Level-Crossings Sampling for Signal Recovery with Slepian Derivatives

Maciej Ordowski*, Mirosław Pawlak*†, Dominik Rzepka*, Marek Miśkowicz*
*Dept. of Measurement and Electronics, AGH University of Science and Technology, Krakow, Poland
†Dept. of Electrical and Computer Engineering, University of Manitoba, Canada, R3T 5V6 Winnipeg, Canada
ordowski@agh.edu.pl, miroslaw.pawlak@umanitoba.ca, drzepka@agh.edu.pl, miskow@agh.edu.pl

***Abstract*—We study the problem of the signal recovery for the class of stationary bandlimited Gaussian processes when the measured information is given in terms of signal level crossings. Level-crossing data collection is based on signal amplitude sampling with error-free timing information and it is a preferable sampling paradigm due to its low-power consumption. In this paper, we develop a reconstruction method that incorporates not only level-crossings samples but also the additional information on the signal derivative. The signal derivative, however, is not observed and we utilize the Slepian model for the process behavior after crossings the given level. The model consists of the linear regression function with the universal Rayleigh component that characterizes the signal slope and a non-stationary Gaussian noise process. The existing methods for signal recovery with derivatives assume that both the signal and its derivatives are observed and they form the observation set. Our approach yields the improved reconstruction accuracy and provides a new strategy for a general class of event-driven data analysis algorithms.**

I. Introduction

The signal recovery and detection from samples is the fundamental problem in signal analysis and communication. When a signal has a known form and specified spectral properties like bandlimitedness its reconstruction from the time-domain regularly spaced samples is the classical topic in signal analysis [1], [2]. When the signal has an unknown form, it is appropriate to consider the signal as a sample function of a random process. When the signal statistics (like correlation structure) are known, this knowledge can often be used to design suitable reconstruction and detection methods from irregularly spaced samples. In this paper we assume that the underlying signals are modelled as stationary band-limited Gaussian processes. The problem of recovery of stochastic signals from time-domain non-uniform samples has received some attention in the literature [3], [4], [5], [6]. An alternative approach for signal acquisition is relied on level-crossing generated samples [7], [8], [9], [10], [11] that provide a low-power consumption strategy for signal sampling. Moreover, the level-crossing samples supply the fine local characterization of the underlying signal with a fewer number of irregularly spaced samples than it is required by the classical time-domain Shannon sampling theorem. The use of stochastic modeling of various signal and system models has been remarkable effective in data science and dynamic systems [4], [5]. The challenging problem within this framework is to incorporate a priori knowledge about the signal shape [6]. The important information about the signal form is provided by its derivative and its incorporation in the recovery algorithm can increase the estimation accuracy and to gain the further knowledge about properties like monotonicity and the signal sign. The problem of signal reconstruction which allows a combination of signal samples and samples of its derivative has been studied in the machine learning and statistical literature [12], [13]. In this paper, we examine the reconstruction problem that incorporates not only signal level-crossings samples but also the additional information on the signal derivative. Nevertheless, contrary to the aforementioned publications the signal derivative is not observed and we approximate the signal slope information (derivative) at the given crossing point by the so-called Slepian model [8], [14], [15], [16], [17]. The Slepian model is the universal signal slope approximation at level-crossings. The model consists of two independent terms where one depends on the level value and the other on the universal Rayleigh distributed random component that characterizes the signal slope. We show that the use of the Slepian model improves the accuracy of signal recovery and provides a promising tool for incorporating other local signal properties like local extrema. The remainder of the paper is organized as follows. Section 2 introduces the basic concepts of the optimal mean-squared-error (**MSE**) stochastic signal recovery theory. This includes both the cases of observed signal samples and its derivatives. Section 3 examines the signal estimation from level-crossings samples. In particular, the Slepian model is introduced and its use in the reconstruction process is examined. Section 4 gives an illustrative numerical example. The detailed proofs of our results and extensive simulations studies will be presented elsewhere.

M. Pawlak, D. Rzepka and M. Miśkowicz were supported by the Polish National Center of Science under Grant DEC-2017/27/B/ST7/03082.

