PAPER *Special Section on Sparsity-Aware Signal Processing*

Sampling Signals with Finite Rate of Innovation and Recovery by Maximum Likelihood Estimation

Akira HIRABAYASHI[†a)], ***Member***, **Yosuke HIRONAGA**[††], ***Student Member***, *and* **Laurent CONDAT**[†††], ***Nonmember***

SUMMARY We propose a maximum likelihood estimation approach for the recovery of continuously-defined sparse signals from noisy measurements, in particular periodic sequences of Diracs, derivatives of Diracs and piecewise polynomials. The conventional approach for this problem is based on least-squares (a.k.a. annihilating filter method) and Cadzow denoising. It requires more measurements than the number of unknown parameters and mistakenly splits the derivatives of Diracs into several Diracs at different positions. Moreover, Cadzow denoising does not guarantee any optimality. The proposed approach based on maximum likelihood estimation solves all of these problems. Since the corresponding log-likelihood function is non-convex, we exploit the stochastic method called particle swarm optimization (PSO) to find the global solution. Simulation results confirm the effectiveness of the proposed approach, for a reasonable computational cost.

key words: *Signals with finite rate of innovation, sequence of Diracs, derivatives of Diracs, piecewise polynomials, maximum likelihood estimation, Cadzow denoising*

1. Introduction

Compression plays the critical role in modern communications systems. In the standard approach, a huge amount of high quality data is first acquired, and then it is compressed to ten or five percent of essential data: only the compressed data is transmitted to receivers. This means not only that most of the acquired data are discarded, but also that both sampler and encoder are necessary to facilitate transmission. If we could directly extract the small amount of the essential data, there would be no data discarded and only sampler is necessary. Sparse sampling is nothing but a technique which enables us to do this [1]. There are two approaches for sparse sampling. One is the so-called compressed sensing, which is a technique for discrete signals [2], [3]. The other is sampling theory for signals with finite rate of innovation [4]–[7], which form a class of continuously-defined signals. This paper focuses on the latter.

Typical examples of signals with finite rate of innovation appear in radar, echo, or sonar. In these techniques, radio or ultrasonic waves are transmitted to a target object, and the reflected waves enable us to compute the distance to the target. Even though channels for reflection are not single, the sum of the reflected waves is sparse. Since the transmitted pulse is known, such sparse signals can be expressed only using a few parameters, i.e., time delays and attenuation coefficients. We can exploit this parametric expression to sample continuously-defined sparse signals at very low frequency compared to the so-called Nyquist frequency.

The most important type of signals with finite rate of innovation is the sequences of Dirac distributions (Diracs, in short). It lies at the heart of the theories formulated for analog signals. This is because simple convolutions of such sequences with particular kernels creates a wide variety of signals of practical interest. An even larger class of signals is generated by convolutions from sequences of *derivatives* of Diracs, including the important cases of piecewise polynomials and piecewise sinusoids with discontinuities [4], [8].

Let $\delta(t)$ denote the Dirac distribution and τ be a positive real. This paper discusses τ-periodic sequences of Diracs and derivatives of Diracs and τ-periodic piecewise polynomials. For example, the sequence of derivatives of Diracs is expressed as $s(t) = \sum_{k'\in\mathbb{Z}} s_0(t - k'\tau)$, where

$$s_0(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} c_{k,r}\delta^{(r)}(t - t_k), \tag{1}$$

for some known integers $K \geq 1$ and $\{R_k\}_{k=0}^{K-1}$. This signal has K degrees of freedom due to the time instants $\{t_k\}_{k=0}^{K-1}$ and $\tilde{K} = \sum_{k=0}^{K-1} R_k$ degrees of freedom due to the coefficients $\{c_{k,r}\}$, per period τ. Thus, the rate of innovation of the signal is $\rho = (K + \tilde{K})/\tau < \infty$. The signal $s(t)$ is sampled using an appropriate kernel, like the Dirichlet kernel [4] or a sum-of-sincs [7]. Then, the sequence can be perfectly reconstructed from the noiseless measurements using the annihilating filter technique [4]. This technique recasts the problem of obtaining t_k as that of computing the filter coefficients. As a result, the conventional approach requires at least $2\tilde{K} + 1$ measurements [9], which is always more than the number of unknown parameters $K + \tilde{K}$ because $\tilde{K} \geq K$. If $\tilde{K}$ is much greater than K, $2\tilde{K} + 1$ is also much greater than $K + \tilde{K}$. If the measurements are corrupted by noise, the annihilating filter approach, a.k.a. least squares [6], yields $\tilde{K}$ locations instead of K. To improve the performance of least squares, Cadzow denoising [10] is usually exploited [6]; this method merely tries to find a Toeplitz matrix of rank K, hence does not guarantee any optimality [11].

Manuscript received January 20, 2013.
Manuscript revised May 3, 2013.
[†]The author is with the College of Information Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan.
[††]The author is with the Graduate School of Medicine, Yamaguchi University, Yamaguchi 755-8611, Japan.
[†††]The author is with GIPSA-lab, Grenoble Institute of Technology, Grenoble, France
a) E-mail: akirahrb@media.ritsumei.ac.jp
DOI: 10.1587/transfun.E0.A.1

Table 1 Comparison of the proposed and conventional approaches.

	Problem to be solved	Solution
Proposed	Maximum likelihood estimation	PSO (Stochastic global optimization)
Conventional [6]	Find a Toeplitz matrix of rank K	Cadzow (Alternative pseudo projection)

To solve these problems, we propose a method that reconstructs the signal using maximum likelihood estimation. The proposed method directly estimates t_k without recasting the problem as the filter coefficients problem. The corresponding likelihood function is non-convex. Hence, to find the global solution, we exploit a heuristic approach called *particle swarm optimization* (PSO) [12]. Even though the proposed method does not guarantee optimality of the solution either, we can expect that it is near optimal. Simulation results show that the proposed method can perfectly reconstruct the signal from less than $2\tilde{K}+1$ measurements, whereas the conventional approach is not applicable in this situation. We should note that, even though both approaches suffer from the non-convexity, we can summarize contributions of the present paper as follows: 1) we exploit the maximum likelihood estimation, which was not used in the conventional approach, 2) we use PSO to find the global optimum solution of the likelihood function. These points are summarized in Table 1.

