RECOVERY OF NONUNIFORM DIRAC PULSES FROM NOISY LINEAR MEASUREMENTS

L. Condat[*]

GIPSA-lab, CNRS–University of Grenoble
Grenoble, France

A. Hirabayashi, Y. Hironaga

Yamaguchi University
Ube, Japan

Contact: Laurent.Condat“at”gipsa-lab.grenoble-inp.fr

ABSTRACT

We consider the recovery of a finite stream of Dirac pulses at nonuniform locations, from noisy lowpass-filtered samples. We show that maximum-likelihood estimation of the unknown parameters can be formulated as structured low rank approximation of an appropriate matrix. To solve this difficult, believed NP-hard, problem, we propose a new heuristic iterative algorithm, based on a recently proposed splitting method for convex nonsmooth optimization. Although the algorithm comes, in absence of convexity, with no convergence proof, it converges in practice to a local solution, and even to the global solution of the problem, when the noise level is not too high. It is also fast and easy to implement.

Index Terms— Recovery of Dirac pulses, finite rate of innovation, maximum likelihood estimation, structured low rank approximation, optimization, Cadzow denoising

1. INTRODUCTION AND PROBLEM FORMULATION

Reconstruction of signals lying in shift-invariant spaces, including bandlimited signals and splines, has received long attention in sampling theory. Recently, analog reconstruction from discrete samples has been enlarged to a broader class of signals, with so-called *finite rate of innovation* (FRI), beyond the classical framework rooted in Shannon's work [1, 2, 3].

In this paper, we focus on the retrieval of a finite stream of Dirac pulses from uniform, noisy, lowpass-filtered samples, a problem at the heart of the FRI theory [4, 1, 5, 6]. More precisely, the sought-after unknown signal s consists of K Dirac pulses in the finite interval $[0, \tau[$, where the real $\tau > 0$ and the integer $K \geq 1$ are known; that is

$$s(t) = \sum_{k=1}^{K} a_k \delta(t - t_k), \quad \forall t \in [0, \tau[, \tag{1}$$

where $\delta(t)$ is the Dirac mass distribution, $\{t_k\}_{k=1}^K$ are the unknown distinct locations in $[0, \tau[$, and $\{a_k\}_{k=1}^K$ are the unknown real nonzero amplitudes. The goal is to obtain estimates of these $2K$ values, which forms a deterministic (non-Bayesian) parametric estimation problem. The available data are, classically, linear uniform noisy measurements $\{v_n\}_{n=0}^{N-1}$ on s, of the form

$$v_n = \int_0^\tau s(t)\varphi\Big(\frac{n\tau}{N} - t\Big)dt + \varepsilon_n \tag{2}$$

$$= \sum_{k=1}^{K} a_k \varphi\Big(\frac{n\tau}{N} - t_k\Big) + \varepsilon_n, \quad \forall n = 0, \ldots, N-1, \tag{3}$$

where $\varphi(t)$ is the sampling function and the $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$ are independent random realizations of Gaussian noise. Note that other noise models could be considered as well, by changing the cost function in eqns. (5), (10), (14) below.

The questions of the choice of the function φ and of the number N of measurements allowing perfect reconstruction, in absence of noise, has been addressed in the literature [5, 6, 7]. In a nutshell, the condition $N \geq 2K + 1$, which we hereafter assume to be true, is necessary and sufficient, provided that φ satisfies some constraints in Fourier domain. In this work, as the emphasis is on appropriately handling the presence of noise, we adopt the simplest choice of the Dirichlet sampling function (which amounts to periodizing the signal s on the real line before sampling it with the sinc function):

$$\varphi(t) = \frac{\sin(N\pi t/\tau)}{N \sin(\pi t/\tau)} = \frac{1}{N} \sum_{m=-M}^{M} e^{j2\pi m t/\tau}, \quad \forall t \in \mathbb{R}, \tag{4}$$

where hereafter we assume, without loss of generality and only to simplify the notations, that N is odd of the form $N = 2M + 1$. The extension of the setting to the reconstruction of pulses with real shape, instead of the ideal Dirac distribution, is of obvious practical interest in ultrawideband communications [2] or to detect impulsive signals in biomedical applications [5]. This extension, or equivalently the choice of another sampling function φ, can be done without difficulty, as detailed in [5].