II. Optimal Signal Recovery with Derivative Information

We represent the uncertainty about the signal form by the stationary Gaussian stochastic process $\{X(t), 0 \leq t \leq T\}$ with (without loss of generality) zero mean and covariance function $R(\tau)$, where $R(0) = \sigma^2$ is the process variance. The signal is assumed to be bandlimited with the bandwidth ω_0. We assume that the process $X(t)$ has smooth trajectories such that its derivative $X^{(1)}(t)$ is well defined. It is known [14] that $X^{(1)}(t)$ is also the stationary Gaussian process with the covariance function $-R^{(2)}(\tau)$, where $\mu_2 = -R^{(2)}(0)$ is the variance of $X^{(1)}(t)$ and it is often named as the second spectral moment of $X(t)$. It worth mentioning that for bandlimited processes μ_2 is proportional to the signal bandwidth ω_0. The measured information about the signal $X(t)$ is represented by the vector $\boldsymbol{X}_N = [X(t_1), \ldots, X(t_N)]^T$ of samples of $X(t)$ at the given time instants $\{t_k\}$. In the signal recovery problem one wishes to predict $X(t)$ based on $\boldsymbol{X}_N$. The stochastic **MSE** signal estimation theory [18], [3] suggests that one should consider the conditional process $Z_N(t) = X(t)|\boldsymbol{X}_N$. In fact, the mean value of $Z_N(t)$ defines the optimal **MSE** predictor $\widehat{X}(t)$ with the minimal MSE given by $\sigma_N^2 = \boldsymbol{E}[\text{var}[Z_N(t)]]$. Under the Gaussian model the optimal **MSE** estimate $\widehat{X}(t)$ is a linear function of $\boldsymbol{X}_N$ and the conditional process $Z_N(t)$ is the nonstationary Gaussian process. In fact, using standard formulas for the conditional multivariate Gaussian distribution we have

$$\widehat{X}(t) = (\boldsymbol{R}^{-1}\boldsymbol{b}(t))^T \boldsymbol{X}_N, \qquad (1)$$

where $\boldsymbol{R} = \boldsymbol{E}[\boldsymbol{X}_N \boldsymbol{X}_N^T]$ and $\boldsymbol{b}(t) = \boldsymbol{E}[X(t)\boldsymbol{X}_N]$, i.e., $\boldsymbol{R}$ represents the covariance matrix of the data vector $\boldsymbol{X}_N$ and $\boldsymbol{b}(t)$ is the vector of covariances between $X(t)$ and $\boldsymbol{X}_N$. Hence, $\boldsymbol{R} = [R(t_i - t_j)]$ and $\boldsymbol{b}(t) = [R(t - t_i)]$. It is also worth observing that $\widehat{X}(t)$ in (1) interpolates data, i.e., $\widehat{X}(t_k) = X(t_k)$. This is the well known fact in the stochastic kriging theory [4]. The interpolation property can be harmful due to the data uncertainty and numerical stability and one needs some degree of smoothing. This is easily avoided by some sort of regularization, i.e., when the matrix $\boldsymbol{R}$ is replaced by its regularized version $(\boldsymbol{R} + \lambda \boldsymbol{I})$ for some $\lambda > 0$. The improved reconstruction accuracy can be obtained by incorporating the additional information on the examined class of signals. In particular, the signal derivative provides the valuable information about the signal local variability. Hence, let the observed information be $[\boldsymbol{X}_N, \mathring{\boldsymbol{X}}_N]$, where $\mathring{\boldsymbol{X}}_N = [X^{(1)}(t_1), \ldots, X^{(1)}(t_N)]^T$ are the samples of the signal derivative. Analogously as before we need to consider the conditional process $W_N(t) = X(t)|\boldsymbol{X}_N, \mathring{\boldsymbol{X}}_N$. Some algebra shows that the mean value of $W_N(t)$ being the optimal **MSE** predictor of $X(t)$ can be written as

$$\widetilde{X}(t) = (\boldsymbol{P}_{11}\boldsymbol{b}(t) + \boldsymbol{P}_{21}\boldsymbol{d}(t))^T \boldsymbol{X}_N + (\boldsymbol{P}_{12}\boldsymbol{b}(t) + \boldsymbol{P}_{22}\boldsymbol{d}(t))^T \mathring{\boldsymbol{X}}_N, \qquad (2)$$

where the vector $\boldsymbol{d}(t) = \boldsymbol{E}[X(t)\mathring{\boldsymbol{X}}_N]$ represents the covariance between $X(t)$ and the derivative data, i.e., it has components $[-R^{(1)}(t - t_i)]$. Also $\{\boldsymbol{P}_{ij}\}$ are the matrices defining the block matrix $\boldsymbol{P}$ that is the inverse of the observed data covariance matrix Σ. The matrix Σ has the following block structure

$$\Sigma = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{Q} \\ \boldsymbol{Q}^T & \boldsymbol{S} \end{bmatrix},$$

