SAMPLING FRI SIGNALS WITH THE SOS KERNEL: BOUNDS AND OPTIMAL KERNEL

Stéphanie Bernhardt, Rémy Boyer*, Sylvie Marcos*, Yonina C. Eldar† and Pascal Larzabal‡*

* Laboratoire des Signaux et Systèmes (L2S)
Univ. Paris Sud-CNRS-CentraleSupélec
Gif-Sur-Yvette, France

† Technion,
Department of EE
Haifa, Israel

‡ SATIE
ENS-Cachan
Cachan, France

ABSTRACT

Recently it has been shown that using appropriate sampling kernel, finite rate of innovation signals can be perfectly reconstructed even tough they are non-bandlimited. In the presence of noise, reconstruction is achieved by an estimation procedure of all the parameters of the incoming signal. In this paper we consider the estimation of a finite stream of pulses using the Sum of Sincs (SoS) kernel. We derive the Cramér Rao Bound (BCRB) relative to the estimated parameters. The SoS kernel is used since it is configurable by a vector of weights: we propose a family of kernels which maximizes the Bayesian Fisher Information (BIM) i.e. the total amount of information about each of the parameter in the measures. The advantage of the proposed family is that it can be user-adjusted to favor one specific parameter. The variety of the resulting kernel goes from a perfect sinusoid to the Dirichlet kernel.

1. INTRODUCTION

Finite streams of filtered pulses are an important class of signals since they appear in many applications including bio-imaging, radar, and spread-spectrum communication. The pulse shape being known, such signals are totally described by the knowledge of the amplitudes and the delays of the pulses. Therefore, even if those signals are not bandlimited, they can be sampled thanks to their inherent sparsity, in the sense that they are fully described by only a small number of parameters per unit of time. This is the key idea of the finite rate of innovation (FRI) framework introduced in [1]. FRI signals can be sampled using a sampling kernel that must verify some specific properties [2]. In [3], a family of adjustable kernels following those properties is proposed: the Sum of Sincs (SoS) kernels. In this paper we focus on those kernels since they allow a large variety of kernel shapes while observing the needed properties. The kernel shape is adjusted using simply a vector of weights which allows to choose the best weights to optimize a given objective.

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM# (Grant agreement no. 318306) and by the following projects: ANR MAGELLAN, MI-CNRS DISPLAY and TITAN.

In this paper we aim to find the best kernel maximizing the estimation performance. To characterize this performance, it is standard to use the Cramér-Rao Bound (CRB) which provides a lower bound on the Mean Squared Error (MSE) of any unbiased estimator [4]. The CRB has already been used to optimize FRI sampling in [5] using exponential reproducing kernels. In [6] we derived the deterministic CRB for a stream of Dirac pulse sampling by an arbitrary kernel. However, the CRB is a function of the unknown parameter vector meaning that the optimal kernel is unknown parameter dependent. To obtain a kernel which is optimal on average, we consider the stochastic framework where the parameters are random variables following a given distribution. Therefore we derive the BCRB (as defined in [7]) for the estimation of a stream of pulses using specifically the SoS kernel, which is already recognized as one of the best [3]. Consequently, it is on that family that the optimum is looked for.

Using the Bayesian Fisher Information Matrix (BIM) we derive the optimum kernel which minimizes the Bayesian CRB (BCRB). We choose an objective function that allows to decide the relative importance of the delays and the amplitudes, which can vary depending on the applications. As an example for a radar application, the main objective is to recognize the delays since it allows to localize the source, in this case it is of interest to favor the delay estimation. Another open parameter allows the user to choose the robustness of the resulting kernel towards the noise. The resulting family of optimal SoS kernels can take the form from a simple sinusoid to the standard Dirichlet function. The following notations are used throughout the paper: $\mathbf{e}_m$ is the vector with a 1 at the m^e coefficient, and zeros otherwise, $\mathbb{1}_L$ is a $L \times L$ matrix full of ones. $\mathrm{tr}(.)$ represents the trace of a matrix. The real inner product is defined as $\langle g(t), x(t) \rangle = \int_{-\infty}^{\infty} g(t)x(t)dt$.