[*]The first author performed this work while at Yamaguchi University, invited by the second author, thanks to a three-months fellowship of the Japanese Society for the Promotion of Science.

The paper is organized as follows. In Sect. 2, we formulate the maximum likelihood estimation problem and in Sect. 3, we show that it amounts to a low rank matrix approximation problem. The new algorithm to solve it is presented in Sect. 4.

2. MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS

A natural approach to solve parametric estimation problem is maximum likelihood (ML) estimation; it consists in selecting the model which is the most likely to explain the observed noisy data. In our case, since we have supposed the noise to be Gaussian, this corresponds to solving the nonlinear least-squares problem [8]:

$$\text{Find } (\tilde{\mathbf{t}}, \tilde{\mathbf{a}}) = \underset{\mathbf{t}' \in [0,\tau[^K, \mathbf{a}' \in \mathbb{R}^K}{\arg\min} \sum_{n=0}^{N-1} \left| v_n - \sum_{k=1}^{K} a'_k \varphi\Big(\frac{n\tau}{N} - t'_k\Big) \right|^2. \tag{5}$$

Almost surely, the solution to this problem is unique, the obtained amplitudes $\tilde{\mathbf{a}} = \{\tilde{a}_k\}_{k=1}^K$ are nonzero, and the obtained locations $\tilde{\mathbf{t}} = \{\tilde{t}_k\}_{k=1}^K$ are distinct.

Now, applying the discrete Fourier transform to the vector of measurements yields the Fourier coefficients defined by

$$\hat{v}_m = \sum_{n=0}^{N-1} v_n e^{-j2\pi mn/N}, \quad \forall m = -M, \dots, M. \tag{6}$$

We define the Fourier coefficients $\{\hat{\varepsilon}_m\}_{m=-M}^M$ similarly. Combining (3) and (4), we get, for every $n = 0, \dots, N-1$,

$$v_n - \varepsilon_n = \frac{1}{N} \sum_{k=1}^{K} a_k \sum_{m=-M}^{M} e^{j2\pi m(n/N - t_k/\tau)} \tag{7}$$

$$= \frac{1}{N} \sum_{m=-M}^{M} e^{j2\pi mn/N} \left(\sum_{k=1}^{K} a_k e^{-j2\pi m t_k/\tau} \right). \tag{8}$$

We recognize the form of the inverse discrete Fourier transform. Thus, by identification, we obtain

$$\hat{v}_m = \sum_{k=1}^{K} a_k e^{-j2\pi m t_k/\tau} + \hat{\varepsilon}_m, \quad \forall m = -M, \dots, M. \tag{9}$$

Since the inverse discrete Fourier transform is unitary, up to a constant, the problem (9) can be rewritten as [8]:

$$\text{Find } (\tilde{\mathbf{t}}, \tilde{\mathbf{a}}) = \underset{\mathbf{t}' \in [0,\tau[^K, \mathbf{a}' \in \mathbb{R}^K}{\arg\min} \sum_{m=-M}^{M} \left| \hat{v}_m - \sum_{k=1}^{K} a'_k e^{-j2\pi m t'_k/\tau} \right|^2. \tag{10}$$

We remark that (10) takes the form of a spectral estimation problem, which consists in retrieving the parameters of a sum of complex exponentials from noisy samples [9]. The optimal statistical properties of ML estimation for spectral estimation are well known [10, 11]. However, solving the problem (10) is a difficult task, as the cost function has a multimodal shape with many local minima [12, 13]. Stochastic optimization approaches can be applied [14, 15, 8], but due to the combinatorial nature of the problem, their computational cost blows up even for moderate values of K and N. Several methods have been proposed to find a local minimum of the cost function in (10), e.g. [16, 17, 18]. They proceed by iteratively refining an initial estimate of the solution, which has to be already of good quality. A major advantage of the approach developed in this paper is that it does not necessitate any initialization. We also note that, when $N \gg K$ and the locations t_k are not too close to each other, classical spectral estimation techniques like MUSIC [19] and ESPRIT [20] can be used; they are fast but statistically suboptimal. In this work, we investigate our problem in its whole generality, without any simplifying assumption.

