PERFORMANCE BOUNDS FOR THE ESTIMATION OF FINITE RATE OF INNOVATION SIGNALS FROM NOISY MEASUREMENTS

Zvika Ben-Haim, Tomer Michaeli, and Yonina C. Eldar

Technion—Israel Institute of Technology, Haifa 32000, Israel
{zvikabh@tx, tomermic@tx, yonina@ee}.technion.ac.il

ABSTRACT

In this paper, we derive lower bounds on the estimation error of finite rate of innovation signals from noisy measurements. We first obtain a fundamental limit on the estimation accuracy attainable regardless of the sampling technique. Next, we provide a bound on the performance achievable using any specific sampling method. Essential differences between the noisy and noise-free cases arise from this analysis. In particular, we identify settings in which noise-free recovery techniques deteriorate substantially under slight noise levels, thus quantifying the numerical instability inherent in such methods. The results are illustrated in a time-delay estimation scenario.

Index Terms— Finite rate of innovation, Time-delay estimation, Cramér–Rao bound, Union of subspaces

1. INTRODUCTION

The goal of sampling theory is to recover a continuous-time signal $x(t)$ from a discrete set of measurements. The archetypical example in this field is the Shannon sampling theorem, which states that a B-bandlimited function can be reconstructed from samples taken at a rate of $2B$. Recently, considerable attention has been devoted to the extension of this theory to arbitrary signals having a finite rate of innovation (FRI), which are functions parameterized by a finite number ρ of parameters per time-unit. For a variety of families of FRI signals, several existing algorithms are guaranteed to recover the signal $x(t)$ from samples taken at rate ρ [1–6].

Real-world signals are often contaminated by noise and thus do not conform precisely to the FRI model. It is therefore of interest to quantify the effect of noise on FRI techniques. In the noisy case, it is no longer possible to perfectly recover the original signal from its samples. Nevertheless, one might hope for an appropriate finite-rate technique which achieves the best possible estimation accuracy, in the sense that increasing the sampling rate confers no further performance benefits. For example, to recover a B-bandlimited signal contaminated by white noise, one can use an ideal low-pass filter with cutoff B prior to sampling at the rate of innovation, which is $\rho = 2B$. It can be shown that this approach achieves an optimal recovery error, even among techniques having arbitrary sampling rates.

By contrast, empirical observations indicate that, for some FRI signals, any increase in the sampling rate improves estimation accuracy [2–6]. In this paper, we provide analytical justification and quantification of these empirical findings. We first derive the Cramér–Rao bound (CRB) for estimating $x(t)$ directly from continuous-time measurements $y(t) = x(t) + w(t)$, where $w(t)$ is a Gaussian white noise process. This setting is to be distinguished from previous bounds in the FRI literature [7, 8] in three respects. First, the measurements are a continuous-time process $y(t)$ and the bound therefore applies regardless of the sampling method. Second, in our model, the noise is added prior to sampling. Consequently, even sampling at an arbitrarily high rate will not completely compensate for measurement noise. Third, we bound the MSE in estimating $x(t)$ and not the parameters defining it; this involves a delicate analysis of the CRB for estimating a continuous-time function.

In practice, rather than processing the continuous-time signal $y(t)$, it is typically desired to estimate $x(t)$ from a discrete set of samples $\{c_n\}$ of $y(t)$. To quantify the extent to which sampling degrades the ability to recover the signal, we derive the CRB for estimating $x(t)$ from the measurements $\{c_n\}$. This analysis reveals two interesting phenomena. First, desirable sampling techniques can be identified as those whose performance approaches the continuous-time bound. Second, as opposed to the noiseless setting in which ρ samples per time unit typically suffice for recovery, in this setting the MSE depends on the specific structure of the set of feasible signals. In particular, we identify settings in which significant oversampling is required in order to optimally estimate the underlying signal.

