Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements

Laurent Condat, Akira Hirabayashi

Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements

Laurent Condat and Akira Hirabayashi

***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 amounts to a difficult, even believed NP-hard, matrix problem of structured low rank approximation. To solve it, 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. Thus, our method improves upon the classical Cadzow denoising method, for same ease of implementation and speed.**

***Index Terms*—Recovery of Dirac pulses, spike train, finite rate of innovation, super-resolution, spectral estimation, maximum likelihood estimation, structured low rank approximation, alternating projections, Cadzow denoising, optimization, splitting**

I. Introduction and Problem Formulation

Reconstruction of signals lying in linear spaces, including bandlimited signals and splines, has long been the dominant paradigm in sampling theory, rooted in Shannon's work [1]. Recently, analog reconstruction from discrete samples has been enlarged to a broader class of signals, with so-called finite rate of innovation, i.e. ruled by parsimonious models [2]–[5]. This emerging theory predates and parallels the framework of sparse recovery and compressed sensing [6].

In this paper, we focus on the prototypical problem of estimating a finite stream of Dirac pulses, a.k.a. spike train, from uniform, noisy, lowpass-filtered samples [2], [3], [7]–[10]. More precisely, the sought-after unknown signal $\check{s}$ 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

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

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

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

$$= \sum_{k=1}^{K} \check{a}_k \varphi\Big(\frac{n\tau}{N} - \check{t}_k\Big) + \varepsilon_n, \quad \forall n = 0,\dots,N-1, \tag{3}$$

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 [9]–[12]. 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. Additionally, we assume, without loss of generality and only to simplify the notations, that N is odd, of the form $N = 2M+1$. Since the emphasis in this paper is on appropriately handling the presence of noise and not on being the most general, we adopt the simplest choice of the Dirichlet sampling function [9], which amounts to periodizing the signal s on the real line before sampling it with the sinc function:

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

The extension of the setting to the reconstruction of pulses with real shape, instead of the ideal Dirac distribution, is of practical interest for a wide range of applications, including ultrawideband communications [4] and the detection of impulsive signals in biomedical applications [9]. This generalization, or equivalently the choice of another sampling function φ, can be addressed without any difficulty, as described in details in [9]; therefore, to keep the derivations and notations simple, we do not consider this general case in the paper. We note that the considered sparse deconvolution problem is sometimes called *super-resolution* [13], [14], because it consists in locating the

L. Condat is with GIPSA-lab, a joint research unit of the CNRS and the University of Grenoble, France. Contact: see http://www.gipsa-lab.fr/∼laurent.condat/

A. Hirabayashi is with Ritsumeikan University, Kyoto, Japan.

The first author initiated this work while at Yamaguchi University in Spring 2012, invited by the second author, thanks to a three-months fellowship of the Japanese Society for the Promotion of Science.

Part of this work has appeared at the conference IEEE ICASSP, 2013, with title "Recovery of Nonuniform Dirac Pulses from Noisy Linear Measurements".

[1]The check symbol ˇ is used to distinguish the true parameter values from other quantities, like their estimates denoted with a tilde ˜.

pulses with precision beyond the resolution limit of τ/N suggested by the classical sampling theory.

The paper is organized as follows. In Sect. II, we formulate the maximum likelihood estimation problem and in Sect. III, we show that it amounts to a low rank matrix approximation problem. The new algorithm to solve it is presented in Sect. IV. Experimental results illustrate the effectiveness of the approach in Sect. V.

II. 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. Since we have assumed Gaussian noise, this corresponds to solving the nonlinear least-squares problem [15], [16]:

$$\text{Find } (\tilde{\mathbf{t}}, \tilde{\mathbf{a}}) = \underset{\mathbf{t}\in[0,\tau[^K, \mathbf{a}\in\mathbb{R}^K}{\arg\min} \sum_{n=0}^{N-1} \left| v_n - \sum_{k=1}^{K} a_k \varphi\Big(\frac{n\tau}{N} - t_k\Big) \right|^2. \tag{5}$$

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 $n = 0, \dots, N-1$,

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

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

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

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

Since the inverse discrete Fourier transform is unitary, up to a constant, the problem (9) can be rewritten as [15]:

Maximum likelihood (ML) estimation problem

$$\text{Find } (\tilde{\mathbf{t}}, \tilde{\mathbf{a}}) = \underset{\mathbf{t}\in[0,\tau[^K, \mathbf{a}\in\mathbb{R}^K}{\arg\min} \sum_{m=-M}^{M} \left| \hat{v}_m - \sum_{k=1}^{K} a_k e^{-j2\pi m\frac{t_k}{\tau}} \right|^2. \tag{10}$$