2. PERFORMANCE BOUNDS

2.1. Sampling model

The signal to be sampled is a continuous-time signal with a finite number of L weighted and delayed pulses:

$$x(t) = \sum_{l=0}^{L-1} a_l h(t - \tau_l) \tag{1}$$

where $\boldsymbol{\tau} = [\tau_0, \ldots, \tau_{L-1}]^T \in \mathbb{R}^L +$ and $\mathbf{a} = [a_0, \ldots, a_{L-1}]^T \in \mathbb{R}^L$ with are the unknown parameters called the time-delays and the amplitudes for each pulse $h(t)$. This type of signal is called Finite Rate of Innovation signal since it can be totally determined with only a finite number of parameter within a given frame of time, here $2L$.

We consider the problem of estimating the unknown vector of parameters $\boldsymbol{\theta} = [\boldsymbol{\tau}^T \ \mathbf{a}^T]^T$ based on a finite number of samples N, uniformly sampled with a sampling interval T_S. We consider the Bayesian framework where a_l are i.i.d. random variables following a centered distribution $p(a_l)$ with variance σ_a^2, and τ_l are i.i.d random variables following a generalized normal (GN) distribution with zero location parameter, scale parameter $\alpha > 0$ and shape parameter $\beta > 0$ denoted $\tau_l \sim \mathcal{GN}(0, \alpha, \beta)$, we note the variance of $p(\tau_l)$ σ_τ^2. The GN distribution encompasses the Laplacian ($\beta = 1$), Gaussian ($\beta = 2$) and uniform ($\beta \to \infty$) distributions [8]. The N samples are a filtered/smoothed version of $x(t)$ with a kernel $g(t)$ according to

$$y_n = \langle g(t - nT_S), x(t)\rangle + w_n = c_n + w_n \tag{2}$$

where w_n is the real white Gaussian digital noise process having mean zero and variance σ^2.

Using the SoS sampling kernel, defined by [3]

$$g(t) = \mathrm{rect}\left(\frac{t}{NT_S}\right) \sum_{k=-K/2}^{K/2} b_k e^{\frac{j2\pi kt}{NT_S}}, \tag{3}$$

the vector of samples $\mathbf{c} = [c_1, \cdots, c_N]$ can be written

$$\mathbf{c} = \mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{x}, \tag{4}$$

where $\mathbf{t}_S = \{nT_S : 0 \le n \le N-1\}$ correspond to the vector of the time samples, $\mathbf{V}(-\mathbf{t}_S)$ is a $N \times (K+1)$ matrix whose (m,k)-th element is $e^{-j2\pi kn/N}$ and $\mathbf{B}$ is the $K+1 \times K+1$ diagonal matrix having $\{b_k\}$ on its diagonal. We consider kernels for which $b_k = b^*_{-k}$ to ensure that the filter is real valued. $\mathbf{x}$ is a vector composed with the $K+1$ continuous time Fourier transform coefficients of $x(t)$. They can be written

$$\mathbf{x} = \mathbf{H}\mathbf{V}(\boldsymbol{\tau})\mathbf{a}, \tag{5}$$

where $\mathbf{V}(\boldsymbol{\tau})$ is the $(K+1) \times L$ matrix whose (k,l)-th entry is $e^{-j2\pi k\tau_l/(NT_S)}$, with $\forall l, \tau_l \le NT_S$. $\mathbf{H}$ is the $(K+1) \times (K+1)$ diagonal matrix whose k-th entry is $h_k = \frac{1}{NT_S} H\left(\frac{2\pi k}{NT_S}\right)$ where $H(w)$ is the continuous time Fourier transform of the pulse $h(t)$.

2.2. Bayesian Fisher Information Matrix

The measurement model (2) can be written in vector form with $\mathbf{y} = [y_1, \cdots, y_N]$ and $\mathbf{w} = [w_1, \cdots, w_N]$:

$$\mathbf{y} = \mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{H}\mathbf{V}(\boldsymbol{\tau})\mathbf{a} + \mathbf{w}. \tag{6}$$

The BIM is defined as the expectation over the parameters of interest of the Fisher Information Matrix. To derive de CRB the $\mathrm{rect}(\cdot)$ function is approximated by the GN with $\beta \to \infty$. The samples follow a normal distribution $\mathbf{y}|\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I}_N)$ where $\boldsymbol{\mu} = \mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{H}\mathbf{V}(\boldsymbol{\tau})\mathbf{a}$. Using the Slepian-Bang formula, the BIM is [4]