3. THE ANNIHILATION PROPERTY: REFORMULATION OF EQN. (10) AS A MATRIX APPROXIMATION PROBLEM

Let us assume temporarily that there is no noise, i.e. $\hat{\varepsilon}_m = 0$ in (9). Then, the sequence of Fourier coefficients $\{\hat{v}_m\}_{m=-M}^M$ can be *annihilated* [21]; that is, its convolution with the sequence $\{h_k\}_{k=0}^K$ is identically zero:

$$\sum_{k=0}^{K} h_k \hat{v}_{m-k} = 0, \quad \forall m = -M+K, \dots, M, \tag{11}$$

where the Z-transform of the annihilating filter is defined, up to a constant, as

$$H(z^{-1}) = \sum_{k=0}^{K} h_k z^k = \prod_{k=1}^{K} (z - e^{j2\pi t_k/\tau}). \tag{12}$$

In matrix form, the annihilation property is

$$\underbrace{\begin{pmatrix} \hat{v}_{-M+K} & \cdots & \hat{v}_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \hat{v}_M & \cdots & \hat{v}_{M-K} \end{pmatrix}}_{\mathbf{T}_K} \begin{pmatrix} h_0 \\ \vdots \\ h_K \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{13}$$

Let us define, for every integer $P = K, \dots, M$, the Toeplitz—i.e. with constant values along its diagonals—matrix $\mathbf{T}_P$, of size $N - P \times P + 1$, obtained by arranging the values $\{\hat{v}_m\}_{m=-M}^M$ in its first row and column; $\mathbf{T}_K$ is depicted in (13). Then, the existence of an annihilating filter of size $K+1$ for the sequence $\{\hat{v}_m\}_{m=-M}^M$ is completely equivalent to the property that $\mathbf{T}_P$ has rank at most K, for every $P = K, \dots, M$.

Hence, turning back to the case when noise is present in the data, we can recast the estimation problem (10) as the following matrix approximation problem:

$$\text{Find } \widetilde{\mathbf{T}}_P \in \underset{\mathbf{T}' \in \mathbb{C}^{N-P\times P+1}}{\arg\min} \|\mathbf{T}' - \mathbf{T}_P\|_w^2 \quad \text{s. t.} \quad \mathbf{T}' \text{ is Toeplitz and } \operatorname{rank}(\mathbf{T}') \leq K, \tag{14}$$

for some chosen $P \in K, \ldots, M$, where the weighted Frobenius norm of a matrix $\mathbf{A} = \{a_{i,j}\} \in \mathbb{C}^{N-P\times P+1}$ is defined by

$$\|\mathbf{A}\|_w^2 = \sum_{i=1}^{N-P} \sum_{j=1}^{P+1} w_{i,j} |a_{i,j}|^2, \tag{15}$$

where $w_{i,j}$ is the inverse of the size of the diagonal going through the position (i, j):

$$w_{i,j} = \begin{cases} 1/(i-j+P+1) \text{ if } i-j \leq 0, \\ 1/(P+1) \text{ if } 1 \leq i-j \leq N-2P-1, \\ 1/(j-i+N-P) \text{ if } i-j \geq N-2P. \end{cases} \tag{16}$$

After the *structured low rank approximation* (SLRA) [22] problem (14) has been solved, the procedure to recover the estimates of the parameters is the following [1]. First, reshape the obtained Toeplitz matrix $\widetilde{\mathbf{T}}_P$ to a Toeplitz matrix $\widetilde{\mathbf{T}}_K$ of size $N-K \times K+1$; that is,

$$\widetilde{\mathbf{T}}_K = \begin{pmatrix} \tilde{v}_{-M+K} & \cdots & \tilde{v}_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \tilde{v}_M & \cdots & \tilde{v}_{M-K} \end{pmatrix}. \tag{17}$$

