SAMPLING AND RECOVERY OF CONTINUOUS SPARSE SIGNALS BY MAXIMUM LIKELIHOOD ESTIMATION

Akira Hirabayashi, Yosuke Hironaga[*]

Yamaguchi University
Ube, Japan

Laurent Condat

GIPSA-lab & CNRS
Grenoble, France

ABSTRACT

We propose a maximum likelihood estimation approach for the recovery of continuously-defined sparse signals from noisy measurements, in particular periodic sequences of derivatives of Diracs and piecewise polynomials. The conventional approach for this problem is based on total-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. Further on, Cadzow denoising does not guarantee any optimality. The proposed parametric approach solves all of these problems. Since the corresponding log-likelihood function is non-convex, we exploit the stochastic method of particle swarm optimization (PSO) to find the global solution. Simulation results confirm the effectiveness of the proposed approach, for a reasonable computational cost.

Index Terms— signals with finite rate of innovation, derivative of Diracs, piecewise polynomials, maximum likelihood estimation, Cadzow denoising

1. INTRODUCTION

Sampling sparse signals defined in continuous or discrete domain is attracting great interest, as can be seen from the huge amount of publications on the topic, see e.g. [1, 2, 3, 4, 5, 6, 7]. The reconstruction of sequences of Dirac distributions (Diracs, in short) lies at the heart of the theories formulated for analog signals, 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 [1, 8].

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

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

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 [1] or a sum-of-sincs [4]. Then, the sequence can be perfectly reconstructed from the noiseless measurements using the annihilating filter technique [1]. This technique, however, requires at least $2\tilde{K}+1$ measurements, which is more than the number of unknown parameters $K+\tilde{K}$. Having the minimum possible number of samples while maintaining a possible reconstruction is crucial, for instance if each measurement is very costly (financially or in time). If the measurements are corrupted by noise, the annihilating filter approach, a.k.a. total least squares (TLS), whose detailed description is in [3] yields $\tilde{K}$ instead of K locations. Also, this method does not give satisfactory results, so that preprocessing is necessary. For that, Cadzow denoising [9] is the standard approach [3]; it is easy to implement but does not guarantee any optimality.

To solve these problems, we propose a method that reconstructs the signal using maximum likelihood estimation, as is in [10, 11]. The corresponding likelihood function is non-convex. Hence, to find the global solution, we exploit a heuristic approach called *particle swarm optimization* (PSO) [12]. 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.

This paper is organized as follows. In Section 2 we describe the sampling setup using the sum-of-sincs kernel and formulate maximum likelihood reconstruction of the sequence of derivatives of Diracs. Section 3 extends our approach to periodic piecewise polynomials.

2. SEQUENCE OF DERIVATIVE OF DIRACS

The τ-periodic sequence of derivatives of Diracs $s(t)$ is sampled using a kernel $\psi(t)$ and yields N noiseless measurements

[*]This work was supported by JSPS-KAKENHI, Grant-in-Aid for Scientific Research (C), 23500212, 2011.The third author thanks to a three-months fellowship from the JSPS for his stay in Yamaguchi University.

$d_n = \langle s, \psi_n \rangle = \int_{-\infty}^{\infty} f(t)\overline{\psi(t - nT)}dt$, for $n = 0, \ldots, N-1$ and $T = \tau/N$. We adopt for $\psi(t)$ the *sum of sincs* (in the Fourier domain) kernel [4], which is defined in time domain by

$$\psi(t) = \frac{\mathrm{rect}(t/\tau)}{\tau} \sum_{p=-P}^{P} b_p e^{i2p\pi t/\tau}, \tag{1}$$

where $\mathrm{rect}(t) = 1$ if $|t| \leq 0.5$ else 0 and $P \leq (N-1)/2$ 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. Then, it follows from (1) that

$$d_n = \sum_{p=-P}^{P} b_p \hat{d}_p e^{i2pn\pi/N}. \tag{2}$$

This admits the matrix representation $\boldsymbol{d} = F^{-1}B\hat{\boldsymbol{d}}$, where B is the diagonal matrix $\mathrm{diag}(b_{-P}, \ldots, b_P)$ and F is the discrete Fourier transform (DFT) matrix, defined accordingly.