We demonstrate our results via the problem of estimating a sequence of pulses having unknown positions and amplitudes [1, 3, 4, 6]. In this case, a simple sufficient condition is obtained for the existence of a sampling scheme whose performance bound coincides with the continuous-time CRB. This scheme is based on sampling the Fourier coefficients of the pulse shape, and is reminiscent of recent time-delay estimation algorithms [6]. However, while the sampling scheme is theoretically sufficient for optimal recovery of $x(t)$, we show that in some cases there is room for substantial improvement in the reconstruction stage of these algorithms.

2. SETTING

A signal $x(t)$ is said to have a rate of innovation ρ if any segment $\{x(t) : t \in [T_1, T_2]\}$ is determined by no more than $(T_2 - T_1)\rho$ parameters. We wish to estimate such signals from noisy measurements. For concreteness, let us focus on the problem of estimating the finite-duration segment $\{x(t) : t \in [0, T]\}$, for some constant T, and assume for simplicity that $K = \rho T$ is an integer. We then have

$$x \in \mathfrak{X} \triangleq \{h_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\} \subset L_2[0, T] \tag{1}$$

where $\{h_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is a set of functions in $L_2[0, T]$ which are parameterized by the deterministic unknown vector $\boldsymbol{\theta}$, and Θ is an open subset of $\mathbb{R}^K$.

We wish to examine the random process

$$y(t) = x(t) + w(t), \quad t \in [0, T] \tag{2}$$

This work was supported in part by a Magneton grant from the Israel Ministry of Industry and Trade, by the Israel Science Foundation under Grant 1081/07, and by the European Commission's FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (grant agreement no. 216715).

where $w(t)$ is continuous-time white Gaussian noise. Since formally it is not possible to define Gaussian white noise over a continuous-time probability space [9], we interpret (2) as a simplified notation for the equivalent set of measurements

$$z(t) = \int_0^t x(\tau)d\tau + \sigma b(t), \quad t \in [0, T] \tag{3}$$

where $b(t)$ is a standard Wiener process (also called Brownian motion) [10]. It follows that $w(t)$ can be considered as a random process such that, for any $f \in L_2$, the integral $\int_0^T f(t)w(t)dt$ is a Gaussian random variable with mean 0 and variance $\sigma^2 \int_0^T |f(t)|^2 dt$ [9].

In this paper, we consider estimators which are functions either of the entire continuous-time process (2) or of some subset of the information present in (2), such as samples of the continuous-time process. To treat these two cases in a unified way, let $(\Omega, \mathscr{F})$ be a measurable space and let $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ be a family of probability measures over $(\Omega, \mathscr{F})$. Let $(\mathcal{Y}, \mathscr{U})$ be a measurable space, and let the random variable $y : \Omega \to \mathcal{Y}$ denote the measurements. An estimator can be defined in this general setting as a measurable function $\hat{x} : \mathcal{Y} \to L_2$. The MSE of an estimator $\hat{x}$ at x is defined as

$$\mathrm{MSE}(\hat{x}, x) \triangleq \int_0^T \mathbb{E}\{|\hat{x}(t) - x(t)|^2\}\, dt. \tag{4}$$

An estimator $\hat{x}$ is said to be unbiased if

$$\mathbb{E}\{\hat{x}(t)\} = x(t) \text{ for all } x \in \mathfrak{X} \text{ and almost all } t \in [0, T]. \tag{5}$$

Our goal in this paper will be to determine the CRB, which is a lower bound on the MSE of unbiased estimators, under several measurement models.

Throughout the paper, we will require the following regularity assumptions.

R1) $h_{\boldsymbol{\theta}}$ is Fréchet differentiable with respect to $\boldsymbol{\theta}$, in the sense that for each $\boldsymbol{\theta}$, there exists a continuous linear operator $\partial h_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta} : \mathbb{R}^K \to L_2$ such that, for any sufficiently small $\boldsymbol{\delta} \in \mathbb{R}^K$,

$$\frac{h_{\boldsymbol{\theta}+\boldsymbol{\delta}} - h_{\boldsymbol{\theta}}}{\|\boldsymbol{\delta}\|} = \frac{\partial h_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|) \quad \text{as } \|\boldsymbol{\delta}\| \to 0. \tag{6}$$

R2) The null space of the mapping $\partial h_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta}$ contains only the zero vector. This assumption is required to ensure that the mapping from $\boldsymbol{\theta}$ to x is non-redundant, in the sense that there does not exist a parametrization of $\mathfrak{X}$ in which the number of degrees of freedom is smaller than K.