Second, compute the right singular vector $\tilde{\mathbf{h}} = \{\tilde{h}_k\}_{k=0}^K$ of $\widetilde{\mathbf{T}}_K$ corresponding to the singular value 0. Since, almost surely, $\widetilde{\mathbf{T}}_K$ has rank exactly K, $\tilde{\mathbf{h}}$ is unique, up to a constant. Third, compute the roots $\{\tilde{z}_k\}_{k=1}^K$ of the polynomial $\sum_{k=0}^K \tilde{h}_k z^k$; the estimates $\{\tilde{t}_k\}_{k=1}^K$ of the locations are given by

$$\tilde{t}_k = \frac{\tau}{2\pi} \arg_{[0,2\pi[}(\tilde{z}_k), \quad \forall k = 1, \ldots, K. \tag{18}$$

Fourth, given the estimates $\{\tilde{t}_k\}_{k=1}^K$, the ML estimates $\{\tilde{a}_k\}_{k=1}^K$ of the amplitudes are obtained by solving the linear system

$$\widetilde{\mathbf{U}}^{\mathrm{H}} \widetilde{\mathbf{U}} \tilde{\mathbf{a}} = \widetilde{\mathbf{U}}^{\mathrm{H}} \hat{\mathbf{v}}, \tag{19}$$

where $\hat{\mathbf{v}} = [\hat{v}_{-M} \ \cdots \ \hat{v}_M]^{\mathrm{T}}$, $\cdot^{\mathrm{H}}$ denotes the Hermitian transpose, and

$$\widetilde{\mathbf{U}} = \begin{pmatrix} e^{j2\pi M \tilde{t}_1/\tau} & \ldots & e^{j2\pi M \tilde{t}_K/\tau} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi M \tilde{t}_1/\tau} & \ldots & e^{-j2\pi M \tilde{t}_K/\tau} \end{pmatrix}. \tag{20}$$

Thus, the process consists in *denoising* the matrix $\mathbf{T}_P$, or equivalently the data $\{v_n\}_{n=0}^{N-1}$, by finding the closest matrix consistent with the model's structure, from which the parameters are estimated. In absence of noise, the parameters are perfectly recovered. The SLRA problem (14), which consists in projecting a matrix in the intersection of a linear subspace and a nonconvex manifold, is believed NP-hard [23, 24]. So, at first glance, we just have replaced the difficult problem (10) by the SLRA problem of same difficulty. However, there are several advantages with the latter. First, the estimation of the locations is decoupled from that of the amplitudes. Second, the initialization problem disappears: an iterative algorithm to solve (14) proceeds directly, with the noisy matrix $\mathbf{T}_P$ as initial estimate of the solution $\widetilde{\mathbf{T}}_P$; moreover, for a low noise level, an algorithm converging to a local solution will find the global solution $\widetilde{\mathbf{T}}_P$, as $\mathbf{T}_P$, $\widetilde{\mathbf{T}}_P$ and the true noiseless matrix will correspond to the same catchment area of the cost function in (14).

We note that the obtained estimates $\{(\tilde{t}_k, \tilde{a}_k)\}_{k=1}^K$ coincide with the ML estimates, solution to (10), only if the roots $\{\tilde{z}_k\}_{k=1}^K$ are distinct and all on the complex unit circle. This is the case in practice, with high probability, if the noise level is not too high. Indeed, the matrices $\mathbf{T}_P$ and $\widetilde{\mathbf{T}}_P$ are centro-Hermitian, i.e. their entries satisfy $\hat{v}_{-m} = \hat{v}_m^*$ and $\tilde{v}_{-m} = \tilde{v}_m^*$, for every m, where $\cdot^*$ denotes complex conjugation. Consequently, the polynomial $\sum_{k=0}^K \tilde{h}_k z^k$ is self-inversive [25], so that its roots $\{\tilde{z}_k\}_{k=1}^K$ are either on the complex unit circle or come by pairs with same complex phase and opposite amplitudes. In essence, when moving continuously the variables $\{v_n\}_{n=0}^{N-1}$ from their noiseless version to their actual noisy version, the corresponding estimated roots $\{\tilde{z}_k = e^{j2\pi\tilde{t}_k/\tau}\}_{k=1}^K$ deviate continuously from the true roots $\{z_k = e^{j2\pi t_k/\tau}\}_{k=1}^K$, while remaining on the complex unit circle; only if the perturbation is large enough, two distinct roots $(\tilde{z}_k, \tilde{z}_{k'})$ will possibly merge and then split in a pair $(\tilde{z}_k, \tilde{z}_{k'} = 1/\tilde{z}_k^*)$ on both sides of the unit circle, yielding $\tilde{t}_k = \tilde{t}_{k'}$.