We can derive the Fourier coefficients of $s(t)$ as $\hat{d}_p = \sum_{k=0}^{K-1}\sum_{r=0}^{R_k-1} \tilde{c}_{k,r} p u_k^p$, where $u_k = e^{-i2\pi t_k/\tau}$, $\tilde{c}_{k,r} = (i2\pi)^r c_{k,r}/\tau^{r+1}$. Let U_t and $\boldsymbol{c}$ be the matrix and the vector defined respectively as

$$\begin{aligned} U_t &= \begin{pmatrix} u_0^{-P} & \cdots & (-P)^R u_{K-1}^{-P} \\ u_0^{-P+1} & \cdots & (-P+1)^R u_{K-1}^{-P+1} \\ \vdots & \ddots & \vdots \\ u_0^{P} & \cdots & (P)^R u_{K-1}^{P} \end{pmatrix}, \\ \boldsymbol{c} &= (\tilde{c}_{0,0}\ \tilde{c}_{0,1}\ \cdots\ \tilde{c}_{K-1,R-1})^T. \end{aligned}$$

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

$$\boldsymbol{d} = F^{-1}BU_t\boldsymbol{c}. \tag{3}$$

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 $n = 0, \ldots, N-1$. We have to estimate the unknowns $\{t_k\}_{k=0}^{K-1}$ and $\{c_{k,r}\}_{k=0,r=0}^{K-1 R_k}$ as precisely as possible from the data $\{y_n\}_{n=0}^{N-1}$. To this end, we exploit the formalism of maximum likelihood estimation. Let $\boldsymbol{y}$ and $\boldsymbol{e}$ be vectors whose n-th elements are y_n and e_n, respectively: $\boldsymbol{y} = \boldsymbol{d} + \boldsymbol{e}$. Assume that the probability density function $p(\boldsymbol{e})$ is known. Then using (3), we can define the log-likelihood function as $L(\boldsymbol{t}, \boldsymbol{c}) = \log p(\boldsymbol{y} - F^{-1}BU_t\boldsymbol{c})$, where $\boldsymbol{t} = [t_0\, t_1\, \cdots\, t_{K-1}]^{\mathrm{T}}$. The most standard model of $p(\boldsymbol{e})$, which we also adopt in this paper, 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

$$L(\boldsymbol{t}, \boldsymbol{c}) = -\frac{\|\boldsymbol{y} - F^{-1}BU_t\boldsymbol{c}\|^2}{2\sigma^2} + \text{Constant}. \tag{4}$$

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

$$f_o(\boldsymbol{t}, \boldsymbol{c}) = \|\hat{\boldsymbol{y}} - BU_t\boldsymbol{c}\|^2, \tag{5}$$

where $\hat{\boldsymbol{y}} = F\boldsymbol{y}$. Thus, 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.

Eqn. (5) 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 [13]. 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 $\boldsymbol{c} = (BU_t)^{\dagger}\hat{\boldsymbol{y}}$.

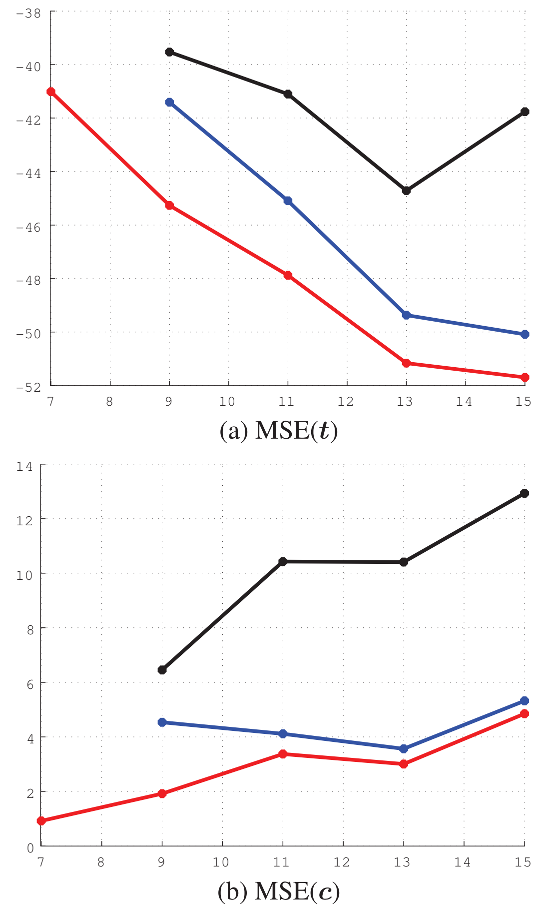
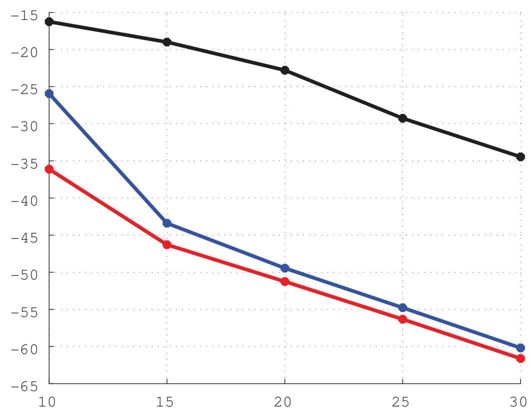
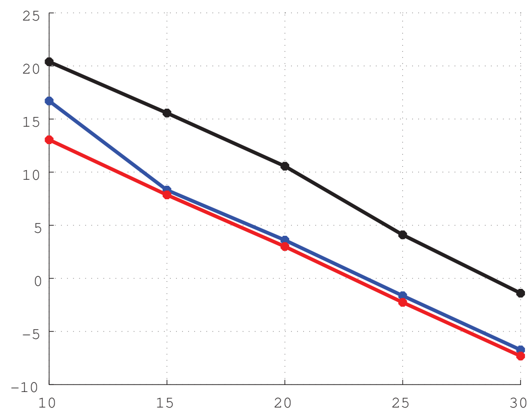


