \theta) + B_1(2\pi - \omega_i - \omega_j; \theta) + B_0(\theta), \quad \omega_i, \omega_j \in \text{OT}$$

For all $l = 1, \ldots, P$ and where IT and OT are, respectively, the inner and outer triangles in the principal domain of the bispectrum PD [6], and dim $\mu(\theta) = P \times 1$, $\Sigma(\theta)$ is asymptotically diagonal [9], and is given by:

$$\Sigma_{ll}(\theta) = \frac{M}{KL^2} S_d(\omega_i; \theta) S_d(\omega_j; \theta) S_d(\omega_i + \omega_j; \theta); \quad \omega_i, \omega_j \in \text{PD}; \ l = 1, \ldots, P.$$
$$(34)$$

Recalling the definitions in [9], P is the number of bispectrum estimation points ($P = M^2/12L^2$ in the PD), K is the number of records, M is the number of samples in each record ($N = MK$), and L is the size of the averaging window. $B_s(\omega_i, \omega_j)$ is the bispectrum in the case of perfect sampling ($\theta = 0$).

The calculation of $J(\theta)$ can be done by writing down the Gaussian p.d.f. of the estimates and proceeding as in (18), or more conveniently using Bangs formula:

$$J(\theta) = 2\mathrm{Re}\left\{\frac{\partial \mu^*(\theta)}{\partial \theta}\, \Sigma^{-1}(\theta)\, \frac{\partial \mu(\theta)}{\partial \theta}\right\} + \mathrm{trace}\left\{\left[\Sigma^{-1}(\theta)\frac{\partial \Sigma(\theta)}{\partial \theta}\right]^2\right\}$$

$$= J_1(\theta) + J_2(\theta). \tag{35}$$

Substitution of (33) and (34) in (35) yields for the first part:

$$J_1(\theta) = 2\mathrm{Re}\left\{\frac{\partial \mu^*(\theta)}{\partial \theta}\, \Sigma^{-1}(\theta)\, \frac{\partial \mu(\theta)}{\partial \theta}\right\}$$

$$= 2\sum_{\omega_i,\omega_j \in \mathrm{PD}}\left|\frac{\partial B_d(\omega_i, \omega_j; \theta)}{\partial \theta}\right|^2 \frac{1}{\frac{M}{KL^2}S_d(\omega_i; \theta)S_d(\omega_j; \theta)S_d(\omega_i + \omega_j; \theta)} \tag{36}$$

and for the second part:

$$J_2(\theta) = \mathrm{trace}\left\{\left[\Sigma^{-1}(\theta)\frac{\partial \Sigma(\theta)}{\partial \theta}\right]^2\right\}$$

$$= \sum_{\omega_i,\omega_j \in \mathrm{PD}}\left(\frac{1}{\frac{M}{KL^2}S_d(\omega_i; \theta)S_d(\omega_j; \theta)S_d(\omega_i + \omega_j; \theta)}\cdot\right.$$

$$\left.\cdot\frac{\partial}{\partial \theta}\left[\frac{M}{KL^2}S_d(\omega_i; \theta)S_d(\omega_j; \theta)S_d(\omega_i + \omega_j; \theta)\right]\right)^2 \tag{37}$$

$$= \sum_{\omega_i,\omega_j \in \mathrm{PD}}\left(\frac{\partial \log(S_d(\omega_i; \theta)S_d(\omega_j; \theta)S_d(\omega_i + \omega_j; \theta))}{\partial \theta}\right)^2$$

$$= \sum_{\omega_i,\omega_j \in \mathrm{PD}}\left(\frac{\partial \log S_d(\omega_i; \theta)}{\partial \theta} + \frac{\partial \log S_d(\omega_j; \theta)}{\partial \theta} + \frac{\partial \log S_d(\omega_i + \omega_j; \theta)}{\partial \theta}\right)^2.$$