We now briefly tackle the state of the art for solving SLRA problems, which have a wide range of applications [22, 26]. To our knowledge, the only publicly available software package for SLRA is the one currently in development by I. Markovsky [27]. However, it only handles real-valued, and not complex-valued, matrices. On the other hand, the popular heuristic *Cadzow denoising* method [28] is promoted in [1, 5] for our problem of recovering Dirac pulses. This algorithm denoises the matrix $\mathbf{T}_P$ iteratively: at each iteration, the matrix is replaced by its closest, in Frobenius norm, matrix of rank at most K, and then the obtained matrix is replaced by its closest Toeplitz matrix. If the algorithm converges, which is not guaranteed [29], it does so to a Toeplitz matrix of rank at most K, which is not even a local minimizer of the cost function $\|\cdot - \mathbf{T}_P\|_w^2$ in (14) [12, 13]. In the next section, we propose a new algorithm to compute a local solution of the SLRA problem (14), which thus improves upon Cadzow denoising, for essentially the same complexity.

4. A NEW ITERATIVE OPTIMIZATION METHOD FOR SLRA

We consider the generic optimization problem:

$$\text{Find } \tilde{x} \in \arg\min_{x \in \mathcal{H}} F(x) \quad \text{s.t.} \quad x \in \Omega_1 \cap \Omega_2, \tag{21}$$

where $\mathcal{H}$ is a real Hilbert space of finite dimension, Ω_1 and Ω_2 are two closed subsets of $\mathcal{H}$, and $F : \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with β-Lipschitz continuous gradient, for some $\beta > 0$; that is, $\|\nabla F(x') - \nabla F(x)\| \leq \beta \|x - x'\|$, $\forall x, x' \in \mathcal{H}$.

We denote by $P_\Omega : \mathcal{H} \to \Omega$ the closest-point projection onto a closed set $\Omega \subset \mathcal{H}$; that is, for every $x \in \mathcal{H}$, $P_\Omega(x) \in \arg\min_{x' \in \Omega} \|x - x'\|$. If Ω is convex, the minimizer in this definition is unique. The proposed algorithm to solve (21) is the following:

Proposed algorithm. Choose the parameters $\mu > 0$, $\gamma \in]0, 1[$, and the initial estimates $x^{(0)}, s^{(0)} \in \mathcal{H}$. Then iterate, for every $i \geq 0$,

$$\left| \begin{array}{l} x^{(i+1)} = P_{\Omega_1}\big(s^{(i)} + \gamma(x^{(i)} - s^{(i)}) - \mu \nabla F(x^{(i)})\big) \\ s^{(i+1)} = s^{(i)} - x^{(i+1)} + P_{\Omega_2}(2x^{(i+1)} - s^{(i)}) \end{array} \right. .$$

The convergence result is a corollary of more general results derived in [30]:

Theorem 1. *In* (21)*, suppose that* (i) *a solution exists;* (ii) *the sets* Ω_1 *and* Ω_2 *are convex;* (iii) $\mathrm{ri}(\Omega_1) \cap \mathrm{ri}(\Omega_2) \neq \emptyset$*, where* ri *denotes the relative interior. In the proposed algorithm, suppose that (iv)* $2\gamma > \beta\mu$*. Then, the sequence* $(x^{(i)})_{i \in \mathbb{N}}$ *converges to some element* $\tilde{x}$ *solution to the problem* (21)*.*