where $\boldsymbol{S} = \boldsymbol{E}[\mathring{\boldsymbol{X}}_N \mathring{\boldsymbol{X}}_N^T]$ is the covariance of the signal derivatives that consists of the elements $\{-R^{(2)}(t_i - t_j)\}$. Also $\boldsymbol{Q} = \boldsymbol{E}[\boldsymbol{X}_N \mathring{\boldsymbol{X}}_N^T]$ is the covariance between the signal samples and its derivative samples, i.e., it has the entries $\{-R^{(1)}(t_i - t_j)\}$, where $R^{(1)}(0) = 0$. The matrices $\boldsymbol{R}, \boldsymbol{S}$ and Σ are symmetric and positively defined, whereas $\boldsymbol{Q}$ is skew-symmetric as $\boldsymbol{Q}^T = -\boldsymbol{Q}$. The **MSE** of the optimal predictor $\widetilde{X}(t)$ is given by $\tau_N^2(t) = \sigma^2 - \boldsymbol{c}(t)^T \Sigma^{-1} \boldsymbol{c}(t)$, where $\boldsymbol{c} = [\boldsymbol{b}(t), \boldsymbol{d}(t)]^T$. It can be shown that $\tau_N^2(t)$ is much smaller than the **MSE** of the predictor in (1) revealing the benefits of using the derivative information in the reconstruction process. Nevertheless, in practical situations it is difficult to obtain the derivative information about the signal and one commonly confines the reconstruction to the signal data only. In this paper we employ the information about the process not in the form of the classical time-domain sampling but from samples generated by level-crossings. The level-crossing sampling provides the universal way of characterizing the signal slope. This is summarized by a Slepian model that describes the behavior of the process in the neighborhood of crossings of the pre-selected levels or other pre-specified events (like local extrema). In the case of level-crossings and stationary Gaussian models the Slepian model consists of a linear regression function with the universal Rayleigh component that characterizes the signal slope and being independent on the selected crossing levels. Hence, our goal is to approximate the conditional process $W_N(t)$ that yields the predictor $\widetilde{X}(t)$ in (2) by the following process

$$W_N^S(t) = X(t)|\text{data at level crossings, crossing rates}. \qquad (3)$$

The information given in this conditional process includes the crossing values, i.e., $X(t) \in \Theta$, where $\Theta = \{u_1, \ldots, u_L\}$ is the set of the predefined levels and crossing rates (derivatives) to be approximated by the Slepian model. The theory of level-crossings including the Slepian model is discussed in Section 3. An example of the multiple-level data set is illustrated in Fig.1.

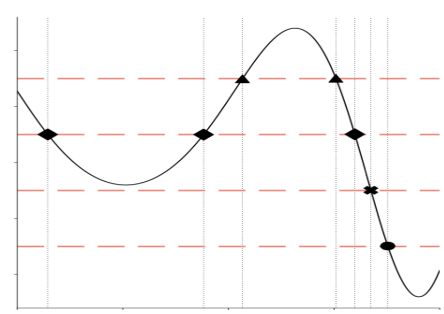

Fig. 1. Multiple-level data set

III. Signal Recovery from Level Crossings

The set of level crossings $\{t \in [0, T] : X(t) = u\}$ for the fixed level u has been examined since the seminal work of Rice [14], [9], [7]. In particular, for the stationary Gaussian signal the average number of the u–crossings (denoted as $N_u(T)$) is given by the Rice formula $\boldsymbol{E}[N_u(T)] = T\pi^{-1}(\mu_2/\sigma^2)^{1/2}\exp(-u^2/2\sigma^2)$. Assuming (without loss of generality) that $\sigma^2 = 1$ this allows us to estimate the critical parameter μ_2 from the zero crossings. In fact, $\pi N_0(T)/T$ is the unbiased estimate of $\sqrt{\mu_2}$ [11]. The problem of signal reconstruction from level crossings requires, however, more detailed information about the behavior of the process at the vicinity of crossing times $\{t_k\}$. This additional information can be provided by the events $\mathcal{E}_k = \{X(t_k), X^{(1)}(t_k)\}$. Using the special case of (2) we note that the optimal predictor of $X(t+t_k)$ based on $\mathcal{E}_k$ is given by