3. CRB FOR CONTINUOUS-TIME MEASUREMENTS

When no constraints are imposed (i.e., $\mathfrak{X} = L_2$), a bound for estimating $x(t)$ from measurements contaminated by colored noise was derived in [11]. However, this bound does not hold when the noise $w(t)$ is white. Indeed, in the white noise case, it can be shown that no finite-MSE unbiased estimators exist, unless further information about $x(t)$ is available. For example, the naive estimator $\hat{x}(t) = y(t)$ has an error $\hat{x}(t) - x(t)$ equal to $w(t)$, whose variance is infinite. In our setting, however, we are given the additional information that $x \in \mathfrak{X}$. As we show below, a finite-valued CRB can then be constructed by requiring unbiasedness only within the constraint set $\mathfrak{X}$, as per (5). As we will see below, the CRB increases linearly with the dimension of the manifold $\mathfrak{X}$. Thus, in particular, the CRB is infinite when $\mathfrak{X} = L_2$. The derivation of our constrained CRB involves careful use of measure theoretic concepts. We thus state the results here without their proofs, which will appear in a forthcoming publication [12].

Theorem 1. *Let x be a deterministic function defined by* (1)*, where $\boldsymbol{\theta} \in \Theta$ is an unknown deterministic parameter and Θ is an open subset of $\mathbb{R}^K$. Suppose that Assumptions R1 and R2 are satisfied. Then, the Fisher information measure (FIM) for estimating $\boldsymbol{\theta}$ from $y(t)$ is given by*

$$\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{cont}} = \frac{1}{\sigma^2} \left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\right)^* \left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\right). \tag{7}$$

Furthermore, the MSE per time unit of any unbiased, finite-variance estimator $\hat{x}$ of x is bounded by

$$\frac{\mathrm{MSE}(\hat{x}, x)}{T} \geq \frac{K}{T}\sigma^2 = \rho\sigma^2. \tag{8}$$

We note that $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{cont}}$ can be used to obtain the CRB for estimating $\boldsymbol{\theta}$ from $y(t)$. However, here we are primarily interested in the MSE (8) in estimating $x(t)$. While the FIM for estimating $\boldsymbol{\theta}$ depends on the structure of the set $\mathfrak{X}$, this dependence vanishes when estimating $x(t)$ itself. Indeed, the bound (8) reveals an intuitive geometric interpretation: The set $\mathfrak{X}$ is a K-dimensional manifold in $L_2[0, T]$, i.e., for any point $x \in \mathfrak{X}$, there exists a K-dimensional subspace $\mathcal{U}$ tangent to $\mathfrak{X}$ at x. We refer to $\mathcal{U}$ as the feasible direction subspace [13]: any perturbation of x within the constraint set $\mathfrak{X}$ must be in $\mathcal{U}$. Formally, $\mathcal{U}$ can be defined as the range space of $\partial h_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta}$. If one wishes to use the measurements y to distinguish between x and its local neighborhood, then it suffices to observe the projection of y onto $\mathcal{U}$. Projecting the measurements onto $\mathcal{U}$ removes most of the noise, retaining only K i.i.d. Gaussian components with a variance of σ^2. Thus, an intuitive explanation is apparent for the bound (8).