In absence of convexity, this result does not apply, so that we will use the method as a heuristic. The SLRA problem (14) can be recast as an instance of (21) as follows: $\mathcal{H} = \mathbb{C}^{N-P \times P+1}$ is the real Hilbert space of complex-valued matrices of size $N - P \times P + 1$ with centro-Hermitian symmetry, endowed with Frobenius inner product $\langle \mathbf{X}, \mathbf{X}' \rangle = \sum_{i,j} x_{i,j} x'^{*}_{i,j}$; Ω_1 is the closed nonconvex subset of $\mathcal{H}$ of matrices with rank at most K; P_{Ω_1} corresponds to SVD truncation, according to the classical Schmidt-Eckart-Young theorem [31, theorem 2.5.3]: if a matrix $\mathbf{X}$ has SVD $\mathbf{X} = \mathbf{L}\mathbf{\Sigma}\mathbf{R}^{\mathrm{H}}$, then $P_{\Omega_1}(\mathbf{X})$ is obtained by setting to zero the singular values in $\mathbf{\Sigma}$, except the K largest; Ω_2 is the linear subspace of $\mathcal{H}$ of Toeplitz matrices; P_{Ω_2} simply consists in averaging along the diagonals of the matrix; $F(\mathbf{X}) = \frac{1}{2}\|\mathbf{X} - \mathbf{T}_P\|_w^2$, with $\nabla F(\mathbf{X}) = \mathbf{W} \circ (\mathbf{X} - \mathbf{T}_P)$, where $\circ$ is the entrywise (Hadamard) product and the entries $\{w_{i,j}\}$ of the matrix $\mathbf{W}$ are defined in (16), with Lipschitz constant $\beta = \max(\{w_{i,j}\}) = 1$.

The proposed algorithm almost always converges: only when the noise level is very high, it possibly gets trapped in a cycle; but in that case, running again the algorithm with lower values of μ and γ seems sufficient to obtain convergence. When the algorithm converges, it does so to a matrix in $\Omega_1 \cap \Omega_2$, which is a local minimizer of the cost function F. In comparison with Cadzow denoising, we obtain estimation errors for the locations of the Dirac pulses, which are about 10% lower in average. An example of experimental setting with corresponding results is depicted in Fig. 1. More results are shown in the extended version of this article [32].

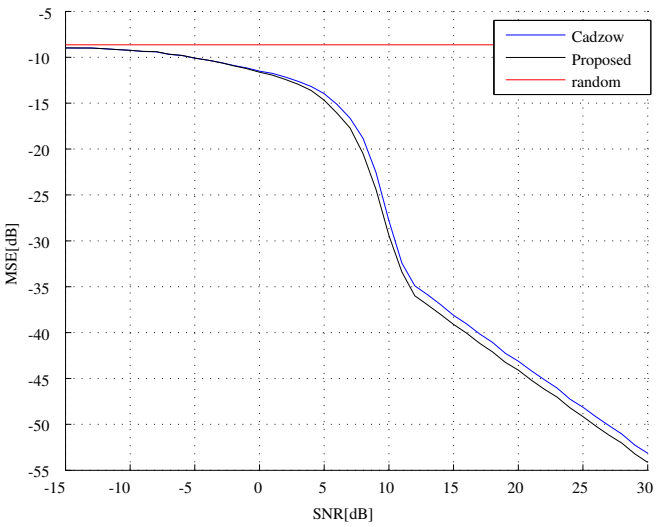

Fig. 1. Plot in log-log scale of the mean squared periodic error $\min\big((\tilde{t}_1 - t_1)_\tau^2 + (\tilde{t}_2 - t_2)_\tau^2, (\tilde{t}_1 - t_2)_\tau^2 + (\tilde{t}_2 - t_1)_\tau^2\big)$, where $(x)_\tau = \big((x + \frac{\tau}{2}) \mod \tau\big) - \frac{\tau}{2}$, for the estimation of the locations of $K = 2$ Dirac pulses, averaged over 10,000 noise realizations for every signal-to-noise ratio (SNR). The true parameters were $t_1 = 0.42$, $t_2 = 0.52$, $a_1 = a_2 = 1$, with $\tau = 1$, $N = 11$. The proposed algorithm is compared to Cadzow denoising, with $P = M = 5$, $\mu = 1$, $\gamma = 0.51\mu$, $x^{(0)} = s^{(0)} = \mathbf{T}_P$. Both algorithms converge within machine precision in about 100 iterations. An upper bound for the error is given by the naive estimator, which sets $\tilde{t}_1$ and $\tilde{t}_2$ randomly and uniformly in $[0, \tau[$.