This paper is organized as follows. Section 2 defines the signals with finite rate of innovation and describes the sampling setup using the sum-of-sincs kernel. Section 3 formulates maximum likelihood estimation for reconstruction of the sequence of Diracs. Section 4 applies the same approach to the sequence of derivatives of Diracs. Section 5 is devoted to the periodic piecewise polynomials. Section 6 concludes the paper.

2. Sampling Signals with Finite Rate of Innovation

Consider a signal represented by linear combination of arbitrary shifts of R known function $\varphi_r(t)$ $(r = 0, \ldots, R-1)$, but the shift amounts t_k and the coefficients $c_{k,r}$ are unknown. We assume $t_k < t'_k$ when $k < k'$. Then, the signal $s(t)$ is represented by

$$s(t) = \sum_{k=-\infty}^{\infty} \sum_{r=0}^{R-1} c_{k,r}\varphi_r(t - t_k). \tag{2}$$

The total number of t_k in period $[t_a, t_b]$ and $c_{k,r}$ with the identical k is denoted by $C_s(t_a, t_b)$. Then, we define a rate of innovation ρ as

$$\rho = \lim_{\tau\to\infty} \frac{1}{\tau} C_s(-\tau/2, \tau/2). \tag{3}$$

Definition 1: [4] A signal with a finite rate of innovation is a signal whose parametric representation is given in Eq. (2) and with a finite ρ, as defined in Eq. (3).

We can also define a local rate of innovation with respect to a moving (yet fixed) window size τ, as

$$\rho_\tau(t) = \frac{1}{\tau} C_s(t - \tau/2, t + \tau/2) \tag{4}$$

In this case, one is often interested in its maximum:

$$\rho_{\max}(\tau) = \max_{t\in\mathbf{R}} \rho_\tau(t)$$

If a signal has a period τ, the local rate of innovation $\rho_\tau(t)$ is useful because it does not depend on t and gets a constant ρ.

This paper also discusses periodic signals $s(t)$, defined by

$$s(t) = \sum_{k'\in\mathbb{Z}} s_0(t - k'\tau), \tag{5}$$

where $s_0(t)$ is the signal in the interval $[0, \tau)$, given as

$$s_0(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} c_{k,r}\varphi_r(t - t_k). \tag{6}$$

In this case, we enforce the condition that $0 \le t_0 < \cdots < t_{K-1} < \tau$. The sequence of Diracs is $s(t)$ in Eqs. (5) and (6) with $R = 1$ and $\varphi_0(t) = \delta(t)$. This is typically sparse, because its value is mostly zero except at positions t_k. Further, this ideal pulse sequence produces the general pulse sequence by convolving with $\varphi(t) \neq \delta(t)$. One generalization of the sequence of Diracs is the sequence of derivatives of Diracs. This is $s(t)$ with $\varphi_r(t) = \delta^{(r)}(t)$. Here the derivatives of Dirac is defined by

$$\int_{-\infty}^{\infty} \delta^{(r)}(t)\phi(t)dt = (-1)^r \phi^{(r)}(0),$$

where $\phi(t)$ is an arbitrary function that has derivatives of any order and tends to zero more rapidly than any power of t, as $|t|$ tends to infinity [13]. This signal is related to piecewise polynomials by $R+1$th derivatives. The present paper discusses these three types of signals.

The target signal is sampled using a kernel $\psi(t)$ and yields N noiseless measurements

$$d_n = \langle s, \psi_n \rangle = \int_{-\infty}^{\infty} s(t)\overline{\psi(t - nT)}dt,$$

for $n = 0, \ldots, N-1$, $T = \tau/N$, and $\bar{z}$ stands for the conjugate of a complex number z. We adopt for $\psi(t)$ the *sum of sincs* kernel [7], which is defined in the frequency domain by

$$\hat{\psi}(\omega) = \frac{\tau}{\sqrt{2\pi}} \sum_{p=-P}^{P} \mathrm{sinc}\left(\frac{\omega}{\frac{2\pi}{\tau}} - p\right), \tag{7}$$

where

$$\mathrm{sinc}(\omega) = \begin{cases} \sin(\pi\omega)/(\pi\omega) & (\omega \neq 0), \\ 1 & (\omega = 0). \end{cases}$$

Its time domain expression is

$$\psi(t) = \frac{\mathrm{rect}(t/\tau)}{\tau} \sum_{p=-P}^{P} b_p e^{i2p\pi t/\tau}, \tag{8}$$

where $\mathrm{rect}(t) = 1$ if $|t| \le 0.5$ else 0 and $P \le (N-1)/2$ is an integer. By setting $b_p = 1$ for all p, this kernel reduces to the standard Dirichlet kernel. Let $\hat{d}_p = \frac{1}{\tau}\int_0^\tau s(t)e^{-i2p\pi t/\tau}dt$ be the Fourier coefficients of $s(t)$. Then, it follows from Eq. (8) that

$$d_n = \sum_{p=-P}^{P} b_p \hat{d}_p e^{i2pn\pi/N}.$$

This admits the matrix representation

$$\boldsymbol{d} = F^{-1}B\hat{\boldsymbol{d}}, \tag{9}$$

where B is the diagonal matrix $\mathrm{diag}(b_{-P}, \ldots, b_P)$ and F is the discrete Fourier transform (DFT) matrix, defined accordingly. In a nutshell, the Fourier coefficients are related to the noiseless measurements acquired using the sinc kernel *exactly* by the DFT.

