Petre Stoica, Xiaolei Shang,
and Yuanbo Cheng

The Cramér–Rao Bound for Signal Parameter Estimation From Quantized Data

Several current ultrawide band applications, such as millimeter-wave radar and communication systems [1]–[3], require high sampling rates and therefore expensive and energy-hungry analog-to-digital converters (ADCs). In applications where cost and power constraints exist, the use of high-precision ADCs is not feasible, and the designer must resort to ADCs with coarse quantization. Consequently, the interest in the topic of signal parameter estimation from quantized data has increased significantly in recent years.

Introduction and relevance

The Cramér–Rao bound (CRB) is an important yardstick in any parameter estimation problem. Indeed, it lower-bounds the variance of any unbiased parameter estimator. Moreover, the CRB is an achievable limit; for instance, it is asymptotically attained by the maximum likelihood estimator (under regularity conditions), and thus it is a useful benchmark to which the accuracy of any parameter estimator can and should be compared.

A formula for the CRB for signal parameter estimation from real-valued quantized data has been presented in [4], but its derivation was somewhat sketchy. The said CRB formula has been extended, for example, in [2], to complex-valued quantized data, but again its derivation was rather rough. The special case of binary (1-bit) ADCs and a signal consisting of one sinusoid has been thoroughly analyzed in [5]. The CRB formula for a binary ADC and a general real-valued signal has been derived, e.g., in [6] and [7].

In this lecture note, we will present a textbook derivation of the CRB for a general signal model and quantizer. We also show that the said CRB monotonically decreases and in the limit converges to the standard CRB (for unquantized data) as the quantization becomes increasingly finer. We then consider the special case of binary quantization and present the corresponding CRB. Finally, we show that if the threshold of the binary ADC is allowed to vary in time, then the optimal threshold that minimizes the CRB is the signal itself, and the corresponding CRB is only $\pi/2$ times larger than the standard CRB (a result derived in a different way in [5] for the special case of one sinusoid in noise).

For clarity, in the main part of the lecture note, we focus on the specific case of normally distributed data, which is most commonly examined in the literature. However, in "Extensions to General Distributions," we consider data with an arbitrary distribution and show that the principal results on the CRB derived in the main article for the normal distribution can be readily extended to the general case.

Prerequisites

While we will try to make this lecture note as self-contained as possible, a basic knowledge of statistical signal processing, estimation theory, and calculus will be beneficial to fully understand it.

Problem statement

Consider the following general model for the noisy measurements of a signal:

$$y_n = s_n(\boldsymbol{\theta}) + e_n, \quad n = 1, \ldots, N, \qquad (1)$$

where the index n indicates the sample number, N is the total number of (temporal or spatial) samples, $s_n(\boldsymbol{\theta})$ is the signal model, which is a known (differentiable) function of the unknown parameter vector $\boldsymbol{\theta}$, and e_n denotes the noise. We assume that $\{e_n\}_{n=1}^{N}$ is a sequence of independent identically distributed normal random variables with zero mean and known variance σ^2. We also assume that all quantities in (1) are real valued (the extension to complex-valued measurements is straightforward under the assumption that the normal distribution of the noise is circular; that is, real(e_n) and imaginary(e_n) are independent of each other; see, e.g., [2]). Because σ^2 is known, we can divide both sides of (1) by σ and thus can assume that the noise variance is equal to one, which we will do in what follows

Digital Object Identifier 10.1109/MSP.2021.3116532
Date of current version: 28 December 2021

to simplify the notation. The extension to the case of unknown σ is not difficult, but it leads to more complicated expressions, and, to keep the exposition here as simple as possible, we will not consider it. When there is no risk of confusion, we will also omit the dependence of different functions [such as $s_n(\boldsymbol{\theta})$] on $\boldsymbol{\theta}$ to simplify the notation and some of the expressions in the following sections.

Consider a quantizer with b bits and $A = 2^b$ adjacent intervals defined as follows:

$$\begin{aligned} &I_k = [l_k, u_k), \quad k = 1, \ldots, A, \\ &l_1 = -\infty, \quad u_A = \infty, \\ &l_{k+1} = u_k \quad (k = 1, \ldots, A-1). \end{aligned} \tag{2}$$

When the input to the quantizer lies in I_k, the output, denoted z_k, is given by

$$z_k = Q(y) \quad \text{if} \quad y \in I_k, \tag{3}$$

and it belongs to a set of size A. When the input is y_n, the output will be one of the elements in the said set with an index that depends on n:

$$z_{k(n)} = Q(y_n) \quad \text{for} \quad y_n \in I_{k(n)}. \tag{4}$$

The discussion in this lecture note is valid for any desired set of the quantizer output; consequently, there is no need to specify the $\{z_k\}$ (for example, in the case of binary ADCs, this set can be $\{-1, 1\}$ or $\{0, 1\}$ and so on).

Let $L(\boldsymbol{\theta})$ denote the likelihood function of $\{z_{k(1)}, \ldots, z_{k(N)}\}$, and let $\boldsymbol{J}$ be the Fisher information matrix (FIM) (see [8] and [9]):

$$\boldsymbol{J} = E\left[\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right]. \tag{5}$$

Because

$$\text{CRB} = \boldsymbol{J}^{-1} \tag{6}$$

(whenever $\boldsymbol{J}$ is nonsingular), our problem is to derive an expression for $\boldsymbol{J}$, which we will do in the next section.