$$\frac{R(t)}{\sigma^2}X(t_k) - \frac{R^{(1)}(t)}{\mu_2}X^{(1)}(t_k). \tag{4}$$

For the data derived from the multiple level-crossings we set $X(t_k) = u(t_k)$, $u(t_k) \in \Theta$, whereas the derivative $X^{(1)}(t_k)$ can be approximated by averaging over a large number of observations of the events $\{\mathcal{E}_k\}$. It was shown by Slepian [14], [8], [15], [17], [16], that the empirical distribution of $X^{(1)}(t_k)$ over an increasing number of the events $\{\mathcal{E}_k\}$ at the points of the crossings tends to the Rayleigh random variable η with the parameter μ_2. The average value of η is $(\pi\mu_2/2)^{1/2}$ or $-(\pi\mu_2/2)^{1/2}$ depending whether we have the upcrossing or downcrossing event, respectively. The aforementioned results and (4) lead to the Slepian model for the optimal predictor of $X(t)$ in the neighborhood of the crossing time t_k

$$\mathcal{S}_k(t) = \frac{R(t-t_k)}{\sigma^2}u(t_k) - m_k\frac{R^{(1)}(t-t_k)}{\mu_2}\eta_k, \tag{5}$$

where $m_k = 1$ or $m_k = -1$ for the upcrossing or the downcrossing events, respectively. It is an important to note that η_k is the universal Rayleigh model of the signal derivative that is independent on the given level set Θ. Thus, the formula in (5) is the linear regression model with the random slope. Note that $\mathcal{S}_k(t_k) = u(t_k)$ and $\mathcal{S}_k^{(1)}(t_k) = m_k\eta_k$. Hence, the Slepian model reproduces the position and the approximate slope of the process. It is worth noting that for signals with large bandwidth (large μ_2) the second term in (5) is of the smaller order than the first one. On the other hand for $u(t_k) = 0$ (zero-crossings) the second term in (5) plays the critical role. The residual process $X(t) - \mathcal{S}_k(t)$ of the Slepian model is the zero mean nonstationary Gaussian process with the variance

$$\tau_k^2(t) = \sigma^2 - \frac{R^2(t-t_k)}{\sigma^2} - \frac{(R^{(1)}(t-t_k))^2}{\mu_2}. \tag{6}$$

Let us note that $\tau_k^2(t_k) = 0$ and $\tau_k^2(t) \to \sigma^2$ as $t - t_k \to \infty$. The latter means that for long times after crossings times the Slepian signal model tends to the original stationary process. The aforementioned results allow us to formulate the reconstruction formula derived from the conditional process in (3). This is based on the universal Slepian model of approximating the crossing rates. The prediction rule is the formal version of (2) where the exact values of $\{X^{(1)}(t_k)\}$ are substituted by the Rayleigh random variables $\{\eta_k\}$, whereas the signal values $\{X(t_k)\}$ by $\{u(t_k)\}$. Hence, we have

$$\begin{aligned}\widetilde{X}_S(t) = &(\boldsymbol{P}_{11}\boldsymbol{b}(t) + \boldsymbol{P}_{21}\boldsymbol{d}(t))^T\boldsymbol{U}_N \\ &+ (\boldsymbol{P}_{12}\boldsymbol{b}(t) + \boldsymbol{P}_{22}\boldsymbol{d}(t))^T\boldsymbol{m}_N^T\boldsymbol{\eta}_N,\end{aligned} \tag{7}$$

where $\boldsymbol{U}_N = [u(t_1), \ldots, u(t_N)]^T$, $u(t_k) \in \Theta$ and $\boldsymbol{\eta}_N = [\eta_1, \ldots, \eta_N]^T$ with $\boldsymbol{m}_N$ being the ± 1 vector of the crossings signs. Also N stands for the total number of level crossings in $[0, T]$. The components of the Rayleigh slope vector $\boldsymbol{\eta}_N$ can be in practical implementations replaced by the average value $\pm(\pi\mu_2/2)^{1/2}$. Also there is a formal way of detecting the crossing signs based on the sign of $R^{(1)}(t)X(0)$, where we assume that $X(0)$ is observed. The theoretical analysis of the estimate $\widetilde{X}_S(t)$ in (7) reveals that we have the following fundamental bounds

$$\mathbf{MSE}(\widetilde{X}) \leq \mathbf{MSE}(\widetilde{X}_S) \leq \mathbf{MSE}(\widehat{X}),$$

where $\widehat{X}(t)$ is the reconstruction based only on the signal observations (see (1)), whereas $\widetilde{X}(t)$ utilizes the additional information of the true signal derivatives (see (2)). The difference between $\widetilde{X}_S(t)$ and $\widetilde{X}(t)$ depends on the number and distribution of the used level-crossings and the signal bandwidth. In fact, the **MSE** between $\{\eta_k\}$ and $\{X^{(1)}(t_k)\}$ is proportional to μ_2 and as we have already mentioned μ_2 depends linearly on the signal bandwidth. Hence, for the small bandwidth we have $\mathbf{MSE}(\widetilde{X}_S) \simeq \mathbf{MSE}(\widetilde{X})$.