We remark that (10) takes the form of a spectral estimation problem, which consists in retrieving the parameters of a sum of complex exponentials or sinusoids from noisy samples [17]–[19]. This classical problem, which has a wide range of applications, e.g. in communications, radar, sonar, and geophysical seismology [19], has been studied extensively [20]–[26]. When $N \gg K$ and the locations t_k are not too close to each other, classical spectral estimation techniques like MUSIC [27] and ESPRIT [28], or greedy strategies [29], can be used; they are fast but statistically suboptimal. Also, we can mention recent developments inspired by the theory of compressed sensing, aiming at spectral estimation from random samples [30]–[34]; this is notably different from our setting, where the Fourier coefficients are successive and not chosen randomly. In this work, we investigate the ML estimation problem (10) in its whole generality, without any simplifying assumption. The optimal statistical properties of ML estimation for spectral estimation are well known [21]–[23], [35]–[38]. However, solving the problem (10) is a difficult task, as the cost function has a multimodal shape with many local minima and a narrow trough around the global minimum [39]–[43], see the example of Fig. 1. Stochastic optimization approaches can be applied [15], [44], [45], 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) [46]–[49]. These approaches proceed in two steps: 1) a method is used to obtain a good initial estimate; 2) this estimate is refined iteratively and converges to the closest local minimizer of the nonconvex cost function in (10). Thus, the quality of the method used for the first step is crucial. A major advantage of the approach developed in the next section is to get rid of this initialization problem.

III. The Annihilation Property: Reformulation of eqn. (10) as a Matrix Denoising Problem

Let us assume temporarily that there is no noise, i.e. $\hat{\varepsilon}_m = 0$ in (9). Then, the sequence of Fourier coefficients $\{\hat{v}_m\}_{m=-M}^M$ satisfies a linear *annihilating* difference equation [50], a known property which dates back to de Prony's work in the eighteenth century [51]. That is, the convolution of the sequence of Fourier coefficients with the annihilating filter $\{h_k\}_{k=0}^K$ is identically zero:

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

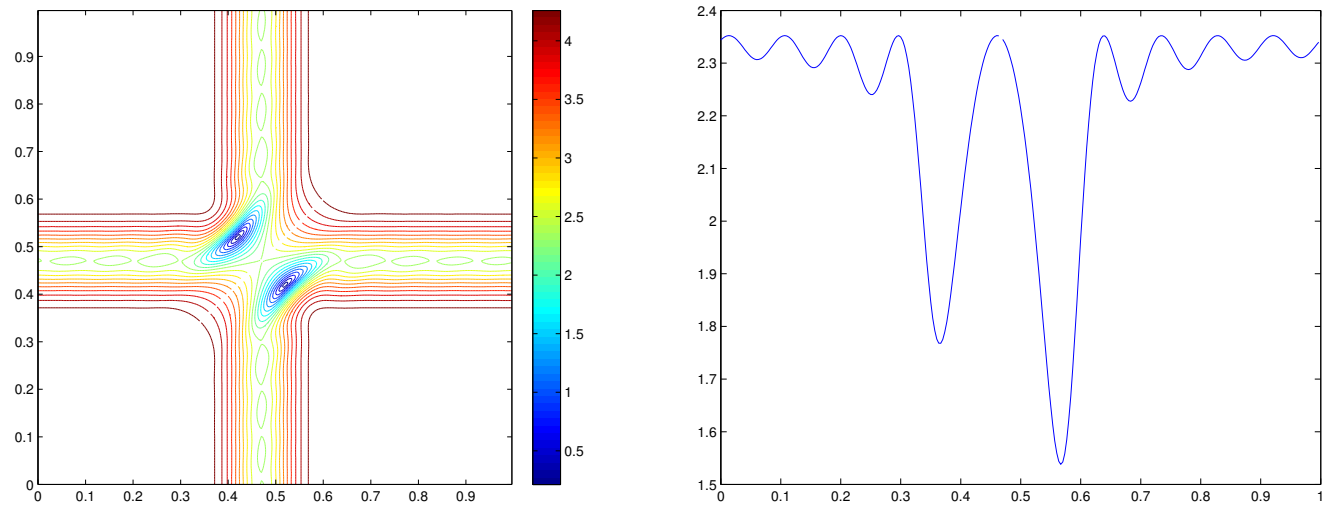
with the (reversed) Z-transform of the annihilating filter defined, up to a constant, as

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

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 any integer P with $K \le P \le M$, the Toeplitz—i.e. with constant values along its diagonals—matrix



(a) (b)

Fig. 1. (a) Contour plot showing the level lines of the cost function in (10), as a function of t_1 and t_2, on the horizontal and vertical axes, respectively. (b) plot of the same cost function as a function of t_2, for $t_1 = 0.47$. For given t_1 and t_2, the values of a_1 and a_2 are the optimal values obtained by solving (20). The setting is the same as in Figs. 3 and 4 (a): the true unknown signal consists of $K = 2$ Dirac pulses with $\tau = 1$, $(\check{t}_1, \check{t}_2) = (0.42, 0.52)$ and $(\check{a}_1, \check{a}_2) = (1, 1)$; we have $N = 11$ noisy measurements with SNR=15dB.

$\mathbf{T}_P$, of size $N - P \times P + 1$, obtained by arranging the values $\{\hat{v}_m\}_{m=-M}^{M}$ in its first row and column; $\mathbf{T}_P$ is depicted in (17) and in (13) for $P = K$. 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.