The summations in (36) and (37) can be converted into integrals, by changing the discrete summation variable into an integration variable

$$\omega_1 = \frac{2\pi L}{M} i; \quad \omega_2 = \frac{2\pi L}{M} j. \tag{38}$$

Therefore, the integral form of (36) is obtained as:

$$J_1(\theta) = \frac{M^2}{4\pi^2 L^2}\frac{2KL^2}{M}\int\!\!\int_{\mathrm{PD}}\left|\frac{\partial B_d(\omega_1, \omega_2; \theta)}{\partial \theta}\right|^2\cdot$$

$$\cdot\frac{1}{S_d(\omega_1; \theta)S_d(\omega_2; \theta)S_d(\omega_1 + \omega_2; \theta)}d\omega_1 d\omega_2 \tag{39}$$

$$= \frac{N}{2\pi^2}\int\!\!\int_{\mathrm{PD}}\left|\frac{\partial B_d(\omega_1, \omega_2; \theta)}{\partial \theta}\right|^2\frac{1}{S_d(\omega_1; \theta)S_d(\omega_2; \theta)S_d(\omega_1 + \omega_2; \theta)}d\omega_1 d\omega_2$$

and for (37) we have:

$$
\begin{aligned}
J_2(\theta) = \frac{M^2}{4\pi^2 L^2} \int \int_{\text{PD}} &\left(\frac{\partial \log S_d(\omega_1; \theta)}{\partial \theta} + \frac{\partial \log S_d(\omega_2; \theta)}{\partial \theta} \right. \\
&\left. + \frac{\partial \log S_d(\omega_1 + \omega_2; \theta)}{\partial \theta} \right)^2 d\omega_1 d\omega_2.
\end{aligned}
\tag{40}
$$

We observe that expression (40) depends on the choice of M and L, which varies according to the estimation method we use. Generally, when a lower bound depends on a parameter, it should be chosen so that it maximizes the bound (minimizes $J(\theta)$), because then the bound is informative. Whenever the bound is minimized, it indicates the optimal performance we might expect, in case the bound is achievable. In our case, the choice of the free parameters is dictated by the condition that the estimator is unbiased and consistent. The estimation bias increases with L, therefore, L should be as small as possible. On the other hand, in order to have a consistent estimate, L should be large enough so that the estimation variance tends to zero, i.e., $L^2 > M$ (see [9]). The choice $L = \sqrt{M}$ satisfies the conditions stated above. Strictly speaking, for consistency one needs $L = M^c$ where $c <$ 0.5. Practically, however, $c = 0.5$ can be assumed. In addition, in order to have maximal resolution in the bispectral estimation (minimal resolution bandwidth), L/M should be minimized; therefore, M should be maximized. Clearly, the relation $N = MK$ leads to the choice $K = 1$, which implies $M = N$. Therefore, the conclusion is that the requirements for unbiasedness, consistency and maximal resolution are satisfied by setting:

$$
M = N; \quad L = \sqrt{N}.
\tag{41}
$$

This choice minimizes the bound and indicates the best performance that may be achieved by using the bispectrum. Substitution of (39), (40) and (41) yields the asymptotic expression for $J(\theta)$:

$$J(\theta) = J_1(\theta) + J_2(\theta)$$

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

$$+ \frac{N}{4\pi^2} \int\int_{PD} \left[\frac{\partial \log S_d(\omega_1; \theta)}{\partial \theta} + \frac{\partial \log S_d(\omega_2; \theta)}{\partial \theta} \right.$$

$$+ \left. \frac{\partial \log S_d(\omega_1 + \omega_2; \theta)}{\partial \theta} \right]^2 d\omega_1 d\omega_2. \tag{42}$$

Analytic calculation of (42) is difficult, but it can be done numerically.

3.2. Numerical examples

To illustrate this result, we will assume that the sampled band limited process has flat spectra

$$B_s(\omega_1, \omega_2) = B; \quad S_s(\omega) = S. \tag{43}$$

We will calculate the bound for the same jitter distributions as in the previous section. The computation is tedious, and it was done by integrating (42) numerically, using expression (33) and (14), respectively, for $B_d(\omega_1, \omega_2; \theta)$ and $S_d(\omega; \theta)$. These computations were carried out with MATLAB.