3. Sequence of Diracs

The Fourier coefficients for the sequence of Diracs can be derived from Eqs. (5) and (6) with $\varphi(t) = \delta(t)$ as

$$\hat{d}_p = \sum_{k=0}^{K-1} \tilde{c}_k u_k^p,$$

where $u_k = e^{-i2\pi t_k/\tau}$, $\tilde{c}_k = c_k/\tau$. These values can be computed from the noiseless measurements d_n by $\hat{\boldsymbol{d}} = B^{-1}F\boldsymbol{d}$. The locations t_k then can be extracted by using the annihilating filter method and c_k are derived by solving the Vandermonde equation [4].

The clean measurements $\{d_n\}_{n=0}^{N-1}$ are corrupted by additive noise, yielding the noisy measurements $y_n = d_n + e_n$, for $n = 0, \ldots, N-1$. Let $\boldsymbol{y}$ and $\boldsymbol{e}$ be vectors whose n-th elements are y_n and e_n, respectively: $\boldsymbol{y} = \boldsymbol{d} + \boldsymbol{e}$. In this case, there does not exist an FIR filter that annihilates the sequence $\hat{\boldsymbol{y}} = B^{-1}F\boldsymbol{y}$ in general. The relevant solution for this situation is to use a filter which minimizes sum of squared residues. This approach is called least square (LS). If noise level is not moderate, some preprocessing is necessary. For this, Cadzow denoising [10] is the standard approach [6], which is one of the technique for structured low-rank approximation of matrix [14]. A data matrix which appears in the process has the Toeplitz structure and should have rank K, but not because of noise. Therefore, the algorithm iteratively find a matrix which has the structure and rank K. Note that the set of Toeplitz matrix is convex while that of rank K is not. Because of this, any optimality is not guaranteed for the solution obtained by the algorithm.

To resolve these difficulties, we exploit the formalism of maximum likelihood estimation. Let U_t and $\boldsymbol{c}$ be the matrix and the vector defined respectively as

$$U_t = \begin{pmatrix} u_0^{-P} & \cdots & u_{K-1}^{-P} \\ u_0^{-P+1} & \cdots & u_{K-1}^{-P+1} \\ \vdots & \ddots & \vdots \\ u_0^{P} & \cdots & u_{K-1}^{P} \end{pmatrix},$$

$$\boldsymbol{c} = [\tilde{c}_0 \ \tilde{c}_1 \ \cdots \ \tilde{c}_{K-1}]^T.$$

Then, we have $\hat{\boldsymbol{d}} = U_t\boldsymbol{c}$ and Eq. (9) yields

$$\boldsymbol{d} = F^{-1}BU_t\boldsymbol{c}. \tag{10}$$

Assume that the probability density function $p(\boldsymbol{e})$ is known. Then using Eq. (10), we can define the log-likelihood function as $L(\boldsymbol{t}, \boldsymbol{c}) = \log p(\boldsymbol{y} - F^{-1}BU_t\boldsymbol{c})$, where $\boldsymbol{t} = [t_0 \ t_1 \ \cdots \ t_{K-1}]^{\mathrm{T}}$. Assume that $p(\boldsymbol{e})$ is the Gaussian distribution with zero mean and covariance matrix $\sigma^2 I$, where σ is a known positive real and I is the identity matrix. Then, the log-likelihood function reads

$$L(\boldsymbol{t}, \boldsymbol{c}) = -\frac{\|\boldsymbol{y} - F^{-1}BU_t\boldsymbol{c}\|^2}{2\sigma^2} - N\log(\sqrt{2\pi}\sigma). \tag{11}$$

This implies that the maximization of the log-likelihood function is equivalent to the minimization of the norm $\|\boldsymbol{y} - F^{-1}BU_t\boldsymbol{c}\|^2$. Further on, F is unitary up to constant. Hence, this minimization is equivalent to that of

$$f_o(\boldsymbol{t}, \boldsymbol{c}) = \|\hat{\boldsymbol{y}} - BU_t\boldsymbol{c}\|^2. \tag{12}$$

Finally, maximum likelihood estimation amounts to estimating the vector $BU_t\boldsymbol{c}$, which is the closest to $\hat{\boldsymbol{y}}$ in the least-squares sense, in Fourier domain.

Eq. (12) is quadratic with respect to $\boldsymbol{c}$, when $\boldsymbol{t}$ is fixed. Therefore, the optimal $\boldsymbol{c}$ for a fixed $\boldsymbol{t}$ is obtained analytically as $\boldsymbol{c} = (BU_t)^\dagger\hat{\boldsymbol{y}}$, where $T^\dagger$ stands for the Moore-Penrose generalized inverse of the bounded operator T [15]. Hence, the minimizer of $f_o(\boldsymbol{t}, \boldsymbol{c})$ is found by searching $\boldsymbol{t}$ that minimizes

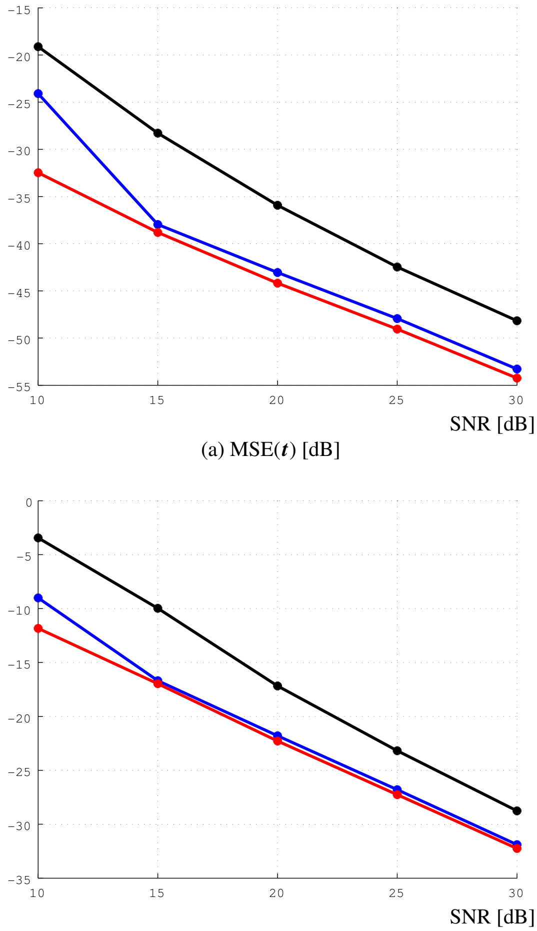
$$f(\boldsymbol{t}) = f_o(\boldsymbol{t}, (BU_t)^\dagger\hat{\boldsymbol{y}}) = \|\hat{\boldsymbol{y}} - (BU_t)(BU_t)^\dagger\hat{\boldsymbol{y}}\|^2,$$

and then by computing $\boldsymbol{c} = (BU_t)^\dagger\hat{\boldsymbol{y}}$.