Hence, turning back to the case when noise is present in the data, we can reformulate the estimation problem (10) as the following matrix approximation problem, named *structured low rank approximation* (SLRA) in the literature [52]–[54]:

Structured low rank approximation (SLRA) problem

$$\text{Find } \widetilde{\mathbf{T}}_P \in \underset{\mathbf{T} \in \mathbb{C}^{N-P \times P+1}}{\arg\min} \|\mathbf{T} - \mathbf{T}_P\|_w^2 \qquad (14)$$
$$\text{s. t. } \mathbf{T} \text{ is Toeplitz and } \operatorname{rank}(\mathbf{T}) \leq K,$$

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

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

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

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

The SLRA problem (14), which consists in projecting a matrix in the intersection of a linear subspace and a nonconvex manifold, is believed to be NP-hard [53], [55], [56]. So, at first glance, we just have replaced the difficult problem (10) by the SLRA problem of same difficulty. However, the parametric problem (10) has been reformulated as a matrix *denoising* problem: the matrix $\mathbf{T}_P$, or equivalently the data $\{v_n\}_{n=0}^{N-1}$, is denoised by finding the closest matrix consistent with the model's structure. The main advantage is that the initialization problem disappears: an iterative algorithm to solve the SLRA problem 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).

Since the desired parameters $\{\tilde{t}_k\}_{k=1}^{K}$ and $\{\tilde{a}_k\}_{k=1}^{K}$ are somehow *encoded* in the matrices, we have to describe the extraction process. The whole reconstruction procedure, also given in [3], is detailed in Fig. 2. In absence of noise, it yields perfect reconstruction of the parameters. We note that the estimation of the locations is decoupled from that of the amplitudes; since finding the amplitudes given the locations is a simple linear regression problem, the emphasis is on the estimation quality of the locations.

We must remark that the estimates $\{(\tilde{t}_k, \tilde{a}_k)\}_{k=1}^{K}$, obtained by the procedure, 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 guaranteed to be the case[2] if the noise level $\|\boldsymbol{\varepsilon}\|$ is below some threshold, which depends on

[2]In details, 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 in $-M, \ldots, M$, where $\cdot^*$ denotes complex conjugation. Consequently, the polynomial $\sum_{k=0}^{K} \tilde{h}_k z^k$ is self-inversive [57], 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. The perturbation due to noise on the coefficients $\hat{v}_m$ yields a proportionally bounded perturbation on the estimated roots $\tilde{z}_k = e^{j2\pi\tilde{t}_k/\tau}$, with respect to the true roots $\check{z}_k = e^{j2\pi\check{t}_k/\tau}$, but the roots remain on the complex unit circle. Only if the perturbation is large enough, two distinct roots $(\tilde{z}_{k_1}, \tilde{z}_{k_2})$ will possibly merge and then split in a pair $(\tilde{z}_{k_1}, \tilde{z}_{k_2} = 1/\tilde{z}_{k_1}^*)$ on both sides of the unit circle, yielding $\tilde{t}_{k_1} = \tilde{t}_{k_2}$.

Reconstruction Procedure

Input: The Fourier coefficients $\{\hat{v}_m\}_{m=-M}^{M}$.
Output: the estimates $\{\tilde{t}_k\}_{k=1}^{K}$ and $\{\tilde{a}_k\}_{k=1}^{K}$ of the unknown locations $\{\check{t}_k\}_{k=1}^{K}$ and amplitudes $\{\check{a}_k\}_{k=1}^{K}$.

Step 1. Choose an integer P such that $K \leq P \leq M$; we recommend $P = M$. Construct the Toeplitz matrix $\mathbf{T}_P$ of size $N-P \times P+1$ by arranging the coefficients $\{\hat{v}_m\}_{m=-M}^{M}$ in its first row and column and repeating them along the diagonals:

$$\mathbf{T}_P = \begin{pmatrix} \hat{v}_{-M+P} & \cdots & \hat{v}_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \hat{v}_M & \cdots & \hat{v}_{M-P} \end{pmatrix}. \tag{17}$$

Step 2. Solve the SLRA problem (14).

Step 3. If $P > K$, reshape the denoised Toeplitz matrix $\widetilde{\mathbf{T}}_P$ to a Toeplitz matrix $\widetilde{\mathbf{T}}_K$ of size $N - K \times K + 1$; that is,

$$\widetilde{\mathbf{T}}_K = \begin{pmatrix} \tilde{\hat{v}}_{-M+K} & \cdots & \tilde{\hat{v}}_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \tilde{\hat{v}}_M & \cdots & \tilde{\hat{v}}_{M-K} \end{pmatrix}. \tag{18}$$

Note that both $\widetilde{\mathbf{T}}_P$ and $\widetilde{\mathbf{T}}_K$ have rank K if (14) is solved exactly.