IV. Numerical Examples

In order to illustrate the above results let us consider the low-pass bandlimited Gaussian signal with $R(\tau) = sinc(\omega_0\tau)$ and the corresponding moments $\sigma^2 = 1$ and $\mu_2 = \omega_0^2/3$. This process represents the case of the lack of a priori information about the signal shape (except the bandlimitedness) since its spectral density is flat. In Fig. 2 we show the reconstruction error as a function of the level u in the case of the three-level symmetric crossings data set, i.e., when $\Theta = \{-u, 0, u\}$. We display the accuracy of the reconstructions methods $\widehat{X}(t)$, $\widetilde{X}(t)$ and $\widetilde{X}_S(t)$. It is plain that the accuracy gets worst with an increasing value of u as the number of crossings points (level-crossing rate) decreases. The small value of u also reduces the accuracy due to the presence of the highly correlated samples and inability of representing peak values of the signal. This suggests an interesting problem of designing the proper level crossing set Θ.

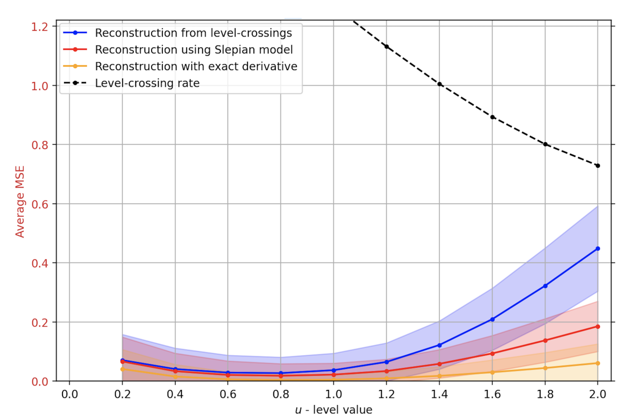

Fig. 2. The reconstruction accuracy versus u for the three-level set.

REFERENCES

[1] R. J. MarksII, *Handbook of Fourier Analysis and Its Applications*. Oxford University Press, 2009.

[2] Y. C. Eldar, *Sampling Theory*. Cambridge University Press, 2015.

[3] H. Choi and D. Munson, "Stochastic formulation of bandlimited signal interpolation," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 1, pp. 82–85, 2000.

[4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2005.

[5] J. Kocijan, *Modelling and Control of Dynamic Systems Using Gaussian Process Models*. Springer, 2016.

[6] L. Swiler, "Survey of constrained Gaussian process regression: approaches and implementation challenges," *J. of Machine Learning for Modelling and Computing*, vol. 1, pp. 119–156, 2021.

[7] M. F. Kratz, "Level crossings and other level functionals of stationary Gaussian processes," *Probab. Survey*, vol. 3, no. 1, pp. 230–288, 2006.

[8] M. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983.

[9] G. Lindgren, "Gaussian integrals and Rice series in crossing distributions - to compute the distribution of maxima and other features of Gaussian processes," *Statistical Science*, vol. 34, no. 1, pp. 100–128, 2019.

[10] M. Miskowicz, *Event-Based Control and Signal Processing*. CRC Press, 2015.

[11] D. Rzepka, M. Pawlak, D. Koscielnik, and M. Miskowicz, "Bandwidth estimation from multiple level-crossings of stochastic signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2488–2502, 2017.

[12] P. Hall and A. Yatchew, "Nonparametric estimation when data on derivatives are available," *Ann. Statist.*, vol. 35, pp. 300–323, 2007.

[13] L. de Roos et al., "High-dimensional Gaussian process inference with derivatives," in *Proceedings of the 38th International Conference on Machine Learning*, 2021, pp. 2535–2545.

[14] G. Lindgren, *Stationary Stochastic Processes: Theory and Applications*. CRC Press, 2012.

[15] G. Lindgren and M. Prevosto, "The relation between wave asymmetry and particle orbits analysed by Slepian models," *J. Fluid Mech.*, vol. 924, no. 1, pp. 1–27, 2021.

[16] M. Grigoriu, "First passage times for Gaussian processes by Slepian models," *Prob. Eng .Mech.*, vol. 61, no. 1, pp. 103–125, 2020.

[17] T. T. Gadrich and R. J. Adler, "Slepian models for non-stationary Gaussian processes," *Journal of Applied Probability*, vol. 30, no. 1, pp. 98–111, 1993.

[18] A. S. T. Kailath and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.