The criterion $f(\boldsymbol{t})$ is non-convex and it is very difficult to find the global minimum solution. We thus exploit the so-called particle swarm optimization (PSO) algorithm [12]. The particles model the parameter $\boldsymbol{t}$ to be optimized. For each particle $j = 1, ..., J$, we first initialize the position $\boldsymbol{t}_j$ and its velocity $\dot{\boldsymbol{t}}_j$ with uniformly distributed random vectors in the domain. We use the particle's and swarm's best known positions $\boldsymbol{b}_j^{(p)}$ and $\boldsymbol{b}^{(s)}$, which are initialized by $\boldsymbol{t}_j$ and the best among the initial positions, respectively. Until a termination criterion is met, the particle's velocity $\dot{\boldsymbol{t}}_j$ and position $\boldsymbol{t}_j$ are updated by

$$\dot{\boldsymbol{t}}_j \leftarrow w\dot{\boldsymbol{t}}_j + c_1 r_1(\boldsymbol{b}_j^{(p)} - \boldsymbol{t}_j) + c_2 r_2(\boldsymbol{b}^{(s)} - \boldsymbol{t}_j)$$
$$\boldsymbol{t}_j \leftarrow \boldsymbol{t}_j + \dot{\boldsymbol{t}}_j,$$

respectively, where c_1, c_2 are pre-defined constants near 1 and r_1, r_2 are uniform random variables within 0 and 1. If $f(\boldsymbol{t}_j) < f(\boldsymbol{b}_j^{(p)})$, then $\boldsymbol{b}_j^{(p)}$ is updated by $\boldsymbol{t}_j$. If $f(\boldsymbol{b}_j^{(p)}) < f(\boldsymbol{b}^{(s)})$, then $\boldsymbol{b}^{(s)}$ is replaced by $\boldsymbol{b}_j^{(p)}$. Finally, $\boldsymbol{b}^{(s)}$ gives the best found solution. Because of its global and random nature, PSO is more robust than gradient approaches, against getting trapped in local minima. The downside is a relatively high computational cost.



(a) MSE($\boldsymbol{t}$) [dB]

(b) MSE($\boldsymbol{c}$) [dB]

Fig. 1 Mean square errors (MSE) [dB] of estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{c}$ of a sequence of Diracs with respect to the SNR [dB]. The number of measurements is 11. The red, blue and black lines show the results by the proposed method, by LS with and without Cadzow denoising, respectively.

In simulations, the parameters are set as $\tau = 1$, $b_p = 1$, $K = 2$ and $N = 11$. The unknown parameters are $\boldsymbol{t} = (t_0, t_1) = (0.42, 0.52)$, and $\boldsymbol{c} = (c_0, c_1) = (1.00, 1.00)$. For PSO, we used $J = 150$ particles and $(w, c_1, c_2) = (0.4, 0, 9, 0.4)$, $(0.9, 0.4, 0.4)$ and $(0.4, 0.4, 0.9)$ for 75, 45 and 30 particles, respectively. One thousand noise vectors $\boldsymbol{e}$ were generated from the Gaussian distribution in which σ was determined so that the SNR† becomes 10, 15, …, 30 [dB]. For each experiment, we computed estimates $\hat{\boldsymbol{t}}$ and $\hat{\boldsymbol{c}}$ of $\boldsymbol{t}$ and $\boldsymbol{c}$, for 1,000 different noise realizations. Accordingly, the mean square errors MSE($\boldsymbol{t}$) and MSE($\boldsymbol{c}$) were defined as the average over the 1,000 trials of $\|\hat{\boldsymbol{t}} - \boldsymbol{t}\|^2$ and $\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|^2$, respectively. The results are shown in Fig. 1, where the red, blue and black lines show the results by the proposed method, by LS with and without Cadzow denoising, respectively. We can see that the proposed method outperforms the conventional methods for every value of the SNR.

†The SNR is defined by $10 \log_{10} \dfrac{\|\boldsymbol{d}\|^2}{\sigma^2 N}$.

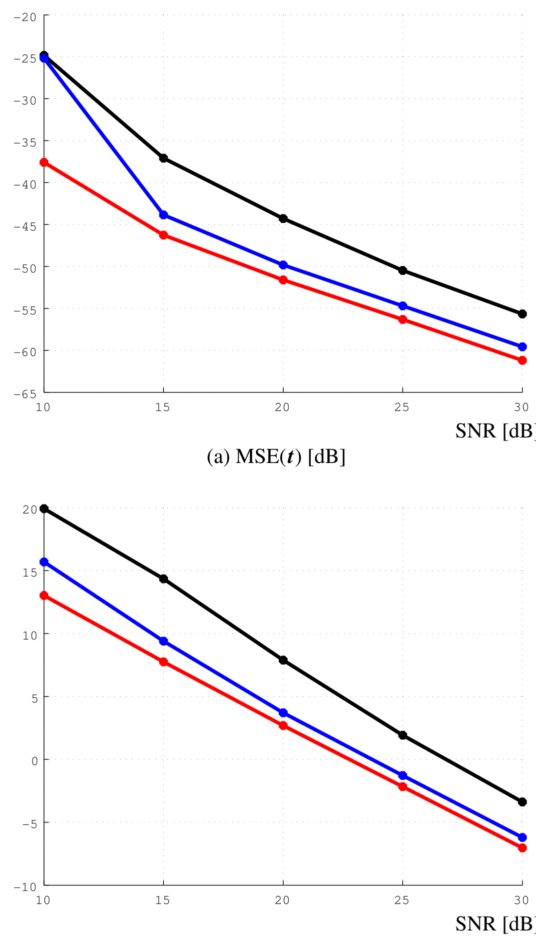
4. Sequence of Derivatives of Diracs

The Fourier coefficients of the sequence of derivatives of Diracs can be obtained from Eq. (6) with $\varphi_r(t) = \delta^{(r)}(t)$, as