Step 4. Compute the right singular vector $\tilde{\mathbf{h}} = [\tilde{h}_0 \ \cdots \ \tilde{h}_K]^{\mathrm{T}}$ of $\widetilde{\mathbf{T}}_K$ corresponding to the singular value zero (the smallest singular value in practice); that is, in the singular value decomposition (SVD) $\widetilde{\mathbf{T}}_K = \mathbf{L}\mathbf{\Sigma}\mathbf{R}^{\mathrm{H}}$, $\tilde{\mathbf{h}}$ corresponds to the last column of $\mathbf{R}$. Then, 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

$$\tilde{t}_k = \frac{\tau}{2\pi} \arg_{[0,2\pi[}(\tilde{z}_k), \quad \forall k = 1, \ldots, K. \tag{19}$$

Step 5. Given the estimates $\{\tilde{t}_k\}_{k=1}^{K}$, the maximum-likelihood estimates $\{\tilde{a}_k\}_{k=1}^{K}$ of the amplitudes are obtained by solving the linear system

$$\widetilde{\mathbf{U}}^{\mathrm{H}}\widetilde{\mathbf{U}}\tilde{\mathbf{a}} = \widetilde{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{v}}, \tag{20}$$

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

$$\widetilde{\mathbf{U}} = \begin{pmatrix} e^{j2\pi M\tilde{t}_1/\tau} & \cdots & e^{j2\pi M\tilde{t}_K/\tau} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi M\tilde{t}_1/\tau} & \cdots & e^{-j2\pi M\tilde{t}_K/\tau} \end{pmatrix}. \tag{21}$$

Fig. 2. Procedure to estimate the parameters of a stream of Dirac pulses $\check{s}(t) = \sum_{k=1}^{K} \check{a}_k \delta(t - \check{t}_k)$ from noisy Fourier coefficients.

the separation $\min_{k_1 \neq k_2} |\check{t}_{k_1} - \check{t}_{k_2}|$.

A. State of the Art for Structured Low Rank Approximation (SLRA)

The problem of reconstructing a low-rank matrix from limited, noisy, linear measurements arises in many areas of engineering and applied sciences, such as signal processing, machine learning, control, and computer vision. The wide range of applications includes low-order system identification, dynamic MRI, quantum state tomography, phase retrieval, and global positioning from local distances, to mention a few; see references in [52], [53], [58], [59]. In many of these problems, the matrix to be recovered is constrained to have some structure, like being Toeplitz, Hankel, Sylvester, or positive definite. The SLRA problems have been studied in the literature under different names; they are equivalent or have tight connections with constrained or structured total least-squares, total least-norm, and errors-in-variables [52], [53], [60]–[65]. Several methods have been proposed, in the community of numerical algebra, to obtain a local solution of a SLRA problem, using high-level optimization routines [42], [62], [66]–[71]. For instance, the iterative approach in [42] uses the Matlab function `fminunc`, which is a BFGS quasi-Newton method with line search. Besides the difficulty of implementation, the algorithm is very costly, as it requires many SVD at each iteration. So, to our knowledge, the only efficient publicly available software package for SLRA is the one currently in development by I. Markovsky [58]. However, it only handles real-valued data, whereas the matrices in (14) are complex-valued with centro-Hermitian symmetry.

Due to the complexity of dedicated solvers for SLRA, a simple alternative is to use the heuristic method of *Cadzow denoising* [72], [73], a.k.a. alternating projections. This method is promoted in [3], [9] for the reconstruction of Dirac pulses, and represents the state of the art for this problem. To describe the method, let us introduce some notations.

- We place ourselves in the real Hilbert space $\mathcal{H} = \mathbb{C}^{N-P\times P+1}$ of complex-valued matrices of size $N - P \times P + 1$, having centro-Hermitian symmetry. The (Frobenius) inner product between two matrices $\mathbf{X}$, $\mathbf{X}'$ of $\mathcal{H}$ is $\langle \mathbf{X}, \mathbf{X}' \rangle = \sum_{i,j} x_{i,j} x'^{*}_{i,j} \in \mathbb{R}$.
- We denote by P_Ω the projection onto a subset Ω of $\mathcal{H}$, which maps an element of $\mathcal{H}$ to the closest element in Ω, with respect to the Frobenius norm.
- We define $\mathcal{T}$, the linear subspace of $\mathcal{H}$ of Toeplitz matrices. $P_{\mathcal{T}}$ consists in averaging along the diagonals; that is, $P_{\mathcal{T}}$ maps a matrix $\mathbf{X}$ with entries $\{x_{i,j}\}_{i=1,\ldots,N-P}^{j=1,\ldots,P+1}$ to a matrix $\mathbf{X}'$ with entries $x'_{i,j} =$

$$\begin{cases} \frac{1}{i-j+P+1} \sum_{k=1}^{i-j+P+1} x_{k,k+j-i} & \text{if } i-j \leq 0, \\ \frac{1}{P+1} \sum_{k=1}^{P+1} x_{k+i-j,k} & \text{if } 1 \leq i-j \leq N-2P-1, \\ \frac{1}{j-i+N-P} \sum_{k=1}^{j-i+N-P} x_{k+i-j,k} & \text{if } i-j \geq N-2P. \end{cases} \tag{22}$$

- We define $\mathcal{R}_K$, the closed but nonconvex manifold of matrices of $\mathcal{H}$ having rank at most K. $P_{\mathcal{R}_K}$ consists in truncating the SVD of the matrix, according to the

classical Schmidt-Eckart-Young theorem [74] [75, theorem 2.5.3]: if a matrix $\mathbf{X}$ has SVD $\mathbf{X} = \mathbf{L}\mathbf{\Sigma}\mathbf{R}^{\mathrm{H}}$, then $P_{\mathcal{R}_K}(\mathbf{X})$ is obtained by setting to zero all except the K largest diagonal elements of $\mathbf{\Sigma}$, which are the singular values of $\mathbf{X}$.