Fig. 1. Mean square errors (MSE) of estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{c}$ with respect to the number of measurements with 20dB noise. The red, blue and black lines show the results by the proposed method, by TLS with and without Cadzow denoising, respectively.



(a) MSE($\boldsymbol{t}$)



(b) MSE($\boldsymbol{c}$)

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

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

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 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. The proposed method requires a number of measurements more than or equal to $K + \tilde{K} + 1 = 7$, while the conventional method needs at least $2\tilde{K} + 1 = 9$ measurements. To see this difference, we reconstructed the sequence of derivative of Diracs from various numbers of measurements, from 7, to 15. The noise level was chosen so that the SNR is 20dB[1]. 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) 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, see the caption for details. We can see that the proposed method outperforms the conventional methods for every number of measurements. Note that the TLS approach cannot be applied to the case of seven measurements, while the proposed method performs the best in this case for the estimation of $\boldsymbol{c}$. In Fig. 2, we show the behavior of the MSE with respect to the SNR; here, the number of measurements is fixed to $N = 13$. Again, the proposed method performs better than the conventional approaches, whatever the SNR.

3. PERIODIC PIECEWISE POLYNOMIALS

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

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

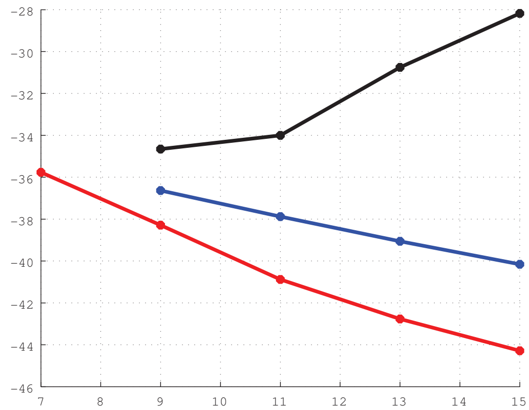
and the function $\varphi_{K-1}(t)$ as

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

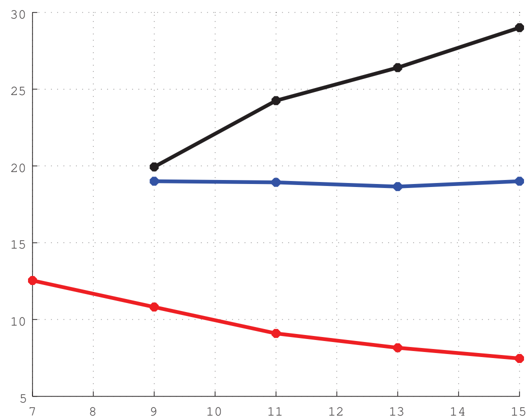
where $v_k(t) = \sum_{r=0}^{R} \alpha_{k,r} t^r$. Then, a τ-periodic piecewise polynomial $s(t)$ of degree R is defined by $s(t) = \sum_{k' \in \mathbb{Z}} s_0(t - k'\tau)$, with $s_0(t) = \sum_{k=0}^{K-1} \varphi_k(t)$. The available samples are inner products $\langle s, \psi_n \rangle$ corrupted by noise. Eqn. (2) still holds for this kind of signal.