4. CRB FOR SAMPLED MEASUREMENTS

In this section, we consider the problem of estimating $x(t)$ of (1) from a finite number of samples of the process $y(t)$ given by (2). Specifically, suppose our measurements are given by

$$c_n = \int_0^T y(t) s_n^*(t) dt, \quad n = 1, \ldots, N \tag{9}$$

where $s_n \in L_2[0, T]$ are sampling kernels. The measurements $c_1, \ldots, c_N$ are jointly Gaussian with $\mu_n \triangleq \mathbb{E}\{c_n\}$ and $\mathrm{Cov}(c_i, c_j) = \int_0^T s_i(t) s_j^*(t) dt$. We will be interested in bounds on the error with which $x(t)$ can be estimated from the measurements $\boldsymbol{c} = (c_1, \ldots, c_N)^T$. In particular, we wish to determine conditions under which estimation from the samples $\boldsymbol{c}$ is just as accurate as estimation from the continuous-time noise signal $y(t)$.

A somewhat unusual aspect of this estimation setting is that the choice of the sampling kernels $s_n(t)$ affects not only the measurements obtained, but also the statistics of the noise. One example of the impact of this fact is the following. Suppose we choose a modified set of sampling kernels $\{\tilde{s}_n(t)\}_{n=1}^N$ which are an invertible linear transformation of $\{s_n(t)\}_{n=1}^N$, i.e.,

$$\tilde{s}_n(t) = \sum_{i=1}^N \boldsymbol{A}_{ni} s_i(t) \tag{10}$$

where $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is an invertible matrix. Then, the resulting measurements $\tilde{\boldsymbol{c}}$ are given by $\tilde{\boldsymbol{c}} = \boldsymbol{A}\boldsymbol{c}$, and similarly the original measurements $\boldsymbol{c}$ can be recovered from $\tilde{\boldsymbol{c}}$. It follows that these settings are equivalent in terms of the accuracy with which x can be estimated. Indeed, any estimator $\hat{x}(\boldsymbol{c})$ can equivalently use the measurements $\tilde{\boldsymbol{c}}$ through $\hat{x}(\boldsymbol{A}\tilde{\boldsymbol{c}})$. This is to be distinguished from an arguably more common estimation scenario, in which a change in the measurement system does not affect the noise; in the latter case, an invertible matrix $\boldsymbol{A}$ can have a profound effect on the estimation accuracy, e.g., by changing the SNR.

The conclusion from the discussion above is that the crucial aspect in choosing sampling kernels $\{s_n(t)\}_{n=1}^N$ is the space $\mathcal{S} = \mathrm{span}\{s_1, \ldots, s_N\}$ spanned by these kernels. Once a space $\mathcal{S}$ is selected, the choice of kernels used to span this space is irrelevant from an estimation perspective.

Concerning the choice of the subspace $\mathcal{S}$, suppose first that there exist elements in the range space of $\partial h_{\boldsymbol{\theta}}/\partial\boldsymbol{\theta}$ which are orthogonal to $\mathcal{S}$. This implies that one can perturb x in such a way that the constraint set $\mathfrak{X}$ is not violated, without changing the distribution of the measurements $\boldsymbol{c}$. This situation occurs, for example, when the number of measurements N is smaller than the dimension K of the parametrization of $\mathfrak{X}$. While it may still be possible to reconstruct some of the information concerning x from these measurements, this is an undesirable situation from an estimation point of view. Thus we will assume henceforth that

$$\mathcal{R}\left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}\right) \cap \mathcal{S}^{\perp} = \{\mathbf{0}\}. \tag{11}$$

We then have the following result.

Theorem 2. *Let x be a deterministic function defined by* (1)*, where $\boldsymbol{\theta} \in \Theta$ is an unknown deterministic parameter and Θ is an open subset of $\mathbb{R}^K$. Assume regularity conditions R1 and R2, and let $\hat{x}$ be an unbiased estimator of x from the measurements $\boldsymbol{c} = (c_1, \ldots, c_N)^T$ of* (9)*. Then, the FIM $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}}$ for estimating $\boldsymbol{\theta}$ from $\boldsymbol{c}$ is given by*

$$\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}} = \frac{1}{\sigma^2}\left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}\right)^{*} \boldsymbol{P}_{\mathcal{S}} \left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}\right) \tag{12}$$

where $\boldsymbol{P}_{\mathcal{S}}$ is the orthogonal projector onto the subspace $\mathcal{S}$ spanned by $\{s_n(t)\}_{n=1}^N$. If (11) *holds, then $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}}$ is invertible. In this case, any finite-variance, unbiased estimator $\hat{x}$ for estimating x from $\boldsymbol{c}$ satisfies*

$$\mathrm{MSE}(\hat{x}, x) \geq \mathrm{Tr}\left[\left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}\right)^{*}\left(\frac{\partial h_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}\right)(\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}})^{-1}\right]. \tag{13}$$