It can easily be seen that for the process defined by (43) the bound (42) depends on the ratio $\gamma = |B|^2/S^3$, which is defined as the *skewness*, so we shall refer to this quantity in the sequel.

Figure 4 presents the normalized bound in the principal domain of the bispectrum (PD), for each of the jitter distributions, with $N = 1E4$, and $\gamma = 1$. We observe that there is much similarity between the bispectral bound and the spectral bound (compare with Figure 2). The explanation given above for the peculiar character of the bound can be applied in this case as well; in the case of continuous distribution, the original bispectral components gradually disappear when θ increases, while the "white noise" component becomes more significant, especially when $\theta > 0.1$. Therefore, the relative estimation variance increases with θ. In the case of discrete distribution, that constant component is identically zero, but the higher order term $B_1(\omega)$ absorbs part of the original bispectral components. As before, part of the information in the bispectrum is preserved even for large values of θ. This has to do with the fact that the characteristic function $\Phi(\omega; \theta)$ is periodic, and does not tend to zero. Therefore, as the figure clearly shows, the relative accuracy of the estimate is improved for large values of θ.

Figure 5 shows the normalized bound in the bispectrum, in the case of Gaussian jitter. In this figure we study the contribution of each subdomain of the bispectrum to the bound, by limiting the integration region in (42) to the inner triangle (IT), the outer triangle (OT), and the whole principal domain (PD). Here too, we set $N = 1E4$ and $\gamma = 1$. The figure shows clearly that most of the information
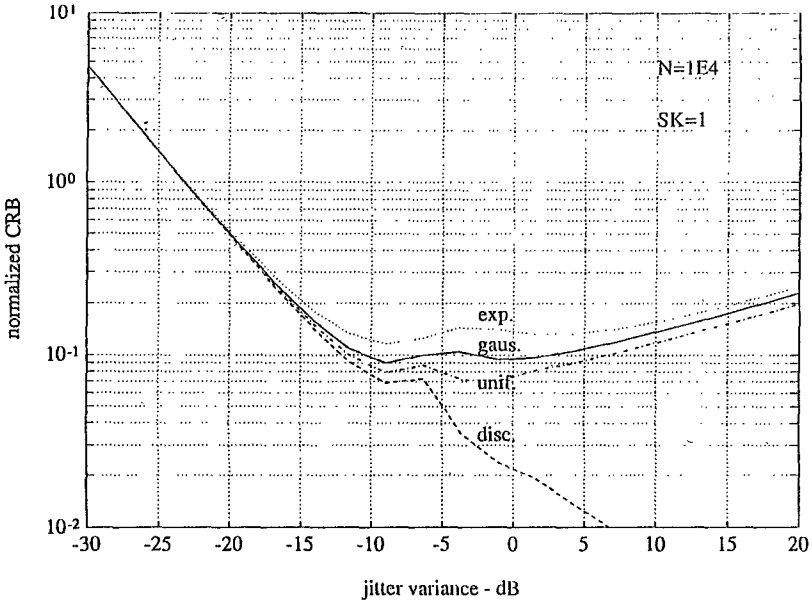
Figure 4. Normalized CRB for jitter estimation in the bispectral domain (BDCRB) for various jitter distributions.
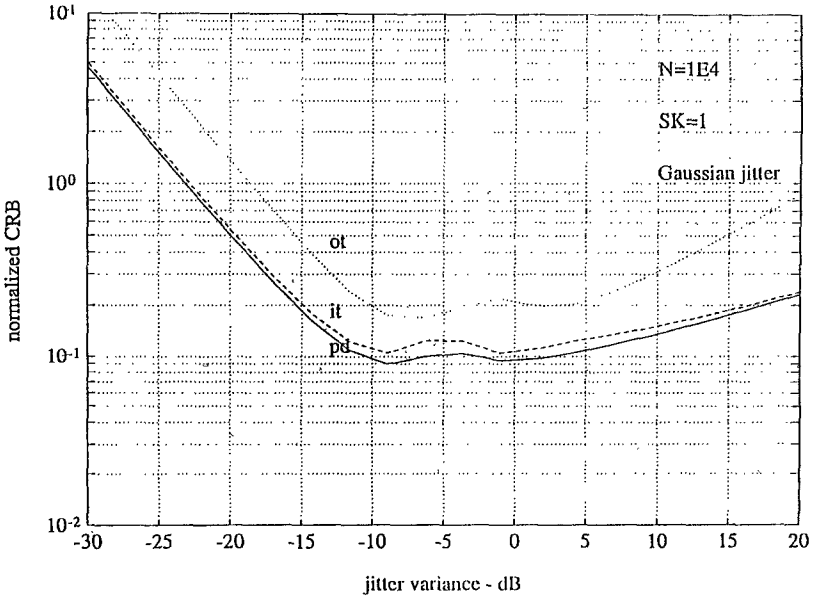