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} E_{\boldsymbol{\theta}}\left\{\frac{\partial \boldsymbol{\mu}^T}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}\right\} - E_{\boldsymbol{\theta}}\left\{\frac{\partial^2 \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right\}. \tag{7}$$

At low noise variance, the second term is negligible [7]. Using the fact that the variables are independent, and that $p(a_l)$ is centered we can separate the Fisher information relative to the amplitude and the Fisher information relative to the delays (see the appendix for the derivations). Finally, we consider that $p(\tau_l)$ tends to a uniform law ($\beta \to \infty$), since it is a natural choice of distribution. The asymptotic BIM is:

$$\begin{aligned}
\mathbf{J}_{\boldsymbol{\tau}} &\approx \left(\frac{\sigma_a^2 4\pi^2}{\sigma^2 NT_S^2} \sum_{k=-K/2}^{K/2} k^2 h_k^2 b_k^2\right) \mathbf{I}_L \\
&= \left(\frac{4\sigma_a^2 \pi^2}{\sigma^2 NT_S^2} \mathbf{b}^T \mathbf{D} \mathbf{b}\right) \mathbf{I}_L
\end{aligned} \tag{8}$$

$$\begin{aligned}
\mathbf{J}_{\mathbf{a}} &\approx \left(\frac{N^2}{\sigma^2} \sum_{k=-K/2}^{K/2} h_k^2 b_k^2\right) \mathbf{I}_L + \frac{N}{\sigma^2}\left(h_0^2 b_0^2 \left(\mathbb{1}_L - \mathbf{I}_L\right)\right) \\
&= \frac{N^2}{\sigma^2}\left(\left(\mathbf{b}^T \bar{\mathbf{D}} \mathbf{b}\right) \mathbf{I}_L + h_0^2 b_0^2 \left(\mathbb{1}_L - \mathbf{I}_L\right)\right)
\end{aligned} \tag{9}$$

where $\mathbf{D}$ and $\bar{\mathbf{D}}$ are diagonal matrices with $[\mathbf{D}]_{kk} = k^2 h_k^2$ and $[\bar{\mathbf{D}}]_{kk} = h_k^2$.

2.3. Bayesian Cramér-Rao Bound

The BCRB is the trace of the inverse of the BIM [7]. Using the Sherman-Morison inversion formula, the BCRBs at high SNR are:

$$\mathrm{BCRB}(\tau_l) \approx \frac{\sigma^2 NT_S^2}{4\pi^2 \sigma_a^2} \frac{1}{\sum_{k=-K/2}^{K/2} k^2 h_k^2 b_k^2} \tag{10}$$

$$\mathrm{BCRB}(a_l) \approx \frac{\sigma^2}{N\left(\sum_{\substack{k=-K/2 \\ k\neq 0}}^{K/2} h_k^2 b_k^2\right)} \cdot \left(\frac{h_0^2 b_0^2 + (K-1)\sum_{\substack{k=-K/2 \\ k\neq 0}}^{K/2} h_k^2 b_k^2}{h_0^2 b_0^2 + K \sum_{\substack{k=-K/2 \\ k\neq 0}}^{K/2} h_k^2 b_k^2}\right). \tag{11}$$

3. OPTIMAL KERNEL

3.1. Objective function

The sum of sinc kernel is a particularly versatile kernel compared to the often used Gaussian and Dirichlet kernels since it can be weighted through the coefficient vector

$\mathbf{b} = \left[b_{-K/2}, \cdots, b_0, \cdots b_{K/2}\right]^T$. It is therefore possible to choose the best kernel depending on a specific pulse shape $h(t)$, and for a specific objective function.

We want to optimize the kernel in a way to be able to favor the delays or the amplitudes. To be able to compare the BCRBs, we use the normalized BCRBs, which are defined following the definition of the normalized mean square error [7] by $\mathrm{NBCRB}(a_l) = \frac{\mathrm{BCRB}(a_l)}{\sigma_a^2}$ for the amplitudes and by $\mathrm{NBCRB}(\tau_l) = \frac{\mathrm{BCRB}(\tau_l)}{\sigma_\tau^2}$ for the delays.