The FIM for general quantizers

Derivation of the FIM formula

Under the assumptions made, $\{y_n\}_{n=1}^N$ are independent random variables, and, therefore, so are $\{z_{k(n)}\}_{n=1}^N$. This observation implies that

$$L(\boldsymbol{\theta}) = p(z_{k(1)}, \ldots, z_{k(N)}) = \prod_{n=1}^{N} p(z_{k(n)}), \tag{7}$$

where

$$\begin{aligned} p(z_{k(n)}) &= p(y_n \in I_{k(n)}) \\ &= \phi(u_{k(n)} - s_n) - \phi(l_{k(n)} - s_n) \end{aligned} \tag{8}$$

with $\phi(x)$ being the cumulative distribution function of the normal standard distribution:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt. \tag{9}$$

From (7), we have that

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{n=1}^{N} \frac{\partial p(z_{k(n)})/\partial \boldsymbol{\theta}}{p(z_{k(n)})}. \tag{10}$$

Extensions to General Distributions

The Fisher information matrix (FIM) formula in (15),

$$\boldsymbol{J} = \sum_{n=1}^{N}\left[\sum_{k=1}^{A} \frac{[\phi'(u_k - s_n) - \phi'(l_k - s_n)]^2}{\phi(u_k - s_n) - \phi(l_k - s_n)}\right] \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T}, \tag{S1}$$

where now $\phi(x)$ denotes a general (differentiable) cumulative distribution function (cdf), holds for any distribution of noise [indeed, only the specific expression for the FIM in (16) relies on the normal noise assumption; all of the other calculations in the proof of (15) are valid for a general cdf].

The inequality in (17), namely

$$\check{\boldsymbol{J}} \geq \boldsymbol{J}, \tag{S2}$$

also holds in general as its derivation did not rely on any distributional assumption.

The standard CRB (for unquantized data), which generalizes (23) to an arbitrary distribution (with zero mean and finite variance), is given by (see, e.g., [S1])

$$\boldsymbol{J}_0 = \rho_0 \sum_{n=1}^{N} \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T}, \tag{S3}$$

where

$$\rho_0 = \int_{-\infty}^{\infty} \frac{[\phi''(x)]^2}{\phi'(x)} dx \tag{S4}$$

and where it was implicitly assumed that $\phi'(x) > 0$ for any $|x| < \infty$. The result that $\rho \leq \rho_0$ and, therefore,

$$\boldsymbol{J} \leq \boldsymbol{J}_0 \tag{S5}$$

also holds in the present general case, as can be seen directly from (26). An interesting fact in this context is that ρ_0 in (S4) is lower bounded by 1 (see [S1]), which means that the normal distribution has the smallest matrix $\boldsymbol{J}_0$ (in the order of positive semidefinite matrices) in the class of standard FIMs. Note that this is not true for quantized data: for a given set of intervals $\{I_k\}$ the matrix $\boldsymbol{J}$ associated with the normal distribution is not necessarily the smallest FIM (indeed we have verified numerically that the FIM for the Laplace distribution, for example, may be smaller for some intervals $\{I_k\}$).

Finally, we show that the convergence result (28) or, equivalently, (29) continues to hold in the general case. To that end, we use the result in (S17) in "The Convergence of ρ," whose derivation did not rely on the normal distribution assumption,

$$\rho \to \int_{-\infty}^{\infty} \frac{[\phi'(x)]^2}{\phi'(x)} dx = \rho_0, \tag{S6}$$

which proves that $\boldsymbol{J}$ converges to $\boldsymbol{J}_0$.

Reference

[S1] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramér-Rao bound [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 132–133, 2011. doi: 10.1109/MSP.2011.940411.

Using (8) and the following standard property of $\phi(x)$:

$$\phi'(x) = \frac{d\phi(x)}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

(the probability density function), (11)

we obtain

$$\frac{\partial p(z_{k(n)})}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{[l_{k(n)} - s_n]^2}{2}} - e^{-\frac{[u_{k(n)} - s_n]^2}{2}} \right] \frac{\partial s_n}{\partial \boldsymbol{\theta}}, \quad n = 1, \ldots, N. \tag{12}$$

Because $\{z_{k(n)}\}$ is a sequence of independent random variables, the terms in (10) are independent of each other. Furthermore, the mean of these terms is equal to zero (for $n = 1, \ldots, N$):

$$E\left[\frac{\partial p(z_{k(n)})/\partial \boldsymbol{\theta}}{p(z_{k(n)})} \right] = \sum_{k=1}^{A} \frac{\partial p(z_k)/\partial \boldsymbol{\theta}}{p(z_k)} p(z_k) = \frac{\partial}{\partial \boldsymbol{\theta}} \underbrace{\left[\sum_{k=1}^{A} p(z_k) \right]}_{=1} = 0. \tag{13}$$

It follows from (5), (10), and the preceding discussion, that

$$\boldsymbol{J} = \sum_{n=1}^{N} \sum_{k=1}^{A} \frac{\partial p(z_k)/\partial \boldsymbol{\theta}}{p(z_k)} \frac{\partial p(z_k)/\partial \boldsymbol{\theta}^T}{p(z_k)} p(z_k) = \sum_{n=1}^{N} \sum_{k=1}^{A} \frac{\partial p(z_k)/\partial \boldsymbol{\theta}\, \partial p(z_k)/\partial \boldsymbol{\theta}^T}{p(z_k)}. \tag{14}$$

Inserting (8) and (12) in (14) yields the following expression for $\boldsymbol{J}$:

$$\boldsymbol{J} = \sum_{n=1}^{N} \left[\sum_{k=1}^{A} \frac{[\phi'(u_k - s_n) - \phi'(l_k - s_n)]^2}{\phi(u_k - s_n) - \phi(l_k - s_n)} \right] \times \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T} \tag{15}$$

or, more explicitly,

$$\boldsymbol{J} = \frac{1}{2\pi} \sum_{n=1}^{N} \left[\sum_{k=1}^{A} \frac{\left[e^{-\frac{(u_k - s_n)^2}{2}} - e^{-\frac{(l_k - s_n)^2}{2}} \right]^2}{\phi(u_k - s_n) - \phi(l_k - s_n)} \right] \times \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T}. \tag{16}$$

We note in passing that (13) is rarely mentioned in the derivations of the CRB in the literature. However, without (13), the expression for $\boldsymbol{J}$ would be more complicated [in particular, it would include all cross-terms in the product of (10) with its transpose].