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. In this section, we show how to directly estimate the piecewise polynomial, without recasting the problem as the estimation of a sequence of derivatives of Diracs. Let us introduce the vector $\boldsymbol{\alpha} = [\alpha_{0,0} \ \cdots \ \alpha_{K-1,R}]^{\mathrm{T}}$

[1]The SNR is defined by $10 \log_{10} \frac{\sum_{n=1}^{N} {d_n}^2}{\sigma^2 N}$.



(a) MSE($\boldsymbol{t}$)



(b) MSE($\boldsymbol{\alpha}$)

Fig. 3. MSE of estimated parameters for $\boldsymbol{t}$ and $\boldsymbol{\alpha}$. The legends are the same as in Fig. 1.

and the matrix $\Phi_t = BDV_tW_t$, where

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

$V_t = \begin{pmatrix} U_t & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}$, $\omega = (\tau/i2\pi)^{R+1}$, $\mathbf{0}$ denotes the zero vector, and W_t is a mapping from $\boldsymbol{\alpha}$ to $\boldsymbol{c}$ (see [14] for details). Note that the relation of differentiation was implicitly used in these formulas. We then have $\boldsymbol{d} = F^{-1}\Phi_t\boldsymbol{\alpha}$. Therefore, the log-likelihood function is defined similarly as in (4) 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 piecewise polynomial of degree $R = 1$ with $K = 2$ discontinuities. The unknown parameters are $\boldsymbol{t} = (0.20, 0.65)$ and $\boldsymbol{\alpha} = (\alpha_{0,0}, \alpha_{0,1}, \alpha_{1,0}, \alpha_{1,1}) = (-1.00, -3.00, 2.00, 4.00)$.

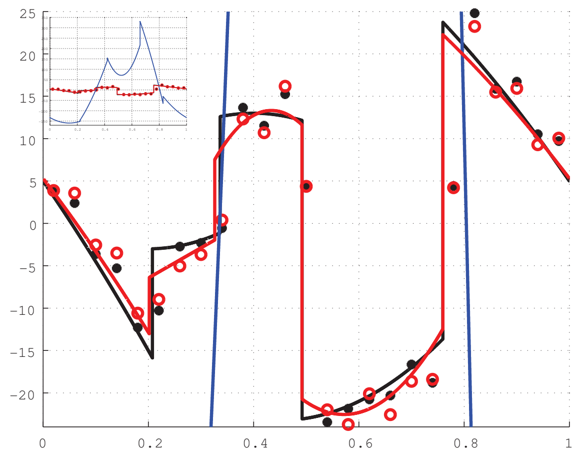


Fig. 4. A simulation example with $K = 4$ and $R = 2$. The black line shows the target signal and the red circles and black dots are measurements with and without 20dB noise. The red and blue lines are reconstructed signals by the proposed method and TLS with Cadzow denoising, respectively. The small figure in the top left corner shows the entire shape of the blue line.

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. 3, with the same legends as in Fig. 1. We can see that the proposed method outperforms the conventional methods in all cases. A simulation example with $K = 4$, $R = 2$, and $N = 25$ is shown in Fig. 4. 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 TLS 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.

4. CONCLUSION

We proposed a maximum likelihood estimation method for the recovery of periodic sequences of derivatives of Diracs and 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. Future work includes further performance analyses of the proposed method, comparison with other stochastic optimization methods [15, 16] and the calculation of the Cramér-Rao bounds for periodic case [17]. A Matlab implementation of the proposed method will be available online at http://sip.csse.yamaguchi-u.ac.jp.

5. REFERENCES

[1] 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.

[2] P.-L. 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.

[3] 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.

[4] R. Tur, Y. C. 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.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[6] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[7] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, 2011.

[8] J. Berent, P.-L. 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] J. A. 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.

[10] S. Umesh and D.W. Tufts, "Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to NMR spectroscopy data," *IEEE Transactions on Signal Processing*, vol. 44, no. 9, pp. 2245–2259, September 1996.

[11] A. Hirabayashi, T. Iwami, S. Maeda, and Y. Hironaga, " Reconstruction of the sequence of Diracs from noisy samples via maximum likelihood estimation, " in Proc. of IEEE ICASSP, 2012, pp. 3805—3808.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the 1995 IEEE International Conference on Neural Networks*, Nov/Dec 1995, vol. 4, pp. 1942 –1948.

[13] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York, 1972.

[14] 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, January 2012.

[15] 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.

[16] 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.

[17] T. Routtenberg and J. Tabrikian, "Non-Bayesian Periodic Cramr-Rao Bound," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 1019–1032, February 2013.