Then, Cadzow denoising simply consists in applying alternately $P_{\mathcal{R}_K}$ and $P_{\mathcal{T}}$ to the matrix $\mathbf{T}_P$:

Cadzow denoising algorithm. Set $\mathbf{T}_P^{(0)} = \mathbf{T}_P$, depicted in (17). Then iterate, for every $l \geq 0$,

$$\left| \mathbf{T}_P^{(l+1)} = P_{\mathcal{T}}\big(P_{\mathcal{R}_K}(\mathbf{T}_P^{(l)})\big)\,. \right.$$

This algorithm seems to always converge in practice to a matrix $\widetilde{\mathbf{T}}_P \in \mathcal{T} \cap \mathcal{R}_K$, i.e. a Toeplitz matrix of rank at most K. But contrary to a common belief, this convergence has not been proved [76]–[79], although some partial convergence results have been derived recently [80]–[82]. Moreover, even if Cadzow denoising converges to a Toeplitz matrix of rank at most K, this matrix is not a local minimizer of the Frobenius distance or of the weighted Frobenius distance $\|\cdot - \mathbf{T}_P\|_w$ in the SLRA problem (14), as shown by examples in [42], [43], [60]. In the next section, we propose a new algorithm to compute a local solution of the SLRA problem, which thus improves upon Cadzow denoising, for essentially the same complexity of one SVD per iteration.

IV. A New Optimization Algorithm for SLRA

Before addressing the particular problem (14), let us consider the generic optimization problem:

$$\text{Find } \tilde{x} \in \operatorname*{arg\,min}_{x \in \mathcal{H}} F(x) \quad \text{s.t.} \quad x \in \Omega_1 \cap \Omega_2, \tag{23}$$

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, $\|\nabla F(x') - \nabla F(x)\| \leq \beta \|x - x'\|$, $\forall x, x' \in \mathcal{H}$.

The following algorithm allows to solve (23):

Optimization 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 $l \geq 0$,

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

This algorithm is a particular instance of a more general proximal splitting algorithm derived recently by the first author [83]. Consequently, as corollaries of the results derived in [83], we get the following convergence results.

Theorem 1. *In* (23)*, 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 [84]. In the algorithm, suppose that (iv)* $2\gamma > \beta\mu$*. Then, the sequence* $(x^{(l)})_{l \in \mathbb{N}}$ *converges to some element* $\tilde{x}$ *solution to the problem* (23)*.*

Theorem 2. *In* (23)*, suppose that* $F = 0$ *and that* (i) *the sets* Ω_1 *and* Ω_2 *are convex;* (ii) $\mathrm{ri}(\Omega_1) \cap \mathrm{ri}(\Omega_2) \neq \emptyset$*. In the algorithm, suppose that* (iii) $\gamma = \mu = 0$*. Then, the algorithm reverts to the Douglas–Rachford splitting algorithm, well known in optimization theory [84], [85]. Consequently, the sequence* $(x^{(l)})_{l \in \mathbb{N}}$ *converges to some element* $\tilde{x} \in \Omega_1 \cap \Omega_2$*.*

Now, we recognize that the SLRA problem (14) is a particular instance of (23); in the notations introduced in the previous section, $\Omega_1 = \mathcal{R}_K$, $\Omega_2 = \mathcal{T}$, and $F = \frac{1}{2}\|\cdot - \mathbf{T}_P\|_w^2$. This motivates the use of the proposed optimization algorithm to solve the SLRA problem. Let us rewrite explicitly the algorithm in this setting. For this, we remark that $\nabla F(\mathbf{X}) = \mathbf{W} \circ (\mathbf{X} - \mathbf{T}_P)$, where $\circ$ is the entrywise (Hadamard) product and the entries $w_{i,j}$ of the matrix $\mathbf{W}$ are defined in (16). The Lipschitz constant of ∇F is $\beta = \max_{i,j} |w_{i,j}| = 1$.