Interval splitting increases the FIM

Let $\{\tilde{I}_k\}_{k=1}^{\tilde{A}}$ denote a set of intervals obtained by splitting some or all of $\{I_k\}_{k=1}^{A}$ in smaller subintervals (hence, $\tilde{A} > A$), and let $\tilde{\boldsymbol{J}}$ be the FIM corresponding to $\{\tilde{I}_k\}$. Intuitively, we would expect that $\tilde{\boldsymbol{J}}$ dominates $\boldsymbol{J}$, i.e.,

$$\tilde{\boldsymbol{J}} \geq \boldsymbol{J} \tag{17}$$

in the sense that $(\tilde{\boldsymbol{J}} - \boldsymbol{J})$ is a positive semidefinite matrix. To prove (17), we introduce the following notation (for simplification, we omit the dependence of some of these variables on k and n):

$$\begin{aligned} &m_k \in [l_k, u_k), \\ &a = \phi'(l_k - s_n) - \phi'(m_k - s_n), \\ &b = \phi'(m_k - s_n) - \phi'(u_k - s_n), \\ &\alpha = \phi(u_k - s_n) - \phi(m_k - s_n), \\ &\beta = \phi(m_k - s_n) - \phi(l_k - s_n). \end{aligned} \tag{18}$$

In what follows, we prove that

$$\frac{(a+b)^2}{\alpha + \beta} \leq \frac{a^2}{\alpha} + \frac{b^2}{\beta}, \tag{19}$$

which clearly implies (17). A simple calculation shows that (19) is equivalent to the following inequalities:

$$\begin{aligned} (19) &\Leftrightarrow \alpha\beta(a+b)^2 \leq (\alpha + \beta)(\beta a^2 + \alpha b^2) \\ &\Leftrightarrow (\alpha b - \beta a)^2 \geq 0. \end{aligned} \tag{20}$$

Because the last inequality is obviously true, the proof of (17) is concluded.

The optimal splitting point m_k that maximizes the increase of the FIM could be determined by maximizing the right-hand side of (19). However, it will depend not only on k but also on s_n; hence, it would be of little use from a practical standpoint. This appears to be a general problem for any attempt to optimize the intervals of the quantizer, and some efforts to circumvent it by assuming that $s_n \approx 0$ (for $n = 1, \ldots, N$) (see, e.g., [4]) are bound to have only limited success. Optimizing the intervals $\{I_k\}$ by maximizing the FIM is an interesting research problem that awaits a practically useful general solution (an efficient global solver for this interval design problem will be presented in a forthcoming article [10]).

Upper and lower bounds on the FIM

Let

$$\rho = \sum_{k=1}^{A} \frac{[\phi'(\tilde{u}_k) - \phi'(\tilde{l}_k)]^2}{\phi(\tilde{u}_k) - \phi(\tilde{l}_k)}, \tag{21}$$

where $\phi(x)$ and $\phi'(x)$ are as defined in (9) and (11), and

$$\tilde{u}_k = u_k - s_n, \ \tilde{l}_k = l_k - s_n \ (\tilde{u}_k > \tilde{l}_k) \tag{22}$$

(we omit the dependence of ρ, $\tilde{u}_k$, and $\tilde{l}_k$ on n to simplify the notation).

In comparison with (15), the standard CRB (for unquantized data) has a simpler expression that corresponds to setting $\rho \equiv 1$ in (15):

$$\boldsymbol{J}_0 = \sum_{n=1}^{N} \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T} \tag{23}$$

(see, e.g., [8] and [9]). Because the output of the quantizer provides less "information" about the signal than the unquantized data $\{y_n\}$, we expect that

$$\boldsymbol{J} \leq \boldsymbol{J}_0. \tag{24}$$

Proving that (24) indeed holds is an interesting exercise that we undertake in what follows. First, we note that

$$\begin{aligned} &[\phi'(\tilde{u}_k) - \phi'(\tilde{l}_k)]^2 \\ &= \left[\int_{\tilde{l}_k}^{\tilde{u}_k} \phi''(x)\, dx \right]^2 \\ &= \left[\int_{\tilde{l}_k}^{\tilde{u}_k} \frac{\phi''(x)}{\sqrt{\phi'(x)}} \sqrt{\phi'(x)}\, dx \right]^2 \\ &\leq \left[\int_{\tilde{l}_k}^{\tilde{u}_k} \phi'(x)\, dx \right] \left[\int_{\tilde{l}_k}^{\tilde{u}_k} \frac{[\phi''(x)]^2}{\phi'(x)}\, dx \right] \\ &= [\phi(\tilde{u}_k) - \phi(\tilde{l}_k)] \int_{\tilde{l}_k}^{\tilde{u}_k} \frac{[\phi''(x)]^2}{\phi'(x)}\, dx. \end{aligned} \tag{25}$$

The inequality in (25) follows from the Cauchy–Schwarz inequality for integrals. Using (25) in the definition of ρ yields the following inequality:

$$\begin{aligned} \rho &\leq \sum_{k=1}^{A} \int_{\tilde{l}_k}^{\tilde{u}_k} \frac{[\phi''(x)]^2}{\phi'(x)} dx \\ &= \int_{-\infty}^{\infty} \frac{[\phi''(x)]^2}{\phi'(x)} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx. \end{aligned} \tag{26}$$

However, the last expression in (26) is nothing but the second-order moment of a normal random variable with zero mean and variance equal to 1. Therefore,

$$\rho \le 1, \tag{27}$$

and (24) is proved.

It follows from (27) and the analysis in the section "Interval Splitting Increases the FIM," that, as the splitting of the intervals becomes finer, the corresponding values of ρ form a monotonically increasing sequence of numbers that is bounded above by 1. The monotonic convergence theorem (see, e.g., [11]) then implies that this sequence converges. In "The Convergence of ρ," we prove that

$$\rho \to 1, \tag{28}$$

and thus

$$\boldsymbol{J} \to \boldsymbol{J}_0 \tag{29}$$

as $A \to \infty$ and the quantization becomes infinitely fine.