$$\hat{d}_p = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} \tilde{c}_{k,r} p u_k^p,$$

where $\tilde{c}_{k,r} = (i2\pi)^r c_{k,r}/\tau^{r+1}$. This is essentially different from that of the sequence of Diracs because there is a polynomial term p in the right-hand side. The sequence of this Fourier coefficients still can be annihilated by an FIR filter that has $\tilde{K}$ coefficients, not K. The classical approach tries to find these coefficients in order to find K locations t_k. This is an augmentation of the number of unknown parameters. In noiseless case, the $\tilde{K}$ coefficients give the K locations t_k, but in noisy case, $\tilde{K}$ locations are retrieved. Cadzow algorithm can be applied to this denoising, but does not improve the situation, as we expect.

To solve these problems, we also exploit maximum



(a) MSE($\boldsymbol{t}$) [dB]

(b) MSE($\boldsymbol{c}$) [dB]

Fig. 2 MSE [dB] of estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{c}$ of a sequence of derivatives of Diracs with respect to the SNR [dB]. The legends are the same as in Fig. 1.

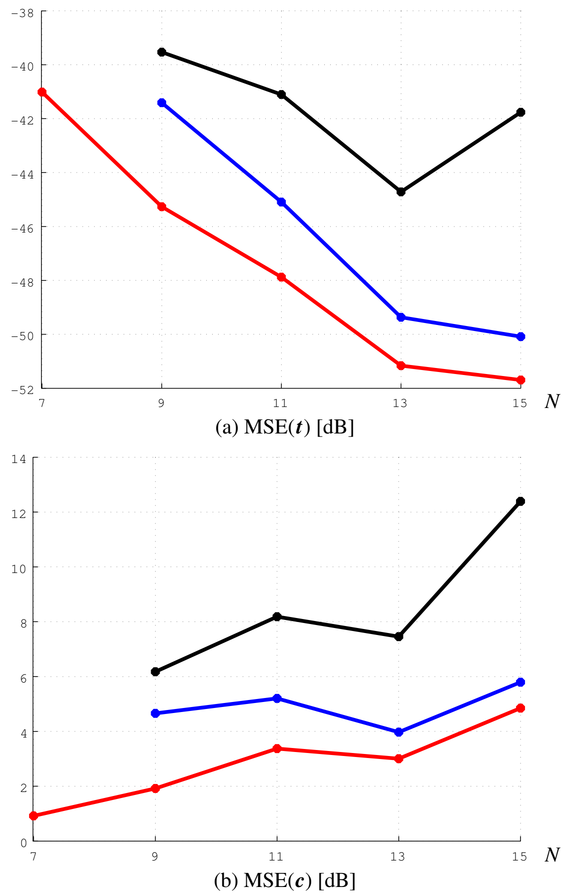


Fig. 3 MSE [dB] of the estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{c}$ of a sequence of derivatives of Diracs with respect to the number N of measurements. The lines show the SNR for 20dB.The legends are the same as in Fig. 1.

likelihood formulation. Let V_t and $\boldsymbol{c}$ be the matrix and the vector defined respectively as

$$V_t = \begin{pmatrix} u_0^{-P} & \cdots & (-P)^R u_{K-1}^{-P} \\ u_0^{-P+1} & \cdots & (-P+1)^R u_{K-1}^{-P+1} \\ \vdots & \ddots & \vdots \\ u_0^{P} & \cdots & (P)^R u_{K-1}^{P} \end{pmatrix},$$

$$\boldsymbol{c} = [\tilde{c}_{0,0}\ \tilde{c}_{0,1}\ \cdots\ \tilde{c}_{K-1,R-1}]^T.$$

Then, we have $\hat{\boldsymbol{d}} = V_t\boldsymbol{c}$ and therefore,

$$\boldsymbol{d} = F^{-1}BV_t\boldsymbol{c}. \tag{13}$$

Similarly to the case of the sequence of Diracs, the log-likelihood function reads

$$L(\boldsymbol{t}, \boldsymbol{c}) = -\frac{\|\boldsymbol{y} - F^{-1}BV_t\boldsymbol{c}\|^2}{2\sigma^2} + N\log(\sqrt{2\pi}\sigma), \tag{14}$$

which is equivalent to the minimization of

$$f_o(\boldsymbol{t}, \boldsymbol{c}) = \|\hat{\boldsymbol{y}} - BV_t\boldsymbol{c}\|^2. \tag{15}$$

The minimizer of this term is found by searching $\boldsymbol{t}$ that minimizes

$$f(\boldsymbol{t}) = f_o(\boldsymbol{t}, (BV_t)^\dagger\hat{\boldsymbol{y}}) = \|\hat{\boldsymbol{y}} - (BV_t)(BV_t)^\dagger\hat{\boldsymbol{y}}\|^2,$$

and then by computing $\boldsymbol{c} = (BV_t)^\dagger\hat{\boldsymbol{y}}$. The search of the minimizer was again conducted by PSO with the same setup for the inner parameters as is in the case of the sequence of Diracs.