We cannot easily compute the optimal kernel which minimizes the total BCRB since the trace of the BCRB matrix is not convex. Therefore, we aim to maximize the total amount of information about each of the parameter in the measures. This information is measured by the trace of $\mathbf{J}_{\boldsymbol{\tau}}$ and the trace of $\mathbf{J}_{\mathbf{a}}$ for the delays and the amplitudes parameter respectively [9]. We remark in the expressions (8) and (9) that all the diagonal elements of the BIM matrices are equal and note J_τ and J_a these constant diagonal terms. The maximization of the trace of $\mathbf{J}_{\boldsymbol{\tau}}$ ($\mathbf{J}_{\mathbf{a}}$ respectively) is equivalent to maximizing J_τ (J_a respectively). By inversion of the NBCRB, the normalized BIM on each parameter (we note it J_τ^N and J_a^N) is the BIM multiplied by the variance of the considered parameter. A user-defined tuning parameter λ is introduced in the objective function to define the relative importance of each parameter. Using the normalized BIM we get that for $\lambda = 0.5$ both parameters have the same relative importance.

The final objective function is:

$$\begin{aligned} f_\lambda(\mathbf{b}) &= \lambda J_\tau^N + (1-\lambda) J_a^N & (12)\\ &= \lambda \sigma_\tau^2 J_\tau + (1-\lambda)\sigma_a^2 J_a & (13)\\ &= \lambda \frac{\sigma_\tau^2 \sigma_a^2 4\pi^2}{\sigma^2 N T_S^2} \mathbf{b}^T \mathbf{D} \mathbf{b} + (1-\lambda)\frac{N^2 \sigma_a^2}{\sigma^2} \mathbf{b}^T \bar{\mathbf{D}} \mathbf{b} & (14)\\ &= \mathbf{b}^T \boldsymbol{\Delta}_\lambda \mathbf{b}. & (15) \end{aligned}$$

Note that the matrices $\mathbf{D}$ and $\bar{\mathbf{D}}$ are positive definite, therefore this criterion is convex.

We have

$$\boldsymbol{\Delta}_\lambda = \mathrm{SNR}\left(\lambda \frac{\sigma_\tau^2 4\pi^2}{N T_S^2}\mathbf{D} + (1-\lambda) N^2 \bar{\mathbf{D}}\right) \quad (16)$$

with $\mathrm{SNR} = \sigma_a^2/\sigma^2$.

A constraint inherent to the problem of sampling finite rate of innovation signals is that the coefficient b_k are different from zero. Since in the objective function the sign of the b_ks has no influence, this constraint is equivalent to the constraint that $b_k \geq \epsilon > 0$, where ϵ is a user-defined parameter. Increasing the filter's amplification will always reduce the BCRB, therefore the energy is normalized using the constraint that $||\mathbf{b}||^2 = K+1$. The resulting optimization problem is

$$\max_{\mathbf{b}} \mathbf{b}^T \boldsymbol{\Delta}_\lambda \mathbf{b} \text{ s.t. } \begin{cases} ||\mathbf{b}||^2 = K+1 \\ \forall k \in [-K/2 \cdots K/2] : b_k \geq \epsilon \end{cases}. \quad (17)$$

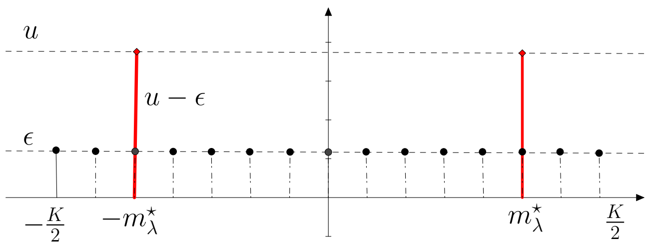

Fig. 1. Resulting optimal kernel

3.2. Optimization results

We first consider the following simplified problem, for $\epsilon = 0$:

$$\max_{\mathbf{b}} \mathbf{b}^T \boldsymbol{\Delta}_\lambda \mathbf{b} \text{ s.t. } ||\mathbf{b}||^2 = K+1. \quad (18)$$

The Lagrangian function is

$$\forall \lambda \in [0,1], \mathcal{L}_\lambda(\mathbf{b}, \alpha) = \mathbf{b}^T \boldsymbol{\Delta}_\lambda \mathbf{b} - \alpha(||\mathbf{b}||^2 - (K+1)). \quad (19)$$

The optimal $\mathbf{b}^\star$ is given by differentiating $\mathcal{L}(\mathbf{b}, \alpha)$ wrt. $\mathbf{b}$ and the optimal $\mathbf{b}_\lambda^\star$ is the eigen-vector associated with the largest eigen-value $\beta_{m_\lambda^\star}(\boldsymbol{\Delta}_\lambda)$ of matrix $\boldsymbol{\Delta}_\lambda$.