It is insightful to compare $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}}$ of (12) with the FIM $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{cont}}$ obtained from continuous-time measurements in Section 3. In particular, if it happens that $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{cont}} = \boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}}$, then by substituting (7) into (13), it is seen that the continuous-time bound of Theorem 1 and the sampled bound of Theorem 2 coincide. Thus, it is possible (at least in terms of the performance bounds) that estimators based on the samples $\boldsymbol{c}$ will suffer no degradation compared with the "ideal" estimator based on the entire set of continuous-time measurements.

The simplest situation in which samples provide all of the information present in the continuous-time signal is the case in which $\mathfrak{X}$ is a K-dimensional subspace of L_2. In this case, $\partial h_{\boldsymbol{\theta}}/\partial\boldsymbol{\theta}$ is a mapping onto the subspace $\mathfrak{X}$. Thus, in light of (12), the optimal choice of a sampling space $\mathcal{S}$ is $\mathfrak{X}$ itself, a choice which requires $N = K$. In this case we obtain $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{cont}} = \boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}}$. Indeed, this choice of a sampling space captures the entire signal x while removing all noise components which are orthogonal to $\mathfrak{X}$. The classical scenario of a bandlimited signal, which was mentioned in Section 1, is the best-known example of such a signal.

A similar situation occurs when $\mathfrak{X}$ is a subset of an M-dimensional subspace $\mathcal{M}$ of L_2 with $M > K$. In this case, $\partial h_{\boldsymbol{\theta}}/\partial\theta_i \in \mathcal{M}$ for all i and all $\boldsymbol{\theta}$, and therefore $\mathcal{R}(\partial h_{\boldsymbol{\theta}}/\partial\boldsymbol{\theta}) \subseteq \mathcal{M}$. Thus, by choosing $N = M$ sampling kernels such that $\mathcal{S} = \mathcal{M}$, we again achieve $\boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{cont}} = \boldsymbol{J}_{\boldsymbol{\theta}}^{\mathrm{samp}}$, demonstrating that all of the information content in x has been captured by the samples. Note, however, that the required sampling rate N/T is potentially much higher than the rate of innovation K/T. One example in which this occurs will be presented in Section 5.

In general, however, the constraint set $\mathfrak{X}$ will not be contained in any finite-dimensional subspace of L_2. In such cases, it will generally not be possible to achieve the performance of the continuous-time bound using any finite number of samples. This situation differs substantially from the noise-free FRI setting, in which typically K samples suffice to reconstruct signals which are parameterized by a vector of length K. It follows that algorithms operating at the rate of innovation are not, in general, optimally suited for estimation when noise is present in the system. This conclusion is compatible with the fact that many FRI algorithms are numerically unstable.

5. EXAMPLE: ESTIMATING MULTI-PULSE SIGNALS

In this section, we focus on a specific application of FRI signals, namely, that of estimating a signal consisting of a number of pulses having unknown positions and amplitudes [4–6]. More precisely, we consider signals of the form

$$x(t) = \sum_{n\in\mathbb{Z}}\sum_{\ell=1}^{L} a_\ell g(t - t_\ell - nT) \tag{14}$$

where $g(t)$ is a known pulse, and $\{a_\ell\}$ and $\{t_\ell\}$ are unknown amplitudes and time delays, respectively. This setting corresponds to numerous channel sounding applications, such as radar, sonar, ultrasound, and cellular channels. By defining the T-periodic function $h(t) = \sum_{n\in\mathbb{Z}} g(t - nT)$, we can write $x(t)$ as