Figure 5. Normalized BDCRB in the entire principal domain (PD), the inner triangle (IT), and the outer triangle (OT).

available in the bispectrum is obtained from the IT. For small values of θ, the performance of the IT bound is approximately 4dB (2.5 times) better than that of the OT bound. This result is not surprising, because in the IT the jitter becomes noticeable, both by its influence on the original bispectrum and by the appearance of the distortion terms $B_1(\omega)$ and B_0, whereas in the OT the estimation is based only on the distortion terms. In addition, there is a 3 to 1 area advantage to the IT. The transition region of the OT bound occurs for slightly larger values of θ than for the IT bound. Later we will compute these bounds when θ tends to zero.

4. Relative performance

In the following figures we will make a direct comparison between the spectral bound (19) and the bispectral bound (42). Figure 6 shows the bounds obtained in each of the two spectral domains, in the case of Gaussian jitter and a Gaussian process, i.e., $\gamma = 0$ and $N = 1E4$. Clearly, a Gaussian process has zero bispectrum, so we do not expect to have a performance gain by using its bispectrum. The figure indeed supports this statement. It shows that although the two bounds have a very similar character, the bispectral bound always lies above the spectral bound in the range of values of θ shown. This result means that if the bounds are achievable, it is not possible to obtain better performance by using the bispectrum instead of the spectrum. This conclusion is true for a Gaussian process, as well as for a non-Gaussian process with zero bispectrum. Since the bispectrum of a Gaussian process is identically zero, the information conveyed by the bispectrum is due to the dependency of the covariance matrix Σ on θ (see (40), and the right-hand side of (42)). Evidently, this information is smaller than the information contained in the spectrum. On the other hand, the spectral CRB is not the global bound because it is not computed directly from the samples, so there may be some significance to the bispectrum, and HOS in general, even in the Gaussian case.