Note that the optimal $\mathbf{b}_\lambda^\star$ is a function of parameter λ. In addition, as $\boldsymbol{\Delta}_\lambda$ is diagonal and $[\boldsymbol{\Delta}_\lambda]_{-m} = [\boldsymbol{\Delta}_\lambda]_m$, the optimal $\mathbf{b}_\lambda^\star$ is given by

$$\mathbf{b}_\lambda^\star = \sqrt{\frac{K+1}{2}}(\mathbf{e}_{m_\lambda^\star} + \mathbf{e}_{-m_\lambda^\star}) \quad (20)$$

since the largest eigen-value of $\boldsymbol{\Delta}_\lambda$ is of multiplicity 2. Therefore the constraint that $b_{-m} = b_m$ is respected. Note that $||\mathbf{b}_\lambda^\star||^2 = K+1$.

Remark 1 *In the case of a Dirac pulse, the frequency coefficients $h_k = h$ are all equal, therefore*

$$J_a = \mathrm{SNR} N^2 h^2 \sum_{k=-K/2}^{K/2} b_k^2$$

does not depend on the chosen kernel $\mathbf{b}$*, since $\sum_{k=-K/2}^{K/2} b_k^2$ is fixed by constraint. Thus, when $\lambda = 0$ i.e. when only the amplitudes are considered, all the kernels are solutions of the optimization problem. Among those solutions, the Dirichlet kernel is of particular interest [3].*

When $\epsilon > 0$, we make the change of variable $\beta_k = b_k^2$ and solve the equivalent problem using CVX, a package for specifying and solving convex programs [10] [11]. We observe in the simulations that the resulting kernel has the same form than for $\epsilon = 0$ but with a floor at the level ϵ (see Fig. 1). The peak is at the level $u = \sqrt{\frac{(1-\epsilon^2)(K+1)}{2}}$ when $\epsilon \neq 1$ such that $\sum_{-K/2}^{K/2} \beta_k = K+1$.

The resulting optimal kernel in time domain is:

$$g^\star(t) = 2(u-\epsilon)\cos\left(\frac{2\pi m^\star t}{N T_S}\right) + \epsilon D_{K/2}\left(\frac{2\pi t}{N T_S}\right) \quad (21)$$

where $D_p(x) = \sum_{k=-p}^{p} e^{jkx}$ is the Dirichlet kernel. If $\epsilon = 0$, this kernel is a simple sinusoidal signal, whereas if $\epsilon = 1$ we obtain that $u = 1$ and the optimum kernel corresponds to the often used Dirichlet kernel. An example of kernel is showed Fig. 2: it is composed of a sum of a sinusoid and a Dirichlet function centered at the origin.

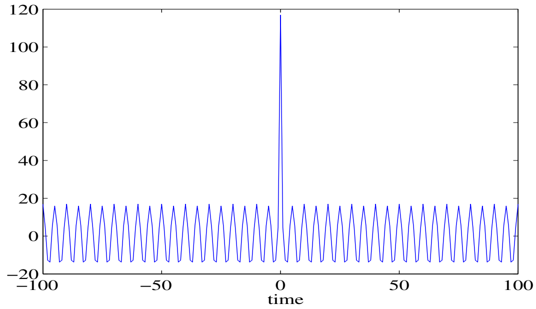

Fig. 2. Optimal kernel in time domain, for $N = K = 200$ and $\epsilon = 0.5$

Remark 2 *The optimization of the BCRB or the BIM gives the best possible kernel theoretically. This kernel is obtained for $\epsilon \to 0$, however using this kernel makes the estimation unfeasible. Choosing appropriately ϵ is important depending on the chosen estimation procedure. The higher ϵ, the higher is the peak which allows to distinguish the delays of the source. If there is no noise, this peak can be infinitely small, and a perfect estimation procedure will recognize the delays. However, when noise is added, it is important to have a higher peak to mitigate the effect of the noise. Therefore the choice of ϵ will depend principally on the signal to noise ratio.*

4. SIMULATION

For all the simulation results we consider the following parameters: $N = K = 200$, $Ts = 1$, $\epsilon = 0.5$ and SNR $= 30$ dB. The BIM is linear with respect to the SNR, therefore we do not plot the results depending on the SNR. We compare the optimized kernel with the standard Dirichlet kernel, which has all b_ks equal to 1.