In simulations, the parameters are set as $\tau = 1$, $b_p = 1$, $K = 2$, and $R_0 = R_1 = 2$. The unknown parameters are $\boldsymbol{t} = (t_0, t_1) = (0.19, 0.63)$, and $\boldsymbol{c} = (c_{0,0}, c_{0,1}, c_{1,0}, c_{1,1}) = (-0.80, 0.65, -1.50, 0.85)$. For each experiment, we computed estimates $\hat{\boldsymbol{t}}$ and $\hat{\boldsymbol{c}}$ of $\boldsymbol{t}$ and $\boldsymbol{c}$, for 1,000 different noise realizations. The mean square errors MSE($\boldsymbol{t}$) and MSE($\boldsymbol{c}$) were defined as the average over the 1,000 trials of $\|\hat{\boldsymbol{t}} - \boldsymbol{t}\|^2$ and $\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|^2$, respectively. The number of measurements are $N = 13$. The results are shown in Fig. 2. As well as in Fig. 1, the red, blue and black lines show the results by the proposed method, by LS with and without Cadzow denoising, respectively. We can see that the proposed method performs better than the conventional approaches, whatever the SNR. Fig. 3 shows MSEs with respect to the number of measurements. The noise level is SNR=20dB. We can see the proposed approach always outperforms the conventional approaches in these simulations as well. Note that the LS approach with/without Cadzow cannot be applied to the case of seven measurements, while the proposed method performs the best in this case for the estimation of $\boldsymbol{c}$. We have not yet clarified the reason why MSE($\boldsymbol{c}$) by the proposed method increases as N increases.

5. Periodic Piecewise Polynomials

For every $k = 0, \ldots, K-2$, let us define the function $\varphi_k(t)$ as

$$\varphi_k(t) = \begin{cases} v_k(t) & (t_k < t < t_{k+1}), \\ 0 & \text{(otherwise)}, \end{cases}$$

and the function $\varphi_{K-1}(t)$ as

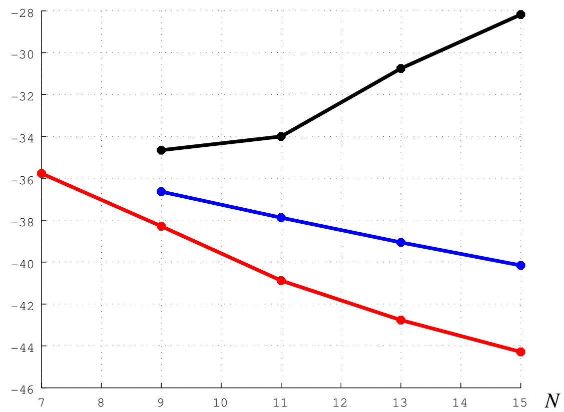
$$\varphi_{K-1}(t) = \begin{cases} v_{K-1}(t+\tau) & (0 \le t < t_0), \\ v_{K-1}(t) & (t_{K-1} < t < \tau), \\ 0 & \text{(otherwise)}, \end{cases}$$

where $v_k(t) = \sum_{r=0}^{R} \alpha_{k,r} t^r$. Then, a τ-periodic piecewise polynomial $s(t)$ of degree R is defined by $s(t) = \sum_{k' \in \mathbb{Z}} s_0(t - k'\tau)$, with

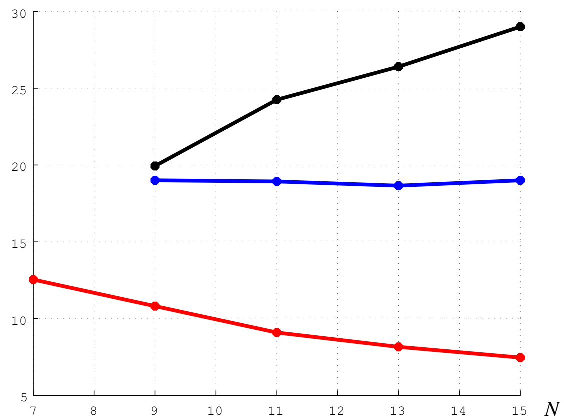
$$s_0(t) = \sum_{k=0}^{K-1} \varphi_k(t).$$

The available samples are $d_n = \langle s, \psi_n \rangle$ corrupted by noise: $y_n = d_n + e_n$.

As mentioned before, the $R+1$th derivative of $s(t)$ is a sequence of derivatives of Diracs. Hence, the classical approach consists in first estimating this sequence and then reconstructing the piecewise polynomial by integration. This implies that we are bound by the difficulties of estimating a sequence of derivatives of Diracs. Further, integration may augment the reconstruction error that was caused by noise

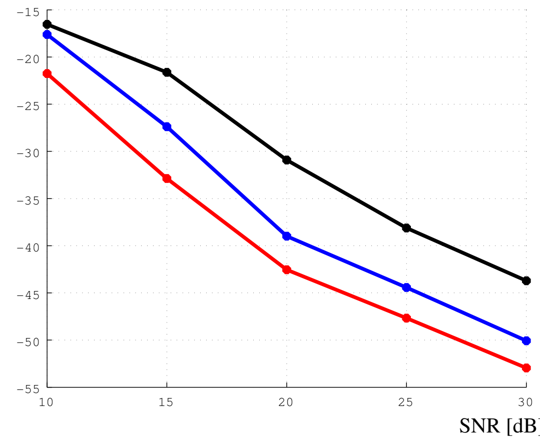


(a) MSE($\boldsymbol{t}$) [dB]

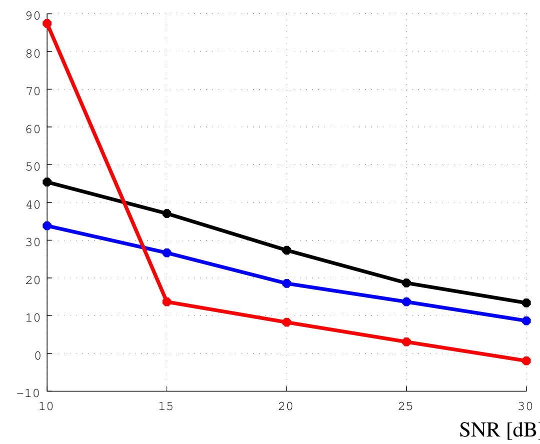


(b) MSE($\boldsymbol{\alpha}$) [dB]

Fig. 4 MSE [dB] of the estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{\alpha}$ of a piecewise polynomial with respect to the number N of measurements. The legends are the same as in Fig. 1.