Figure 7 presents the bispectral bound for several values of the skewness γ, with Gaussian jitter and $N = 1E4$. The bound was calculated for three values of γ: 0, 1 and 10. Obviously, when γ increases the bound is lowered, because the term which depends on the bispectrum in (42) increases as well. For the purpose of comparison, the spectral bound is shown as well. It is quite clear from the figure that when $\gamma < 1$, the bispectral bound is still higher than the spectral bound, whereas at $\gamma \simeq 1$ the bounds are about the same. When γ increases further, the bispectral bound becomes lower. In the vicinity of $\gamma = 1$, the two bounds may intersect in some points, but when γ is much larger or much smaller than 1 there is no intersection of the two curves, which means that one of the bounds is superior for all values of θ. The conclusion that can be drawn from the discussion above, is that it may be advantageous to estimate parameters directly from the bispectral estimates, if the signal has sufficiently large skewness, or better yet, to use a combination of the spectrum and bispectrum to improve estimation performance. How to combine both spectra in an optimal fashion has not yet been explored.
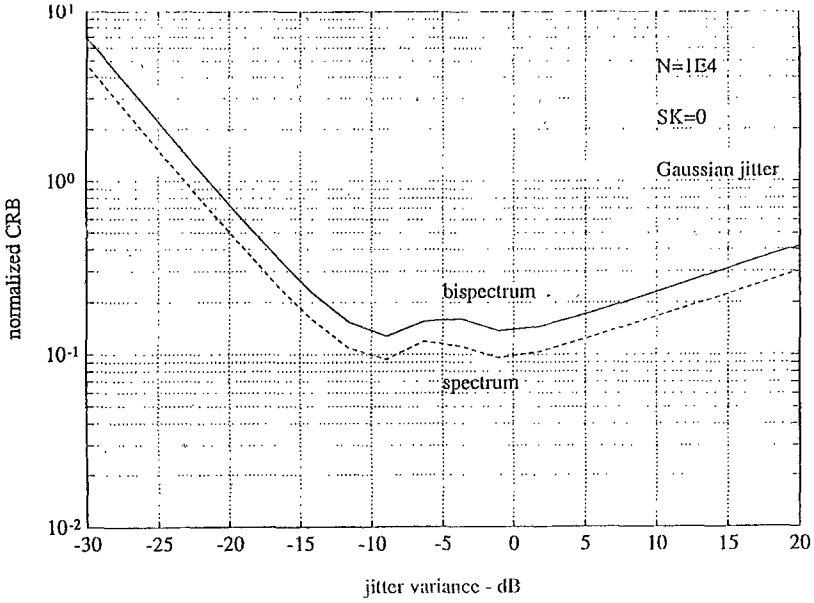
Figure 6. Normalized SDCRB and BDCRB for Gaussian process.
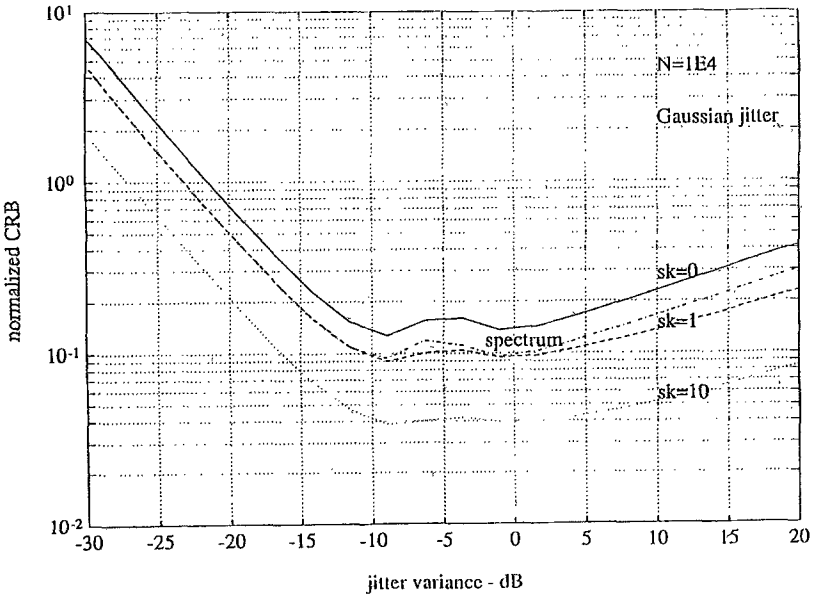

Figure 7. Normalized SDCRB and BDCRB with varying skewness.

As was the case with the spectrum, here too $J(\theta)$ tends asymptotically to a constant value, as θ approaches zero. Again, we approximate the characteristic function $\Phi(w; \theta)$ using (27), assuming the process has flat spectra (see (43)). In the evaluation of (39) we neglect the dependency of $S_d(\omega; \theta)$ on θ in the denominator. In addition, the derivative of $B_0(\theta)$ can be neglected, because $B_0(\theta)$ is proportional to θ^3 (see [8]). Writing down the bispectrum derivatives with respect to θ, we obtain:

$$
\frac{\partial B_2(\omega_1\omega_2; \theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[B\left(1 - \frac{\theta\omega_1^2}{2}\right)\left(1 - \frac{\theta\omega_2^2}{2}\right)\left(1 - \frac{\theta(\omega_1 + \omega_2)^2}{2}\right)\right]
$$

$$
\simeq -\frac{B(\omega_1^2 + \omega_2^2 + (\omega_1 + \omega_2)^2)}{2} \tag{44a}
$$

$$
\frac{\partial B_1(\omega; \theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[B\theta\left(\frac{\pi^2}{3} - \frac{\omega\pi}{2} + \frac{\omega^3}{12\pi}\right)\right] = B\left(\frac{\pi^2}{3} - \frac{\omega\pi}{2} + \frac{\omega^3}{12\pi}\right). \tag{44b}
$$