The optimization results for the case where $h(t)$ is the Dirac delta function are shown Fig. 3. Since h_k are all equal, the normalized BCRB on the amplitude is constant with respect to the parameter λ and equal for all the kernels. The optimal kernel choice is therefore driven by the delays and the result of the optimization is $m^\star = K$ for all λ except $\lambda = 0$ where only the amplitudes are considered. Since the performance are constant for all λ instead $\lambda = 0$, we wrote in Fig. 3 the BCRB on the amplitude and the delays of the Dirichlet kernel, and the optimized kernel for $\lambda = 0$ and $\lambda > 0$. When $\lambda = 0$ the performance are approximately the same as the Dirichlet kernel's for both parameters but for $\lambda > 0$ the BCRB on the delays and consequently the total BCRB are improved.

	Amplitude	Delays
Dirichlet kernel	4.975×10^{-6}	8.984×10^{-8}
Optimized kernel $\lambda = 0$	4.975×10^{-6}	8.950×10^{-8}
Optimized kernel $\lambda > 0$	4.975×10^{-6}	3.626×10^{-8}

Fig. 3. Normalized BCRBs for the Dirac pulse

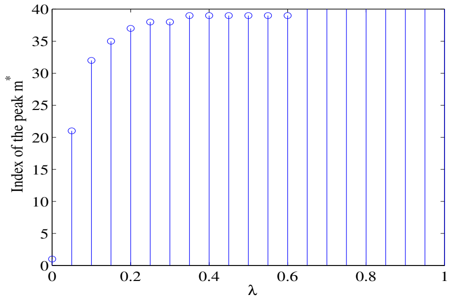

Fig. 4. Index of the peak $m^\star_\lambda$ for a Gaussian pulse.

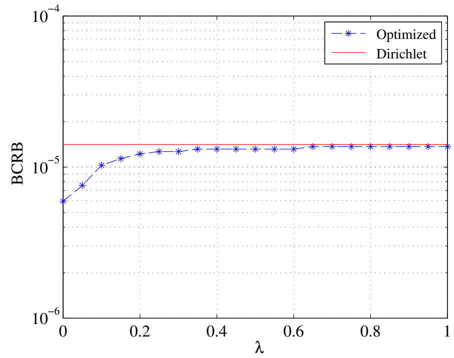

Fig. 5. Normalized BCRB on the amplitudes for a gaussian pulse.

In a second time, we consider the optimization of the kernel in the case where $h(t)$ is a Gaussian pulse. Since the pulse frequency coefficients are not constant, the choice of the peak depends on λ: the higher λ the more important the delays compared to the amplitudes. The peak is near zero when λ is small, since the maximum of the Gaussian is at the origin. When λ increases, the peak is chosen as the trade-off between h_k^2 and $h_k^2 k^2$, therefore the index increases while λ increases (see Fig. 4). The BCRB on the amplitude is minimized when λ is small, and while λ increases and the importance of the amplitude parameter decreases, the BCRB on the amplitude increases to join the performance of the Dirichlet kernel (see Fig. 5). On the other hand, the BCRB on the delays is bad when $\lambda = 0$, since the delays are not considered, the performance are worse than the Dirichlet kernel. When λ increases the BCRB on the delays increases to reach a constant level for $\lambda > 0.1$, which surpasses the Dirichlet kernel's performance (see Fig. 6).

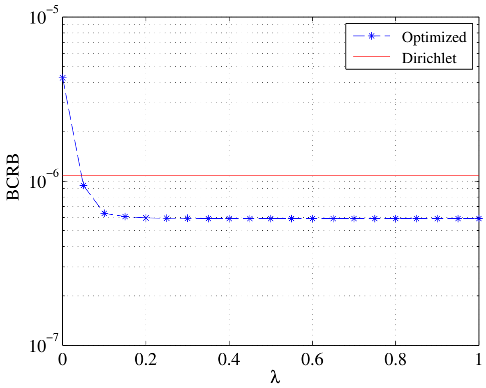

Fig. 6. Normalized BCRB on the delays for a gaussian pulse.