However, for a finite A, in general, $\boldsymbol{J}$ will be strictly less than $\boldsymbol{J}_0$. In the last part of this section, we will derive a lower bound on $\boldsymbol{J}$ that can be used to determine how large the difference $(\boldsymbol{J}_0 - \boldsymbol{J})$ can be. To that end, we will use the notation introduced in (22) and the following result, which is a consequence of the Cauchy–Schwarz inequality:

$$\left(\sum_{k=1}^{A} |a_k|\right)^2 = \left(\sum_{k=1}^{A} \frac{|a_k|}{\sqrt{b_k}}\sqrt{b_k}\right)^2 \le \left(\sum_{k=1}^{A} \frac{a_k^2}{b_k}\right)\left(\sum_{k=1}^{A} b_k\right) \text{ for } \{b_k > 0\}_{k=1}^{A},$$

which implies that

$$\sum_{k=1}^{A} \frac{a_k^2}{b_k} \ge \left(\sum_{k=1}^{A} |a_k|\right)^2 \text{ for } \{b_k > 0\}_{k=1}^{A} \text{ and } \sum_{k=1}^{A} b_k = 1. \tag{30}$$

Using the following definitions:

$$|a_k| = \left|e^{-\frac{\tilde{u}_k^2}{2}} - e^{-\frac{\tilde{l}_k^2}{2}}\right|, \quad b_k = \phi(\tilde{u}_k) - \phi(\tilde{l}_k), \tag{31}$$

in (30), we obtain the inequality

$$\rho = \frac{1}{2\pi}\sum_{k=1}^{A} \frac{a_k^2}{b_k} \ge \frac{1}{2\pi}\left[\sum_{k=1}^{A} |a_k|\right]^2. \tag{32}$$

Now, let d be such that $s_n \in [l_d, u_d)$ or, equivalently,

$$\begin{cases} \tilde{l}_k \le 0 & \text{for } k = 1, \ldots, d \\ \tilde{l}_k > 0 & \text{for } k = d+1, \ldots, A \end{cases} \tag{33}$$

(the dependence of d on n is omitted to simplify the notation; this dependence on n will be reinstated when it becomes important). Using the preceding definition of d, it is straightforward to check that

$$\begin{aligned} \eta &\triangleq \frac{1}{2}\sum_{k=1}^{A} |a_k| \\ &= \frac{1}{2}\Big[\Big(e^{-\frac{\tilde{u}_1^2}{2}} - e^{-\frac{\tilde{l}_1^2}{2}}\Big) + \Big(e^{-\frac{\tilde{u}_2^2}{2}} - e^{-\frac{\tilde{l}_2^2}{2}}\Big) \\ &\quad + \cdots + \Big|e^{-\frac{\tilde{u}_d^2}{2}} - e^{-\frac{\tilde{l}_d^2}{2}}\Big| + \Big(e^{-\frac{\tilde{l}_{d+1}^2}{2}} \\ &\quad - e^{-\frac{\tilde{u}_{d+1}^2}{2}}\Big) + \Big(e^{-\frac{\tilde{l}_{d+2}^2}{2}} - e^{-\frac{\tilde{u}_{d+2}^2}{2}}\Big) + \cdots \\ &\quad + \Big(e^{-\frac{\tilde{l}_A^2}{2}} - e^{-\frac{\tilde{u}_A^2}{2}}\Big)\Big] \\ &= \begin{cases} e^{-\frac{\tilde{u}_d^2}{2}} & \text{if } \tilde{u}_d^2 \le \tilde{l}_d^2 \\ e^{-\frac{\tilde{l}_d^2}{2}} & \text{if } \tilde{u}_d^2 > \tilde{l}_d^2. \end{cases} \end{aligned} \tag{34}$$

The Convergence of ρ

We will use the following notation:

$$\mu_k = \phi(\tilde{l}_k), \quad \nu_k = \phi(\tilde{u}_k) \tag{S7}$$

and note that

$$0 = \mu_1 < \nu_1 \ldots \mu_A < \nu_A = 1 \,(\text{i.e.}, [0,1) = \sqcup_{k=1}^{A}[\mu_k, \nu_k)). \tag{S8}$$

Since $\phi(t)$ is a strictly monotonically increasing function, its inverse function $\phi^{-1}(t)$ exists, and therefore

$$\tilde{l}_k = \phi^{-1}(\mu_k), \quad \tilde{u}_k = \phi^{-1}(\nu_k). \tag{S9}$$

Using this notation, ρ can be written as

$$\begin{aligned} \rho &= \sum_{k=1}^{A} \frac{[\phi'(\tilde{u}_k) - \phi'(\tilde{l}_k)]^2}{\phi(\tilde{u}_k) - \phi(\tilde{l}_k)} \\ &= \sum_{k=1}^{A}\left[\frac{\phi'(\phi^{-1}(\nu_k)) - \phi'(\phi^{-1}(\mu_k))}{\nu_k - \mu_k}\right]^2 \cdot (\nu_k - \mu_k). \end{aligned} \tag{S10}$$

Next, we define a function $\psi : [0,1] \to \mathbb{R}$ as follows:

$$\psi(t) = \phi'(\phi^{-1}(t)) \tag{S11}$$

and rewrite (S10) as

$$\rho = \sum_{k=1}^{A}\left[\frac{\psi(\nu_k) - \psi(\mu_k)}{\nu_k - \mu_k}\right]^2 \cdot (\nu_k - \mu_k). \tag{S12}$$

Making use of the Lagrange mean value theorem yields

$$\begin{aligned} \rho &= \sum_{k=1}^{A} [\psi'(\zeta_k)]^2 \cdot (\nu_k - \mu_k), \quad \zeta_k \in (\mu_k, \nu_k) \\ &\to \int_0^1 [\psi'(t)]^2 dt, \quad (\text{as } A \to \infty \text{ and } \max_{1 \le k \le A}\{\nu_k - \mu_k\} \to 0). \end{aligned} \tag{S13}$$

To evaluate the preceding integral, we use the following facts:

$$(\phi^{-1})'(t) = \frac{1}{\phi'(\phi^{-1}(t))} \tag{S14}$$

and

$$\begin{aligned} \psi'(t) &= \phi''(\phi^{-1}(t)) \cdot (\phi^{-1})'(t) \quad \text{(the chain rule)} \\ &= \frac{\phi''(\phi^{-1}(t))}{\phi'(\phi^{-1}(t))}. \end{aligned} \tag{S15}$$

Consider the following change of variable:

$$x = \phi^{-1}(t) \Leftrightarrow t = \phi(x) \tag{S16}$$

(hence, $t_0^1 \Rightarrow x|_{-\infty}^{\infty}$ and $dt = \phi'(x)dx$). Using (S15), we can rewrite (S13) as

$$\rho \to \int_{-\infty}^{\infty}\left[\frac{\phi''(x)}{\phi'(x)}\right]^2 \phi'(x)dx = \int_{-\infty}^{\infty} \frac{[\phi''(x)]^2}{\phi'(x)}dx. \tag{S17}$$

Making use of the result in (26), we get

$$\rho \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}dx = 1, \tag{S18}$$

and the proof of (28) is finished.