Substitution of the approximation (44) to (39), and integration over the inner triangle, yields after lengthy calculations:

$$
J_1(\theta)_{\mathrm{IT}} = \frac{N|B|^2}{2\pi^2 S^3}\int\int_{\mathrm{IT}}\left(-\frac{\omega_1^2 + \omega_2^2 + (\omega_1 + \omega_2)^2}{2} + \pi^2\right.
$$

$$
\left. + \frac{\omega_1^3 + \omega_2^3 + (\omega_1 + \omega_2)^3}{12\pi} - (\omega_1 + \omega_2)\pi\right)^2 d\omega_1 d\omega_2 \tag{45}
$$

$$
= \frac{N\gamma\pi^4}{7 \cdot 2^9}\left(97 - \frac{307}{30}\right) \simeq 2.358\,N\gamma.
$$

For the outer triangle we get:

$$
J_1(\theta)_{\mathrm{OT}} = \frac{N|B|^2}{2\pi^2 S^3}\int\int_{\mathrm{OT}}\left(\frac{\omega_1^3}{12\pi} + \frac{\omega_2^3}{12\pi} + \frac{(2\pi - \omega_1 - \omega_2)^3}{12\pi}\right)^2 d\omega_1 d\omega_2
$$

$$
\simeq 0.042\,N\gamma. \tag{46}
$$

To calculate (40), we use (29) for the spectral derivative and we obtain in a similar way the asymptotic value of $J(\theta)$, again neglecting the dependency of $S_d(\omega; \theta)$ on θ in the denominator:

$$
J_2(\theta)_{\mathrm{IT}} = \frac{N}{4\pi^2}\int\int_{\mathrm{IT}}(-\omega_1^2 - \omega_2^2 - (\omega_1 + \omega_2)^2 + \pi^2)^2 d\omega_1 d\omega_2 = \frac{N\pi^4}{60}. \tag{47}
$$

$$
J_2(\theta)_{\mathrm{OT}} = \frac{N}{4\pi^2}\int\int_{\mathrm{OT}}(-\omega_1^2 - \omega_2^2 - (2\pi - \omega_1 - \omega_2)^2)^2 d\omega_1 d\omega_2 = \frac{N\pi^4}{180}. \tag{48}
$$

Combining equations (45)–(48) we finally get the asymptotic $J(\theta)$ in the bispectrum as:

$$
J(\theta) = 2.4\,N\gamma + \frac{N\pi^4}{45}, \quad \theta \to 0. \tag{49}
$$

We can now compare the asymptotic $J(\theta)$ in the spectrum to that of the bispectrum ((31) and (49)), in order to find out for what skewness γ they become equal. This turns out to be:

$$\gamma = \frac{\pi^4}{45 \cdot 2.4} \cong 0.90. \tag{50}$$

This result agrees with what we have seen earlier in Figure 7. It is interesting to note that in the case of a Gaussian process, the asymptotic bispectral bound is exactly twice the spectral bound.

Finally, it should be pointed out that a direct comparison of the spectral and bispectral bounds has to be done with caution. The reason is that the result pertaining to the asymptotic spectral bound is valid only in the case of a Gaussian process, while the result concerning the asymptotic bispectral bound is valid for the non-Gaussian as well as the Gaussian case. Therefore, the comparison that we have made in this section involved two *different* processes, and not the same process. Generally, the Gaussianity assumption is taken as the worst case, in terms of the information we have about the process. It is quite possible that the spectral bound for a certain non-Gaussian process may be lower than the Gaussian bound. Therefore, we may need a higher skewness γ than the one calculated in our example (50) in order to justify the use of the bispectrum. If this is the case, then the use of the bispectrum may be less favorable than we might have expected. Clearly, this is an important issue that has to be checked before considering the use of the bispectrum, and HOS in general, in place of or in addition to the spectrum.