5. CONCLUSION

In this work we derived analytical expressions of the BCRB for the estimation of the parameters used for the sampling of finite rate of innovation signals. Based on the BIM relative to these parameters, we propose an optimization framework which chooses the best SoS kernel with respect to the Fisher information. The user can choose the importance of the time domain peak, which adjusts the robustness of the estimation regarding the noise level and the relative importance of the delay and the amplitude parameters. We show by simulation that the optimized kernel gives better results than the standard Dirichlet kernel, for both Dirac and Gaussian stream of pulses.

6. APPENDIX

The derivatives of μ are $\frac{\partial \boldsymbol{\mu}}{\partial \tau_l} = a_l \mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{H}\dot{\mathbf{g}}(\tau_l)$ and $\frac{\partial \boldsymbol{\mu}}{\partial a_l} = \mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{H}\mathbf{g}(\tau_l)$ where $\mathbf{g}(\tau_l) = e^{\frac{-j2\pi k \tau_l}{NT_S}}$ for $-K/2 \leq k \leq K/2$ and $\dot{\mathbf{g}}(\tau_l) = \frac{\partial \mathbf{g}(\tau_l)}{\partial \tau_l}$.

$[\mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{H}]^H \mathbf{V}(-\mathbf{t}_S)\mathbf{B}\mathbf{H}$ is a diagonal matrix whose k^e element is $Nb_k^2 h_k^2$.

Since the distribution $p(a_l)$ is zero-mean, taking the expectation of the FIM annuls the cross-terms. Furthermore for $l \neq l' : [\mathbf{J}_\tau]_{ll'} = 0$. The remaining non-zero terms in the matrix are:

$$[\mathbf{J}_\tau]_{ll} = N\sigma_a^2 \sum_{k=-K/2}^{K/2} k^2 b_k^2 h_k^2,$$

$$[\mathbf{J}_\mathbf{a}]_{ll} = N \sum_{k=-K/2}^{K/2} b_k^2 h_k^2$$

$$[\mathbf{J}_\mathbf{a}]_{ll'} = N \sum_{k=-K/2}^{K/2} b_k^2 h_k^2 E_\tau \left\{ e^{\frac{j2\pi k \tau}{NT_S}} \right\} \cdot E_\tau \left\{ e^{\frac{-j2\pi k \tau}{NT_S}} \right\}.$$

We consider the limit case of the GN distribution when $\beta \rightarrow \infty$ and approximate it by the uniform distribution of τ between $-N/2$ and $N/2$. The last expression becomes:

$$E_{\boldsymbol{\theta}} \left\{ \left[\frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[\frac{\partial \boldsymbol{\mu}}{\partial a_l'} \right] \right\} = N b_0^2 h_0^2.$$

REFERENCES

[1] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *Signal Processing, IEEE Transactions on*, vol. 50, no. 6, pp. 1417–1428, 2002.

[2] P.L. Dragotti, M. Vetterli, and T. Blu, “Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang fix,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[3] R. Tur, Y. C. Eldar, and Z. Friedman, “Innovation rate sampling of pulse streams with application to ultrasound imaging,” *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1827–1842, 2011.

[4] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*, Fundamentals of Statistical Signal Processing. Prentice-Hall PTR, 1993.

[5] J.A. Uriguen, T. Blu, and P.L. Dragotti, “Fri sampling with arbitrary kernels,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 21, pp. 5310–5323, Nov 2013.

[6] S. Bernhardt, R. Boyer, S. Marcos, Y. Eldar, and P. Larzabal, “Cramer-Rao Bound for Finite Streams of an Arbitrary Number of Pulses,” in *EUSIPCO'14*, Lisbonne, Portugal, Sept. 2014, p. nc.

[7] H. L. Van Trees, *Detection, estimation, and modulation theory*, John Wiley & Sons, 2004.

[8] S. Nadarajah, “A generalized normal distribution,” *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.

[9] F. S. Alshunnar, M. Z. Raqab, and D. Kundu, “On the comparison of the fisher information of the log-normal and generalized rayleigh distributions,” *Journal of Applied Statistics*, vol. 37, no. 3, pp. 391–404, 2010.

[10] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” `http://cvxr.com/cvx`, Mar. 2014.

[11] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited, 2008.