Combining (32) and (34) yields the following lower bound on $\boldsymbol{J}$:

$$\boldsymbol{J} \geq \frac{2}{\pi} \sum_{n=1}^{N} \eta^2(n) \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T}. \quad (35)$$

As already mentioned, s_n lies in the interval $[l_{d(n)}, u_{d(n)})$. The smaller this interval the larger $\eta(n)$, and thus the larger the lower bound in (35).

The FIM for a binary quantizer

Derivation of the FIM formula

For $b = 1$, we have $l_1 = -\infty$, $u_1 = l_2 = 0$, and $u_2 = \infty$. Consequently,

$$\boldsymbol{J}_1 = \frac{1}{2\pi} \sum_{n=1}^{N} \left[\frac{e^{-s_n^2}}{\phi(-s_n)} + \frac{e^{-s_n^2}}{1-\phi(-s_n)} \right] \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T} \quad (36)$$

or, equivalently,

$$\boldsymbol{J}_1 = \frac{1}{2\pi} \sum_{n=1}^{N} \frac{e^{-s_n^2}}{\phi(s_n)\phi(-s_n)} \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T}. \quad (37)$$

A proof of (37) was presented in [7], but it relied on a specific choice of the set $\{z_k\}$ (namely $z_1 = -1$ and $z_2 = 1$), while the preceding derivation holds for any choice of $\{z_k\}$.

Upper and lower bounds on the FIM

The analysis in [5] of the special case of one sinusoid in noise found that

$$\boldsymbol{J}_1 \leq \frac{2}{\pi} \boldsymbol{J}_0, \quad (38)$$

where $\boldsymbol{J}_0$ is the standard FIM; see (23). We will show that (38) also holds for the general signal model considered here, but before doing so, we should like to determine if the upper bound in (38) is achievable. To provide an answer to this question, we will consider quantizers with time-varying thresholds, e.g., like those in [6]. In other words, instead of comparing y_n to zero, we compare it to a known threshold denoted h_n. The only modification of the expression of $\boldsymbol{J}_1$ that this change of threshold entails is that the first factor in (37) should be replaced by

$$\rho_1 = \frac{e^{-(s_n - h_n)^2}}{\phi(s_n - h_n)\phi(h_n - s_n)}. \quad (39)$$

Clearly, for $h_n = s_n$, we get $\rho_1 = 4$, and thus the bound in (38) is attained. Consequently, the optimal threshold that minimizes $\mathrm{CRB} = \boldsymbol{J}_1^{-1}$ is the signal itself. While $\{s_n\}$, of course, is unknown (at least initially), this choice of threshold could be implemented sequentially in n as more data are collected and better estimates of $\{s_n\}$ become available. However, we should keep in mind the fact that generating a time-varying threshold $\{h_n\}$ will require digital-to-analog converters (DACs), which can be as expensive and energy hungry as the ADCs and therefore may also be limited in the number of bits they can use. Therefore, an interesting question is the following: Does a simple binary ADC combined with a DAC that uses b bits to generate an approximation of the optimal threshold $h_n = \hat{s}_n$ (from the most recent estimate $\hat{s}_n$ of s_n) yield a better (or worse) performance than using only one b-bit ADC? (See the discussion at the end of this section and the section "Numerical Example" for a partial answer.)

We now return to the inequality in (38) and present a proof of it based on elementary arguments. Clearly, (38) follows if we can show that

$$\frac{e^{-s^2}}{\phi(s)\,\phi(-s)} \leq 4, \forall s \quad (40)$$

or, equivalently,

$$f(s) \triangleq 4\phi(s)\phi(-s) - e^{-s^2} \geq 0, \forall s. \quad (41)$$

For later use, note that

$$f(-\infty) = 0, \quad f(0) = 0, \quad \text{and} \quad f(\infty) = 0. \quad (42)$$

A simple calculation yields

$$\begin{aligned} f'(s) &= \frac{4}{\sqrt{2\pi}} \left[e^{-\frac{s^2}{2}} \phi(-s) - e^{-\frac{s^2}{2}} \phi(s) \right] + 2se^{-s^2} \\ &= 2e^{-\frac{s^2}{2}} g(s), \end{aligned} \quad (43)$$

where

$$g(s) = \frac{2}{\sqrt{2\pi}} [\phi(-s) - \phi(s)] + se^{-\frac{s^2}{2}}. \quad (44)$$

Clearly,

$$f'(s) = 0 \Leftrightarrow g(s) = 0 \text{ and } \operatorname{sign}[f'(s)] = \operatorname{sign}[g(s)], \forall\, |s| < \infty. \quad (45)$$

Also,

$$g(-\infty) = \sqrt{\frac{2}{\pi}}, g(0) = 0, \text{ and } g(\infty) = -\sqrt{\frac{2}{\pi}}. \quad (46)$$

Next, we note that

$$\begin{aligned} g'(s) &= \frac{2}{2\pi} \left[-e^{-\frac{s^2}{2}} - e^{-\frac{s^2}{2}} \right] + e^{-\frac{s^2}{2}} - s^2 e^{-\frac{s^2}{2}} \\ &= e^{-\frac{s^2}{2}} \left[\left(1 - \frac{2}{\pi} \right) - s^2 \right]. \end{aligned} \quad (47)$$

Let

$$s_1 = -\sqrt{1 - 2/\pi}, s_2 = \sqrt{1 - 2/\pi} \quad (48)$$

denote the finite roots of (47) and observe that

$$g''(s_i) = -2s_i e^{-\frac{s_i^2}{2}}, \quad (i = 1, 2), \quad (49)$$

which implies that

$$\begin{aligned} s_1 &= \text{min point of } g(s), \\ s_2 &= \text{max point of } g(s). \end{aligned} \quad (50)$$

Combining the facts shown previously (see Figure 1 and its caption) illustrates that $f(s)$ has a minimum at $s = 0$ [where $f(0) = 0$] and two maximum points, and therefore it satisfies (41). With this observation, the proof is concluded.