(a) MSE($\boldsymbol{t}$) [dB]



(b) MSE($\boldsymbol{\alpha}$) [dB]

Fig. 5 MSE [dB] of the estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{\alpha}$ of a piecewise polynomial with respect to the SNR [dB]. The legends are the same as in Fig. 1.

in measurements.

To solve these problems, we propose a direct estimation of the piecewise polynomial, without recasting the problem as the estimation of a sequence of derivatives of Diracs. To this end, we first introduce matrices D and $\tilde{V}_t$ as

$$D = \begin{pmatrix} \left(\frac{\tau}{i2\pi}\mathrm{diag}\left(\frac{1}{-P}, \frac{1}{-P+1}, \ldots, \frac{1}{P}\right)\right)^{R+1} & \begin{matrix}\mathbf{0}\\ 1\\ \mathbf{0}\end{matrix} \end{pmatrix},$$

$$\tilde{V}_t = \begin{pmatrix} V_t & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix},$$

with $\mathbf{0}$ indicating the zero vector. The matrix D is the mapping from Fourier coefficients of the sequence of derivatives of Diracs to those of the piecewise polynomial. The relation of differentiation was exploited here. We further introduce a matrix W_t which maps $\alpha_{k,r}$ to $c_{k,r}$. For instance when $R = 1$ and $K = 2$, W_t is given by

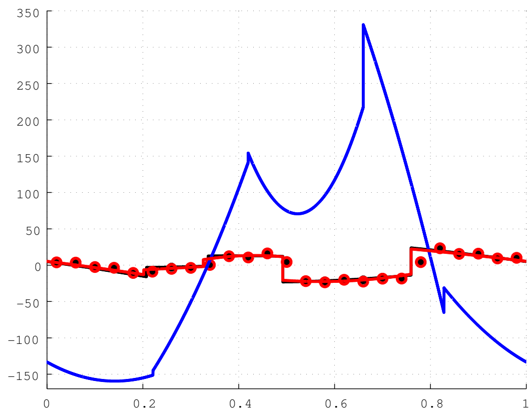
$$W_t = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 1 & -1 & t_0 & -(t_0+\tau) \\ 0 & 0 & -1 & 1 \\ -1 & 1 & -t_1 & t_1 \\ \frac{t_1-t_0}{\tau} & \frac{t_0+\tau-t_1}{\tau} & \frac{t_1^2-t_0^2}{2\tau} & \frac{(t_0+\tau)^2-t_1^2}{2\tau} \end{pmatrix}.$$

We refer [9] for further details on the matrix W_t. By using these matrix, we introduce $\Phi_t = BD\tilde{V}_t W_t$. We then have

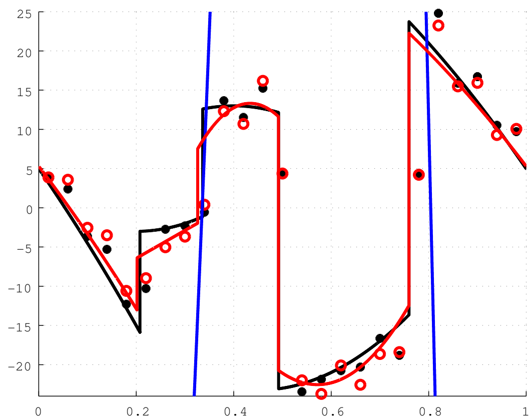
$$\boldsymbol{d} = F^{-1}\Phi_t \boldsymbol{\alpha},$$

where $\boldsymbol{\alpha} = (\alpha_{0,0} \ \cdots \ \alpha_{K-1,R})^{\mathrm{T}}$. Finally, the noiseless measurements of the piecewise polynomial is expressed by using the locations t_k and the coefficients $\alpha_{k,r}$. Because of this expression, the log-likelihood function is defined similarly as in Eq. (11) and its maximization is equivalent to the minimization of $\|\hat{\boldsymbol{y}} - \Phi_t\boldsymbol{\alpha}\|^2$. We find the minimizer of this term by searching $\boldsymbol{t}$ minimizing $\|\hat{\boldsymbol{y}} - \Phi_t\Phi_t^{\dagger}\hat{\boldsymbol{y}}\|^2$, and then calculating $\boldsymbol{\alpha} = \Phi_t^{\dagger}\hat{\boldsymbol{y}}$. The search of the minimizer was again conducted by PSO.

The performance of the proposed method was evaluated by simulations. The target signal is a $\tau = 1$-periodic



(a) Entire view



(b) Enlarged view

Fig. 6 A simulation example with $K = 4$ and $R = 2$. The black line shows the target signal and the red circles and black dots are measurements with and without 20dB noise. The red and blue lines are reconstructed signals by the proposed method and LS with Cadzow denoising, respectively.

piecewise polynomial of degree $R = 1$ with $K = 2$ discontinuities. The unknown parameters are $\boldsymbol{t} = (0.20, 0.65)$ and $\boldsymbol{\alpha} = (\alpha_{0,0}, \alpha_{0,1}, \alpha_{1,0}, \alpha_{1,1}) = (-1.00, -3.00, 2.00, 4.00)$. We reconstructed the signal from 7, 9, …, 15 measurements with 20dB noise. The estimation errors MSE($\boldsymbol{t}$) and MSE($\boldsymbol{\alpha}$) were obtained by averaging $\|\hat{\boldsymbol{t}} - \boldsymbol{t}\|^2$ and $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2$ over 1,000 noise realizations, respectively. The results are shown in Fig. 4, with same legends as in Fig. 1. We can see that the proposed method outperforms the conventional methods in all cases. Fig. 5 shows MSEs in terms of the SNR[dB]. Again, we can see that the proposed method outperforms the conventional methods in all cases except for the case of MSE($\boldsymbol{\alpha}$) in 10dB noise. In this case, the noise gets very large and PSO occasionally behaves unstably.