Proposed algorithm for SLRA. Choose the parameters $\mu > 0$, $\gamma \in \,]0, 1[$, such that $2\gamma > \mu$, and set $\mathbf{T}_P^{(0)} = \mathbf{S}_P^{(0)} = \mathbf{T}_P$. Then iterate, for every $l \geq 0$,

$$\left| \begin{array}{l} \mathbf{T}_P^{(l+1)} = P_{\mathcal{R}_K}\big(\mathbf{S}_P^{(l)} + \gamma(\mathbf{T}_P^{(l)} - \mathbf{S}_P^{(l)}) - \mu \mathbf{W} \circ (\mathbf{T}_P^{(l)} - \mathbf{T}_P)\big), \\ \mathbf{S}_P^{(l+1)} = \mathbf{S}_P^{(l)} - \mathbf{T}_P^{(l+1)} + P_{\mathcal{T}}(2\mathbf{T}_P^{(l+1)} - \mathbf{S}_P^{(l)}). \end{array} \right.$$

However, we remark that one of the assumption in Theorem 1 is not met in our context: the set $\mathcal{R}_K$ is not convex. Therefore, the convergence of the proposed algorithm to a local solution of the problem (14) is not guaranteed, and we will use the algorithm as a heuristic. In fact, Cadzow denoising and the Douglas–Rachford algorithm are also heuristic methods for nonconvex problems, but a large body of empirical evidence shows that these two algorithms actually work in a surprisingly broad range of applications and achieve comparable results to much more sophisticated, special purpose algorithms. They have been used for protein fold prediction, phase retrieval in optics, crystallography, graph coloring, and solving Sudokus [86]–[89]. The Douglas–Rachford iteration generally outperforms Cadzow denoising for these applications [86], [88], [89]. The motivation for applying the new proposed algorithm to SLRA stems from this practical success, since it can be viewed as a Douglas–Rachford iteration with an adequately plugged gradient descent with respect to the cost function. However, there are almost no convergence results for the Douglas–Rachford algorithm in the nonconvex case, as the behavior of such iterative methods is particularly difficult to analyze [89]. Still, we can mention the following good property: for every fixed point $(\widetilde{\mathbf{T}}_P, \widetilde{\mathbf{S}}_P)$ of the proposed algorithm, $\widetilde{\mathbf{T}}_P \in \mathcal{R}_K \cap \mathcal{T}$.

In practice, when using the proposed algorithm to solve the SLRA problem (14), we observed that it almost always converges. Only when the noise level is very high, and in rare cases, it gets trapped in a cycle and does not converge; if this happens, running again the algorithm with lower values of the parameters μ and γ is sufficient to obtain convergence. More extensive use of the algorithm will be necessary to understand its behavior and the appropriate values of μ and γ to use. When the algorithm converges, it always does so to a Toeplitz matrix of rank K, which is a local minimizer of the weighted Frobenius cost function. More precisely, we checked that the locations $\{\tilde{t}_k\}_{k=1}^K$ and amplitudes $\{\tilde{a}_k\}_{k=1}^K$ obtained at convergence of the algorithm are local solutions of the ML

problem (10) as follows. As remarked in [60, Sect. IV.C], in the case of $K = 1$ Dirac pulse, all the local minima of the least squares cost function (10) with respect to the location are roots of a particular polynomial Q constructed from the data. That is, setting the partial derivative with respect to t to zero in $\sum_{m=-M}^{M} |\hat{v}_m - ae^{-j2\pi mt/\tau}|^2$ yields $\sum_{m=-M}^{M} mz^m \hat{v}_m^* = 0$, where $z = e^{-j2\pi mt/\tau}$. Equivalently, t is a local solution only if z is a root on the unit circle of the polynomial

$$Q(z) = \sum_{m=0}^{2M} (m - M)\hat{v}_{m-M}^* z^m. \tag{24}$$

Therefore, in our case of multiple Dirac pulses, we can subtract to the data $\{\hat{v}_m\}_{m=-M}^{M}$ the sequences $\{\tilde{a}_k e^{-j2\pi m\tilde{t}_k/\tau}\}_{m=-M}^{M}$ for all values of k except one value k_0 and check whether $\tilde{z}_{k_0}$, defined in (19), is a root of the corresponding polynomial Q. We observed empirically that this is always the case.

A. A Convex Variant of the Approach

To circumvent the difficulty of achieving the ML estimate for spike train recovery from noisy Fourier coefficients, a convex surrogate of the problem has been proposed in recent papers, based on the minimization of the so-called total variation measure [13], [14], [90]. The main limitation of this approach is the necessity of a minimal separation between the pulses, like $\min_{k_1 \neq k_2} |\check{t}_{k_1} - \check{t}_{k_2}| \geq 2\tau/M$. In this section, we propose another convex relaxation of the problem, which is free of such assumption of minimal separation. It is obtained by replacing the rank in the SLRA problem (14) by its convex surrogate, the nuclear norm [91]:

Convex approximation of the SLRA problem

$$\text{Find } \widetilde{\mathbf{T}}_P \in \underset{\mathbf{T} \in \mathbb{C}^{N-P \times P+1}}{\arg\min} \tfrac{1}{2}\|\mathbf{T} - \mathbf{T}_P\|_w^2 + \lambda \|\mathbf{T}\|_* \tag{25}$$
$$\text{s. t.} \quad \mathbf{T} \text{ is Toeplitz,}$$

where the nuclear norm $\|\cdot\|_*$ of a matrix is the sum of its singular values and the Lagrangian parameter $\lambda > 0$ controls the tradeoff between the antagonist terms.