At the end of this section, we present a lower bound on $\boldsymbol{J}_1$ and compare it with the lower bound in (35) on $\boldsymbol{J}$. In the case of a binary quantizer, d in (33) is either 1 or 2, depending on whether $s_n < 0$ or $s_n \geq 0$. In either case, η in (34) is given by $e^{-(s_n^2/2)}$, and, hence, (35) reduces to

$$\boldsymbol{J}_1 \geq \frac{2}{\pi} \sum_{n=1}^{N} e^{-s_n^2} \frac{\partial s_n}{\partial \boldsymbol{\theta}} \frac{\partial s_n}{\partial \boldsymbol{\theta}^T}. \quad (51)$$

The lower bound in (51) will typically be significantly smaller than that in (35), in agreement with the fact that the estimation performance corresponding to a binary quantizer is inferior to that of a quantizer using more bits [see (17)]. On the other hand, if $e^{-s_n^2}$ in (51) is replaced

by $e^{-(h_n - s_n)^2}$ [as in (39)], with the threshold $h_n = \hat{s}_n$ generated using a DAC with b bits, then the lower bounds in (35) and (51) become quite similar to one another. In such a case, if b is sufficiently large (for example, $b \geq 4$), the two bounds are well approximated by $(2/\pi)\boldsymbol{J}_0$. For the binary ADC, this matrix is also an upper bound; see (38). Hence, it corresponds to the apex performance of this quantizer. However, the FIM for an ADC with $b > 1$ can be larger than $(2/\pi)\boldsymbol{J}_0$, which means that the estimation performance afforded by a b-bit ADC can, in principle, be better than that of a 1-bit ADC, even if the latter uses an optimal threshold generated by a b-bit DAC. (See the section "Numerical Example" for an illustration.)

We remark in this context that a tighter upper bound on $\boldsymbol{J}$ than $\boldsymbol{J} \leq \boldsymbol{J}_0$ [similar to the bound in (38)] is not available. The derivation of such a bound on $\boldsymbol{J}$, which would generalize (38) to the case of $b > 1$, is an open research problem. However, we must note that, while such a bound would be theoretically interesting, its practical usefulness would be limited. Indeed, ρ is quite close to one (and hence $\boldsymbol{J}$ to $\boldsymbol{J}_0$), even for relatively small values of b, such as $b = 3$ or $b = 4$ (for which $\max_{\{I_k\}}\rho$ is 0.97 and 0.99, respectively), and for all practical purposes, $\max_{\{I_k\}}\rho$ becomes indistinguishable from one as b increases. Consequently, an efficient and reliable algorithm for computing the optimum $\{I_k\}$ that maximize ρ appears to be a more useful practical contribution than an upper bound on ρ (see the forthcoming article [10] for such an algorithm).

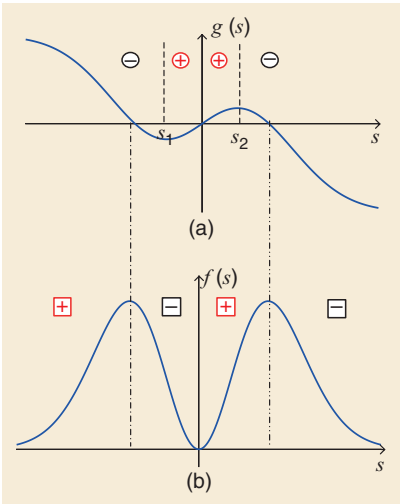

FIGURE 1. (a) $g(s)$ versus s. $\text{sign}[g'(s)]$ (indicated by circles) is positive in $[s_1, s_2]$ and negative elsewhere [see (47)]. This observation along with the fact that s_1 is a minimum of $g(s)$ and s_2 a maximum [see (50)] and that $g(s)$ takes on the values in (46) lead to the plot of $g(s)$ in this part of the figure. (b) $f(s)$ versus s. Using the fact that $\text{sign}[f'(s)] = \text{sign}[g(s)]$ [see (45)], we get the variation of $\text{sign}[f'(s)]$ (indicated by squares). This variation along with the values of $f(s)$ in (42) and the fact that $f'(s) = 0 \Leftrightarrow g(s) = 0$ for any $|s| < \infty$ [see (45)] lead to the plot of $f(s)$ in this figure, which shows that $f(s) \geq 0\ \forall s$.

Numerical example

We consider a signal comprising two sinusoids:

$$s_n(\boldsymbol{\theta}) = a_1 \sin(\omega_1 n + \varphi_1) + a_2 \sin(\omega_2 n + \varphi_2), \quad n = 1, 2, \ldots, N, \tag{52}$$

where

$$\begin{aligned} &N = 100 \text{ or } 512, \\ &\boldsymbol{\theta} = [a_1 \ \ a_2 \ \ \omega_1 \ \ \omega_2 \ \ \varphi_1 \ \ \varphi_2]^T, \\ &\omega_1 = 0.25, \quad \omega_2 = 0.4, \\ &a_1 = 1, \quad a_2 = 1/r, \\ &\varphi_1 = \pi/3, \quad \varphi_2 = \pi/4. \end{aligned} \tag{53}$$

We vary r from 1 to 200. We also vary the noise variance to maintain the same signal-to-noise ratio $(\text{SNR}_2) = 0$ dB for the weaker sinusoid, for all values of r, where

$$\text{SNR}_2 = \frac{a_2^2}{2\sigma^2}. \tag{54}$$

The intervals $\{I_k\}$ are chosen as suggested in [12] for a Lloyd–Max quantizer. In the case of $b = 4$ considered here, and for a signal with unit power, these intervals are given by

$$\{-\infty, -2.401, -1.844, -1.437, -1.099, -0.7996, -0.5224, -0.2582, 0.0000, 0.2582, 0.5224, 0.7996, 1.099, 1.437, 1.844, 2.401, \infty\}.$$

Figures 2 and 3 show four CRBs for ω_1 and ω_2; see the explanations in the figure captions. The standard CRB for ω_1, $\boldsymbol{J}_0^{-1}(\omega_1)$, decreases with r because σ^2 decreases as r increases (as explained previously); hence, SNR_1 increases with r. For the second sinusoid, $\boldsymbol{J}_0^{-1}(\omega_2)$ is constant as r varies because SNR_2 is the same for all values of r.