$$x(t) = \sum_{\ell=1}^{L} a_\ell h(t - t_\ell). \tag{15}$$

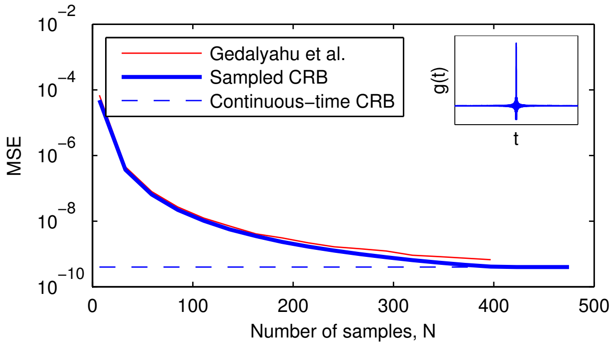
Our goal is now to estimate $x(t)$ from samples of the noisy process $y(t)$ of (2). Since $x(t)$ is T-periodic, it suffices to recover the signal in the region $[0, T]$. In particular, we would like to compare existing algorithms with the CRB in order to determine when these algorithms approach the optimal estimation performance.

Let $\{\tilde{h}_k\}_{k\in\mathbb{Z}}$ be the Fourier series of $h(t)$. The Fourier series of $x(t)$ is then given by

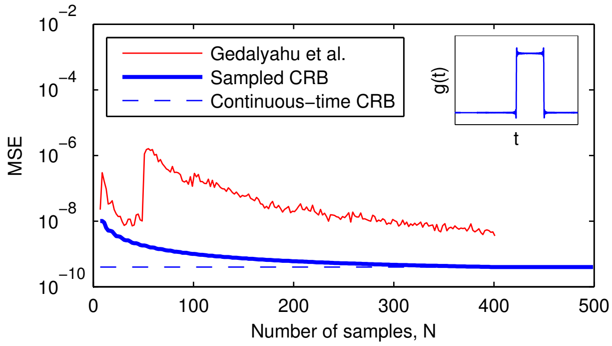
$$\tilde{x}_k \triangleq \frac{1}{T}\int_0^T x(t)e^{-j2\pi kt/T}dt = \tilde{h}_k \sum_{\ell=1}^{L} a_\ell e^{-j2\pi kt_\ell/T}, \quad k \in \mathbb{Z}. \tag{16}$$

Let $\mathcal{K} = \{k \in \mathbb{Z} : \tilde{h}_k \neq 0\}$ denote the indices of the nonzero Fourier coefficients of $h(t)$. Suppose for a moment that $\mathcal{K}$ is finite. It then follows from (16) that $x(t)$ also has a finite number of nonzero Fourier coefficients. Consequently, the set $\mathfrak{X}$ of possible signals $x(t)$ is contained in the subspace $\mathcal{M} = \mathrm{span}\{e^{j2\pi kt/T}\}_{k\in\mathcal{K}}$. Therefore, as explained in Section 4, choosing the $N = |\mathcal{K}|$ sampling kernels $\{s_n(t) = e^{j2\pi nt/T}\}_{n\in\mathcal{K}}$ results in a sampled CRB which is equivalent to the continuous-time bound. This result is compatible with recent work demonstrating successful performance of FRI recovery algorithms using exponentials as sampling kernels [5].

Note, however, that to achieve the performance obtainable from the entire continuous-time signal $y(t)$, the number of samples required is $N = |\mathcal{K}|$, which is potentially much higher than the number of degrees of freedom in the signal $x(t)$. This provides a theoretical explanation of the empirically recognized fact that sampling



(a) The pulse $g(t)$ is a filtered Dirac with 401 Fourier coefficients.



(b) The pulse $g(t)$ is a filtered $\mathrm{rect}(\cdot)$ with 401 Fourier coefficients.