5. CONCLUSION

We showed that the maximum-likelihood estimation of the parameters of Dirac pulses from noisy lowpass-filtered samples can be recast as a structured low-rank approximation problem. We proposed a new heuristic optimization algorithm, which provides a local solution of the difficult nonconvex problem. Preliminary results show that it outperforms the state-of-the-art approach based on Cadzow denoising, for same ease of implementation and complexity, essentially one SVD per iteration. More in-depth analysis of the performances is currently led by the authors. A Matlab implementation is available on the webpage of the first author.

6. REFERENCES

[1] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008, Special issue on Compressive Sampling.

[2] M. Mishali and Y. C. Eldar, "Sub-Nyquist sampling: Bridging theory and practice," *IEEE Signal Processing Mag.*, vol. 28, no. 6, pp. 98–124, Nov. 2011.

[3] J. Urigüen, Y. C. Eldar, P. L. Dragotti, and Z. Ben-Haim, "Sampling at the rate of innovation: Theory and applications," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012.

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

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

[6] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Trans. Signal Processing*, vol. 59, no. 4, pp. 1491–1504, Apr. 2011.

[7] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, "Performance bounds and design criteria for estimating finite rate of innovation signals," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 4993–5015, Aug. 2012.

[8] 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.

[9] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, NJ, 2005.

[10] C. Radhakrishna Rao and L. C. Zhao, "Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals," *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1461–1464, Mar. 1993.

[11] T.-H. Li and K.-S. Song, "On asymptotic normality of nonlinear least squares for sinusoidal parameter estimation," *IEEE Trans. Signal Processing*, vol. 56, no. 9, pp. 4511–4515, Sept. 2008.

[12] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, "Structured low rank approximation," *Linear Algebra Appl.*, vol. 366, pp. 157–172, 2003.

[13] J. Gillard and A. Zhigljavsky, "Analysis of structured low rank approximation as an optimization problem," *Informatica*, vol. 22, no. 4, pp. 489–505, 2011.

[14] V. Y. F. Tan and V. K. Goyal, "Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5135–5146, Oct. 2008.

[15] 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, Jan. 2010.

[16] C. J. Ying, L. C. Potter, and R. Moses, "On model order determination for complex exponential signals: Performance of an FFT-initialized ML algorithm," in *Proc. of IEEE SP Workshop on SSAP*, Quebec City, Canada, 1994, pp. 43–46.

[17] 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 Trans. Signal Processing*, vol. 44, no. 9, pp. 2245–2259, 1996.

[18] P. Stoica and Y. Selén, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: A refresher," *IEEE Signal Processing Mag.*, vol. 21, no. 1, pp. 112–114, Jan. 2004.

[19] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[20] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[21] T. Blu, "The generalized annihilation property—A tool for solving finite rate of innovation problems," in *Proc. of Int. Workshop on Sampling Theory and Appl. (SampTA)*, Marseille, France, May 2009.

[22] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica J. IFAC*, vol. 44, no. 4, pp. 891–909, 2008.

[23] S. Shalev-Shwartz, A. Gonen, and O. Shamir, "Large-scale convex minimization with a low-rank constraint," in *Proc. of ICML*, 2011, pp. 329–336.

[24] N. Gillis and F. Glineur, "Low-rank matrix approximation with weights or missing data is NP-hard," *SIAM. J. Matrix Anal. & Appl.*, vol. 32, no. 4, pp. 1149–1165, 2011.

[25] C. D. Sinclair and J. D. Vaaler, "Self-inversive polynomials with all zeros on the unit circle," in *London Mathematical Society Lecture Note Series: Number Theory and Polynomials*, J. McKee and C. Smyth, Eds., vol. 352, pp. 312–321. 2008.

[26] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*, Springer, 2012.

[27] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," documention of a software package, downloadable at `http://homepages.vub.ac.be/~imarkovs/slra/software.html`, 2012.

[28] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 1, pp. 49–62, Jan. 1988.

[29] D. R. Luke, "Prox-regularity of rank constraint sets and implications for algorithms," *J. Math. Imaging and Vision*, 2012.

[30] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, 2013, in press.

[31] G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins Univ. Press, 3rd edition, 1996.

[32] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements," preprint hal-00759253, submitted to *IEEE Trans. Sig. Proc.*, 2012.