The main advantage of this formulation is to be a strongly convex semi-definite program, for which many efficient algorithms exist to compute the unique solution. Also, replacing the rank by the nuclear norm has proved successful for several problems, like matrix completion [59], [92], [93]. However, in our setting, two pulses can be arbitrarily close to each other, so that the measurements can be highly coherent. Therefore, the quality of the solution to a problem like (25) can be significantly degraded in comparison with the SLRA solution, as shown in another context in [94], [95]. For our problem, the solution of (25) seems significantly worse than the estimate given by the algorithm proposed in Sect. IV, as shown by the example in Fig. 7. Moreover, the choice of the value of λ to use is problematic; a stability analysis of the kind of the one for total variation minimization in [96] would be desirable, but we leave this difficult study for future work.

We note that a minor modification of the proposed algorithm for SLRA can compute the solution of (25): we simply have to replace the operator $P_{\mathcal{R}_K}$, which performs hard-thresholding of the singular values, by soft-thresholding $S_{\mu\lambda}$ of the singular values: in the SVD $\mathbf{X} = \mathbf{L}\mathbf{\Sigma}\mathbf{R}^{\mathrm{H}}$, $S_{\mu\lambda}(\mathbf{X})$ is obtained by replacing each of the diagonal elements $\{\sigma_{i,i}\}_{i=1}^{P+1}$ of the matrix $\mathbf{\Sigma}$ by $\max(\sigma_{i,i} - \mu\lambda, 0)$. As a consequence of the results in [83], the sequence $(\mathbf{T}_P^{(l)})_{l \in \mathbb{N}}$ generated by this algorithm converges to some matrix $\widetilde{\mathbf{T}}_P$ solution to (25).

V. Experimental Results

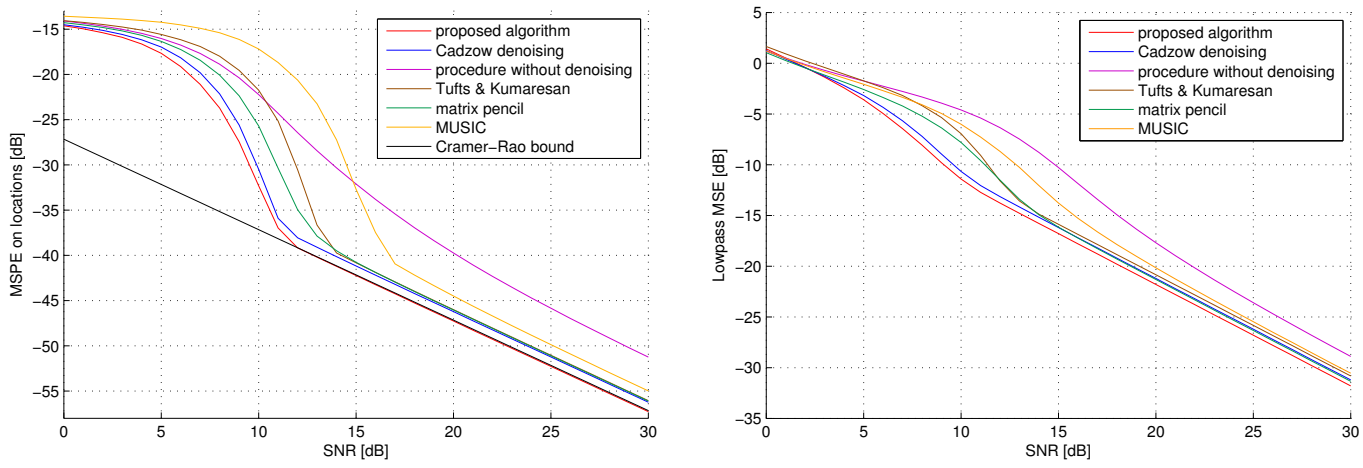
The reconstruction procedure described in Sect. III was implemented in Matlab, with the choice between Cadzow denoising and the algorithm proposed in Sect. IV to solve the SLRA problem (with $P = M$), or with no denoising (starting directly the procedure described in Fig. 2 at Step 3, with $\widetilde{\mathbf{T}}_K = \mathbf{T}_K$; this method is called *total least squares* in [3]). We also implemented several methods of spectral estimation: MUSIC [27] (we used the function `pmusic` of Matlab), the method of Tufts and Kumaresan [21], and the matrix pencil method [97][3]. The results obtained with this Matlab code[4] are illustrated for several experiments in Figs. 3–9.