Fig. 1. Comparison of the CRB and the performance of a practical estimator, as a function of the number of samples.

above the rate of innovation improves the performance of FRI techniques in the presence of noise [6], a fact which stands in contrast to the noise-free performance guarantees of many FRI algorithms.

On the other hand, if there exists an infinite number of nonzero coefficients $\tilde{h}_k$, then in general it will not be possible for an algorithm based on a finite number of samples to achieve the performance obtainable from the complete signal $y(t)$. This occurs, for example, whenever $g(t)$ of (14) is time-limited. In such cases, any increase in the sampling rate will potentially continue to reduce the CRB, although the sampled CRB will converge to the asymptotic value of $\rho\sigma^2$ in the limit as the sampling rate increases.

These effects are demonstrated in Fig. 1, which documents several experiments comparing the CRB with the time-delay estimation technique of Gedalyahu et al. [5], whose performance is identical in the present setting to the method of Vetterli et al. [1]. In these experiments, a signal containing $L = 2$ pulses with random delays and amplitudes was constructed. The pulse $h(t)$ consisted of $|\mathcal{K}| = 401$ nonzero Fourier coefficients at positions $\mathcal{K} = \{-200, \ldots, 200\}$. The CRB is plotted as a function of the number of samples N, where the sampling kernels are given by $s_n(t) = e^{j2\pi nt/T}$ with $n \in \{-\lfloor N/2 \rfloor, \ldots, \lfloor N/2 \rfloor\}$.

In Fig. 1(a), we chose $\tilde{h}_k = 1$ for $-200 \le k \le 200$ and $\tilde{h}_k = 0$ elsewhere; these are the low-frequency components of a Dirac delta function. The noise standard deviation was $\sigma = 10^{-5}$. As expected, the sampled CRB achieves the continuous-time bound $\rho\sigma^2$ when $N \ge |\mathcal{K}|$. However, the CRB obtained at low sampling rates is higher by several orders of magnitude than the continuous-time limit. This indicates that the maxim of FRI theory, whereby sampling at the rate of innovation suffices for reconstruction, may not always hold in the presence of mild levels of noise. Observe that in this scenario, existing algorithms come very close to the CRB. Thus, the previously observed improvements achieved by oversampling are a result of fundamental limitations of low-rate sampling, rather than drawbacks of the specific technique used.

Fig. 1(b) plots the results of a similar experiment, in which $\tilde{h}_k$ equals the 401 lowest-frequency Fourier coefficients of a rectangular pulse. In this case, some of the higher Fourier coefficients have small magnitude, and result in low-SNR measurements. Adding these measurements can actually be detrimental to the performance of the examined estimation algorithm. Yet information is clearly present in these high-frequency samples, as indicated by the continual decrease of the CRB with increasing N. Thus, our analysis indicates that improved estimation techniques should be achievable in this case, in particular by careful utilization of low-SNR measurements.

6. REFERENCES

[1] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[2] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[3] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 1741–1757, 2007.

[4] K. Gedalyahu and Y. C. Eldar, "Time delay estimation from low rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3017–3031, Jun. 2010.

[5] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Trans. Signal Process.*, submitted. [Online]. Available: http://arxiv.org/abs/1004.5070

[6] R. Tur, Y. C. Eldar, and Z. Friedman, "Low rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, submitted. [Online]. Available: http://arxiv.org/abs/1003.2822

[7] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, 1989.

[8] J. Berent, P. L. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with finite rate of innovation methods," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 613–625, Feb. 2010.

[9] H.-H. Kuo, "White noise theory," in *Handbook of Stochastic Analysis and Applications*, D. Kannan and V. Lakshmikantham, Eds. New York: Marcel Dekker, 2002, ch. 3.

[10] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*. New York: Springer, 1981.

[11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, vol. 1.

[12] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, "Performance bounds and design criteria for the estimation of finite rate of innovation signals from noisy measurements," *IEEE Trans. Signal Process.*, 2010, in preparation.

[13] Z. Ben-Haim and Y. C. Eldar, "The Cramér–Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.