A simulation example with $K = 4$, $R = 2$, and $N = 25$ is shown in Fig. 6. We can see that the proposed method gives much better results than the classical approach. We should note that $N = 25$ is the minimum for the classical approach and the proposed method can reconstruct the signal from fewer samples. It took 19.12s for the proposed method to reconstruct the signal, while LS with Cadzow denoising required 0.06s only, but Matlab is far from optimal for the implemention of algorithms like PSO, whose potential for parallelization is not exploited at all.

6. Conclusion

We proposed a maximum likelihood estimation method for the recovery of periodic sequences of Diracs and derivatives of Diracs, and periodic piecewise polynomials. The method is able to reconstruct the signals from a number of measurements equal to the number of unknown parameters, while the conventional approaches are not applicable in that case. Simulations results showed that the proposed method outperforms the conventional methods. Unfortunately, when noise level gets large, the proposed method tends to behave unstable. To overcome such difficulty is one of our future works, which also include comparison with other stochastic optimization methods [16], [17], the use of other sampling kernels, and the calculation of the Cramér-Rao bounds for periodic variables [18].

Acknowledgement

The authors are grateful to the anonymous reviewers for their precious comments. This work was supported by JSPS-KAKENHI, Grant-in-Aid for Scientific Research (C), 23500212, 2011. The third author acknowledges a three-months fellowship from the JSPS for his stay in Yamaguchi University.

References

[1] Y. Eldar and G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, Cambridge, 2011.
[2] D. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol.52, no.4, pp.1289 –1306, april 2006.
[3] E. Candes and M. Wakin, “An introduction to compressive sampling,” IEEE Signal Processing Magazine, vol.25, no.2, pp.21–30, March 2008.
[4] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” IEEE Transactions on Signal Processing, vol.50, no.6, pp.1417–1428, June 2002.
[5] P. Dragotti, M. Vetterli, and T. Blu, “Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix,” IEEE Transactions on Signal Processing, vol.55, no.5, pp.1741–1757, May 2007. Part I.
[6] T. Blu, P.L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” IEEE Signal Processing Magazine, vol.25, no.2, pp.31–40, March 2008.
[7] R. Tur, Y. Eldar, and Z. Friedman, “Innovation rate sampling of pulse streams with application to ultrasound imaging,” IEEE Transactions on Signal Processing, vol.59, no.4, pp.1827 –1842, april 2011.
[8] J. Berent, P. Dragotti, and T. Blu, “Sampling piecewise sinusoidal signals with finite rate of innovation methods,” IEEE Transactions on Signal Processing, vol.58, no.2, pp.613–625, February 2010.
[9] A. Hirabayashi, “Sampling and reconstruction of periodic piecewise polynomials using sinc kernel,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E95-A, no.1, pp.322–329, 2012.
[10] J. Cadzow, “Signal enhancement—a composite property mapping algorithm,” IEEE Transactions on Acoustic, Speech, and Signal Processing, vol.36, no.1, pp.49–62, January 1988.
[11] L. Condat, A. Hirabayashi, and Y. Hironaga, “Recovery of nonuniform dirac pulses from noisy linear measurements,” Proceedings of

International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, 2013. (accepted).
[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings of the 1995 IEEE International Conference on Neural Networks, pp.1942 –1948, nov/dec 1995.
[13] A. Papoulis, The Fourier Integral and Its Applications, McGraw Hill, New York, 1962.
[14] I. Markovsky, "Structured low-rank approximation and its applications," Automatica J. IFAC, vol.44, no.4, pp.891–909, 2008.
[15] A. Albert, Regression and the Moore-Penrose Pseudoinverse, Academic Press, New York, 1972.
[16] V.Y.F. Tan and V.K. Goyal, "Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm," IEEE Transactions on Signal Processing, vol.56, no.10, pp.5135–5146, October 2008.
[17] A. Erdozain and P.M. Crespo, "A new stochastic algorithm inspired on genetic algorithms to estimate signals with finite rate of innovation from noisy samples," Signal Processing, vol.90, no.1, pp.134–144, January 2010.
[18] T. Routtenberg and J. Tabrikian, "Periodic CRB for non-Bayesian parameter estimation," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.2448–2451, 2011.

Akira Hirabayashi received his B.E., M.E., and D.E. degrees in Computer Science in 1993, 1995, and 1999, respectively, from the Tokyo Institute of Technology, Tokyo, Japan. After working as a research associate of Tokyo Institute of Technology from 1995 to 2000 and a lecturer and an associate professor of Yamaguchi University, Yamaguchi, Japan, from 2000 to 2013, he is now a professor of the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. From 2004 to 2005, he was an invited professor at the Biomedical Imaging Group, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland. From 2009 to 2010, he was a visiting professor at the Communications and Signal Processing Group, Imperial College London, London, U.K. His research interests include sampling theory, optical measurement, and signal and image processing. Dr. Hirabayashi is also a member of the Institute of Electrical and Electronics Engineers (IEEE), and the Society of Instrument and Control Engineers (SICE), Japan.

Yosuke Hironaga received his B.E. in Computer Science in 2012 from Yamaguchi University. He is now pursuing the M.E. in the Graduate School of Medicine, Yamaguchi University, Yamaguchi, Japan. His research interests include sparse sampling and its applications to signal/image processing.

Laurent Condat received the M.S. degree in computer science from the Ecole Nationale Superieure d'Informatique et de Mathematiques Appliquees de Grenoble (ENSIMAG), th M.S. degree in applied mathematica from the Joseph Fourier University, Grenoble, France, in 2003, and the PhD degree in applied mathematics from the Grenoble Institute of Technology (INPG) in 2006. From 2006 to 2008, he was a postdoctorate in the team MAMEBA of the Helmholtz Zentrum Mnchen, Munich, Germany. Since 2008, he has been a CNRS research fellow in the GREYC lab, Caen, France. After 2012, he is working at GIPSA-lab, in Grenoble, France. In 2012, he stayed in Yamaguchi University in for three months. His area of interest includes sampling theory and inverse problems of image processing.