The 1-bit CRB, $\boldsymbol{J}_1^{-1}$, is the largest of the considered bounds, which is the price paid for the simplicity of the binary quantizer. Moreover, the degradation of $\boldsymbol{J}_1^{-1}$ as r increases is significant and can reach unacceptable levels if N is not large enough. Both $\boldsymbol{J}_4^{-1}$ and $\boldsymbol{J}_{14}^{-1}$ are much smaller than $\boldsymbol{J}_1^{-1}$, and both degrade more gracefully in the case of

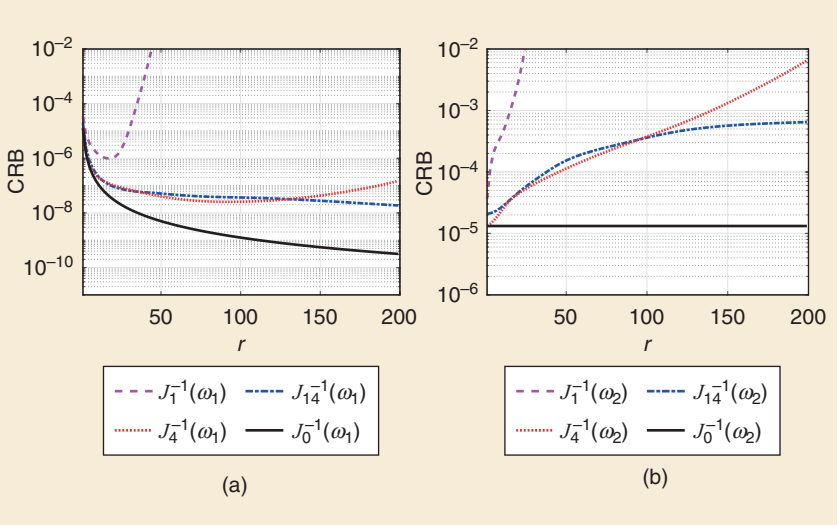

FIGURE 2. CRB versus r for (a) CRB (ω_1), $N = 100$ and (b) CRB (ω_2), $N = 100$. $\boldsymbol{J}_0^{-1}$ = standard CRB (for unquantized data), $\boldsymbol{J}_1^{-1}$ = CRB for 1-bit ADC, $\boldsymbol{J}_4^{-1}$ = CRB for 4-bit ADC, and $\boldsymbol{J}_{14}^{-1}$ = CRB for 1-bit ADC using the optimal threshold $h_n = s_n$ generated via a 4-bit DAC.

ω_2, as r increases, and remain relatively close to $\boldsymbol{J}_0^{-1}$. As one can see from Figures 2 and 3, typically, $\boldsymbol{J}_4^{-1}$ is not far from $\boldsymbol{J}_{14}^{-1}$, despite the fact that the latter is based on information about the signal.

The bowl-shaped plot of $\boldsymbol{J}_1^{-1}(\omega_1)$, in Figures 2(a) and 3(a), can be explained as follows. The second sinusoid acts as (unknown) jitter on the first sinusoid, and it is a well-known fact that jitter can improve the accuracy of parameter estimation from binary data. The optimal power of the jitter (or the corresponding value of r), which minimizes $\boldsymbol{J}_1^{-1}(\omega_1)$, is problem dependent. A similar effect occurs because of the noise (we remind the reader that the noise variance decreases as r increases), which acts as dither on the signal. This jittering/dithering effect causes the decrease of $\boldsymbol{J}_1^{-1}(\omega_1)$ in both Figures 2(a) and 3(a), but as r increases, the said effect vanishes, and $\boldsymbol{J}_1^{-1}(\omega_1)$ starts to increase. If we continue to increase r beyond the range in the figure, $\boldsymbol{J}_1^{-1}(\omega_1)$ increases without bound, a fact which indicates that parameter identifiability has been lost.

Note that the curves of $\boldsymbol{J}_{14}^{-1}(\omega_1)$ and $\boldsymbol{J}_4^{-1}(\omega_1)$ are much more stable for r in the range considered in the figure. However if r is increased beyond a certain level (admittedly too large to be of practical interest), then $\boldsymbol{J}_4^{-1}(\omega_1)$ also starts to increase, and eventually this scheme also loses parameter identifiability. By contrast, in our experiments, $\boldsymbol{J}_{14}^{-1}(\omega_1)$ has continued to decrease, even when we let r take on quite large values.

In sum, the main findings of this numerical evaluation are as follows:

- A 4-bit quantizer appears to offer a satisfactory estimation performance in a wide range of situations, including cases with signal components of rather different powers (but excluding cases in which the signal is almost noise free, a situation that obviously is unlikely to occur in practical applications).
- The use of a 1-bit ADC along with a 4-bit DAC for threshold generation does not appear to offer any advantage, either in hardware or estimation performance, over employing a single 4-bit ADC, even when the threshold generation block had full information about the signal (unless there is very little noise in the data, which is a case of theoretical rather than practical interest).
- When a 1-bit quantizer is the only feasible option, the user should keep in mind that the estimation performance offered by this quantizer can degrade very quickly as the dynamic range of the signal components increases, unless N is sufficiently large. A study of the CRB using the formulas presented in this lecture note can offer guidelines for the cases in which the use of a 1-bit quantizer can be a viable solution.