5. Discussion and conclusions

In [9] we have shown that if we are given samples of a strictly bandlimited, continuous, ergodic process taken at a rate that is nominally not smaller than the Nyquist rate, then it is possible to detect the presence of jitter in the sampling process by testing the resultant bispectrum. In this paper we show that if the samples are taken from a process with known spectrum/bispectrum, then the variance of the jitter process can, in principle, be estimated with an accuracy that increases with the signal skewness. The question is, is it possible to estimate parameters of the jitter distribution from the sampled bispectrum even if the signal spectra (spectrum/bispectrum) are unknown, but the process is known to be strictly bandlimited? Our conjecture is that the answer is positive, at least in the case of small jitter. A possible procedure is to detect the presence of the jitter on the outer triangle of the bispectrum and use the statistic to obtain a 1st order estimate to its variance. Then, using this estimate, the original spectra can be recovered. These spectra can be used again to improve the estimate of the parameter, by taking advantage of the whole principal domain and not just the outer triangle. This procedure can be done iteratively and we believe that at least for processes with smooth spectrum/bispectrum, this proposed algorithm converges in probability to the correct value of the parameter. However, this conjecture has still to be investigated.

References

[1] H. Akaike, Effect of timing-error on the power spectrum of sampled data, *Ann. Inst. Statist. Math.* **11**, No. 3, pp. 145–165, 1960.

[2] A. V. Balakrishnan, On the problem of time jitter in sampling, *IRE Trans. Inform. Theory* **8**, pp. 226–236, 1962.

[3] D. R. Brillinger, An introduction to polyspectra, *Ann. Math. Statist.* **36**, pp. 1351–1374, 1965.

[4] D. R. Brillinger and M. Rosenblatt, Asymptotic theory of estimation of k-th order spectra, in *Spectral Analysis of Time Series*, B. Harris, ed., Wiley, New York, pp. 135–188, 1967.

[5] M. J. Hinich and C. S. Clay, The application of the discrete Fourier transform in the estimation of power spectra, coherence and bispectra of geophysical data, *Rev. of Geophys.* **6**, No. 3, pp. 347–363, 1968.

[6] M. J. Hinich and M. A. Wolinsky, A test for aliasing using bispectral analysis, *J. Amer. Statist. Assoc.* **83**, No. 402, pp. 499–502, 1988.

[7] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, 2nd ed., Hafner, New York, 1963.

[8] I. Sharfer and H. Messer, The effect of jittered time-samples on the discrete bispectrum, Proc. IEEE ICASSP-91, Toronto, Canada, May 1991.

[9] I. Sharfer and H. Messer, The bispectrum of sampled data: part I – detection of the sampling jitter, to be published in the *IEEE Trans. on Signal Processing* **141**, pp. 296–312, Jan. 1993.

[10] I. Sharfer and H. Messer, A sub-optimal estimate of the sampling ji Herz variance using the bispectrum, *Signal Processing* **37**(3), 1994.

[11] J. W. van Ness, Asymptotic normality of bispectral estimates, *Ann. Math. Statist.* **37**, pp. 1257–1275, 1966.

[12] H. L. van Trees, *Detection, Estimation and Modulation Theory, Part I*, Wiley, New York, 1968.

[13] A. Zeira and A. Nehorai, Frequency domain Cramér–Rao bound for Gaussian processes, *IEEE Trans. ASSP* **38**, No. 6, pp. 1063–1066, 1990.

[14] I. Sharfer and H. Messer, The bispectrum of sampled-data: part II — Monte Carlo simulations of detection and estimation of the sampling jitter, to be published in the *IEEE Trans. on Signal Processing*.