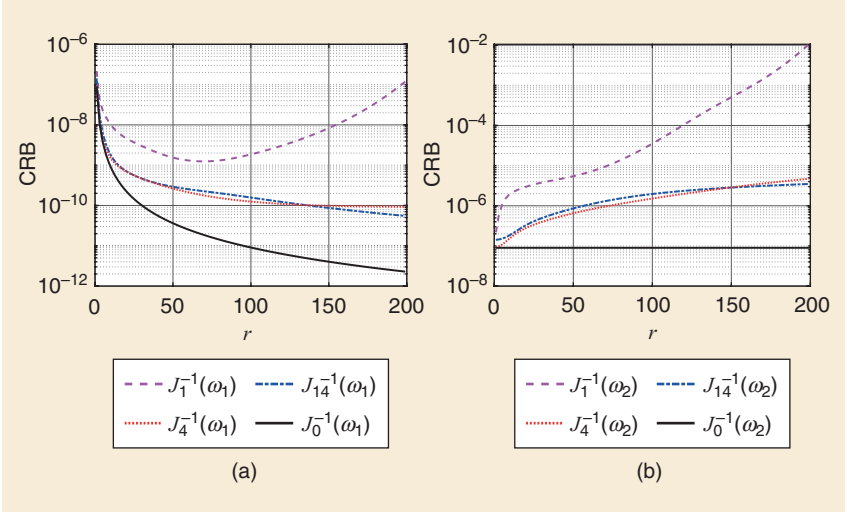

FIGURE 3. CRB versus r for (a) CRB (ω_1), $N = 512$ and (b) CRB (ω_2), $N = 512$. $\boldsymbol{J}_0^{-1}$ = standard CRB (for unquantized data), $\boldsymbol{J}_1^{-1}$ = CRB for 1-bit ADC, $\boldsymbol{J}_4^{-1}$ = CRB for 4-bit ADC, and $\boldsymbol{J}_{14}^{-1}$ = CRB for 1-bit ADC using the optimal threshold $h_n = s_n$ generated via a 4-bit DAC.

What we have learned

It is hoped that the clear derivation and simple form of the CRB expression for signal parameter estimation from quantized data, presented in this lecture note, will encourage the more frequent use of the CRB in applications where coarse-quantization ADCs are the only feasible choice, such as in ultrawide band radar and communication systems or large sensor networks. Proving that the said CRB monotonically decreases and in the limit converges to the standard CRB (for the unquantized data) as the quantization becomes increasingly finer can be viewed as an interesting exercise in statistical signal processing and calculus—and so can the analysis leading to the upper bound on the CRB as well as the achievable lower bound on the CRB associated with 1-bit ADCs. The discussion on the optimal threshold in applications using 1-bit ADCs as well as the performance comparison between a system employing a b-bit ADC and another system using a 1-bit ADC in combination with a b-bit DAC (for threshold generation) are also of potential interest for applications.

Finally, this lecture note has illustrated the fact that, as the number of bits of the quantizers decreases, the estimation accuracy for the small components degrades much more than for the large components of the signal (as intuitively expected: when $A = 2^b$ is small, the few available intervals of the quantizer are normally chosen sufficiently large to cover the variation of the large components in the signal; hence, most information about the small components is lost in the quantized data). This dynamic range problem also appears to be practically relevant and worthy of further study.

Acknowledgments

This work was supported in part by the Swedish Research Council (VR grants 2017-04610 and 2016-06079), in part by the National Natural Science Foundation of China under grant 61771442, and in part by the Key Research Program of Frontier Sciences of the Chinese Academy of Sciences under grant QYZDY-SSW-JSC035.

Authors

Petre Stoica (ps@it.uu.se) is currently a senior professor in signals and systems modeling at Uppsala University, Uppsala, Sweden. He is a member of the Royal Swedish Academy of Engineering Sciences, Stockholm, an international member of the United States National Academy of Engineering, Washington, D.C., an honorary member of the Romanian Academy, Bucharest , a member of the European Academy of Sciences, Liege, Belgium, and a member of the Royal Society of Sciences, Uppsala, Sweden. He is also a fellow of the European Association for Signal Processing and the Royal Statistical Society, a distinguished fellow of the International Engineering and Technology Institute, and a Fellow of IEEE.

Xiaolei Shang (xlshang@mail.ustc.edu.cn) received his B.E. degree from Xidian University, Xi'an, China, in 2018. He is currently working toward his Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, 230027, China. His current research interests include spectral estimation, array signal processing, and their applications. He is a Student Member of IEEE.

Yuanbo Cheng (cyb967@mail.ustc.edu.cn) received his B.S. degree from the University of Science and Technology of China, Hefei, China, in 2019. He is currently working toward his M.S. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, 230027, China. His current research interests include spectral estimation, array signal processing, and their applications.

References

[1] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, 2017. doi: 10.1109/TSP.2017.2781644.

[2] C.-Y. Wu, J. Li, and T. F. Wong, "A Cramér-Rao bound analysis for mmWave PMCW MIMO radar with quantized observations," in *Proc. 54th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, Nov. 2020, pp. 612–616.

[3] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 98–117, 2020. doi: 10.1109/MSP.2020.2978507.

[4] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. Int. ITG Workshop on Smart Antennas (WSA)*, Bremen, Germany, Feb 2010, pp. 143–150.

[5] A. Host-Madsen and P. Handel, "Effects of sampling and quantization on single-tone frequency estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 650–662, 2000. doi: 10.1109/78.824661.

[6] C. Gianelli, L. Xu, J. Li, and P. Stoica, "One-bit compressive sampling with time-varying thresholds: Maximum likelihood and the Cramér-Rao bound," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, Nov. 2016, pp. 309–403.

[7] C. D. Gianelli, "One-bit compressive sampling with time-varying thresholds: The Cramér-Rao bound, maximum likelihood, and sparse estimation," Ph.D. dissertation, Univ. of Florida, 2019.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[9] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice Hall, 2005.

[10] Y. Cheng, X. Shang, and P. Stoica, "Interval design for signal parameter estimation from quantized data," 2021.

[11] W. Rudin, *Principles of Mathematical Analysis*, vol. 3. McGraw-Hill: New York, 1976.

[12] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, 1960. doi: 10.1109/TIT.1960.1057548.

SP