Before discussing the results, we have to choose a criterion to quantify the estimation error between the true unknown signal $\check{s}$ and the estimated signal $\tilde{s} = \sum_{k=1}^{K} \tilde{a}_k \delta(t - \tilde{t}_k)$. We note that this question is related to the notion of distance between discrete measures in the theory of optimal transportation. Since the problem is parametric, it is natural to think of a kind of mean squared error (MSE) between the sets $\{t_k\}_{k=1}^{K}$ and $\{\tilde{t}_k\}_{k=1}^{K}$ on one hand, and $\{a_k\}_{k=1}^{K}$ and $\{\tilde{a}_k\}_{k=1}^{K}$ on the other hand. But the problem of assigning each estimated pulse to a true pulse is not trivial, especially because of the intrinsic periodicity of the setting: the boundaries at $t = 0$ and $t = \tau$ are identified. For the setting of Fig. 3, with two pulses of same amplitude, we considered the best assignment between the two possible permutations and computed the *mean squared periodic error* (MSPE) [98]. But, in general, this best assignment is not appropriate, because it does not depend on the amplitudes. Therefore, we adopt as error criterion the squared L_2 distance between the lowpass-filtered signals $\check{s} * \varphi$ and $\tilde{s} * \varphi$, where φ is the Dirichlet kernel defined in (4). From Parseval's identity, this is equivalent to considering the squared distance $\frac{1}{N}\sum_{m=-M}^{M} |\check{\hat{v}}_m - \tilde{\hat{v}}_m|^2$ between the true Fourier coefficients $\check{\hat{v}}_m$ of $\check{s}$ and the denoised Fourier coefficients $\tilde{\hat{v}}_m$, which are Fourier coefficients of $\tilde{s}$. We call this error criterion the *lowpass mean squared error*.

The first experiment, illustrated in Figs. 3 and 4 (a), corresponds to $K = 2$ pulses with same amplitude. In that case, the performances can be evaluated by comparing the MSPE on the locations with the corresponding Cramér–Rao bound (CRB), which is a lower bound for the expectation of the mean squared error achievable by any unbiased estimator, calculated like in [3]. The proposed method outperforms all the other methods over the whole range of signal-to-noise

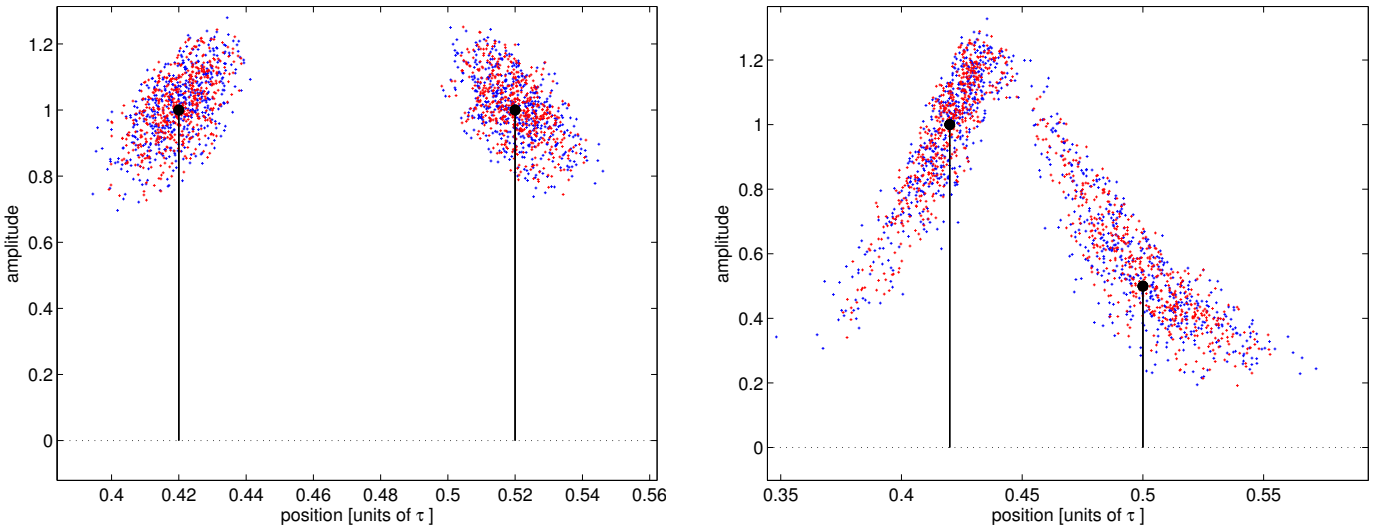
[3]For the method of Tufts and Kumaresan [21] and the matrix pencil method [97], we we used the implementation of Kfir Gedalyahu, found on his webpage at `https://sites.google.com/site/kfirgedal/`.

[4]The Matlab code implementing all the methods mentioned is available on the webpage of the first author at `http://www.gipsa-lab.grenoble-inp.fr/~laurent.condat/publications.html`.



(a) (b)

Fig. 3. Plot in log-log scale of (a) the mean squared periodic error (MSPE) $\min\big((\tilde{t}_1-t_1)_\tau^2+(\tilde{t}_2-t_2)_\tau^2,(\tilde{t}_1-t_2)_\tau^2+(\tilde{t}_2-t_1)_\tau^2\big)/2$, where $(x)_\tau=\big((x+\frac{\tau}{2})\bmod\tau\big)-\frac{\tau}{2}$, and (b) the lowpass MSE $\|\tilde{\mathbf{v}}-\tilde{\mathbf{v}}\|^2/N$, for several reconstruction methods. The true signal $\check{s}$ consists of $K=2$ Dirac pulses, with $(t_1,t_2)=(0.42,0.52)$, $(a_1,a_2)=(1,1)$, $\tau=1$, $N=11$. The values are averaged over 10,000 noise realizations for every integer value of the signal-to-noise ratio (SNR).



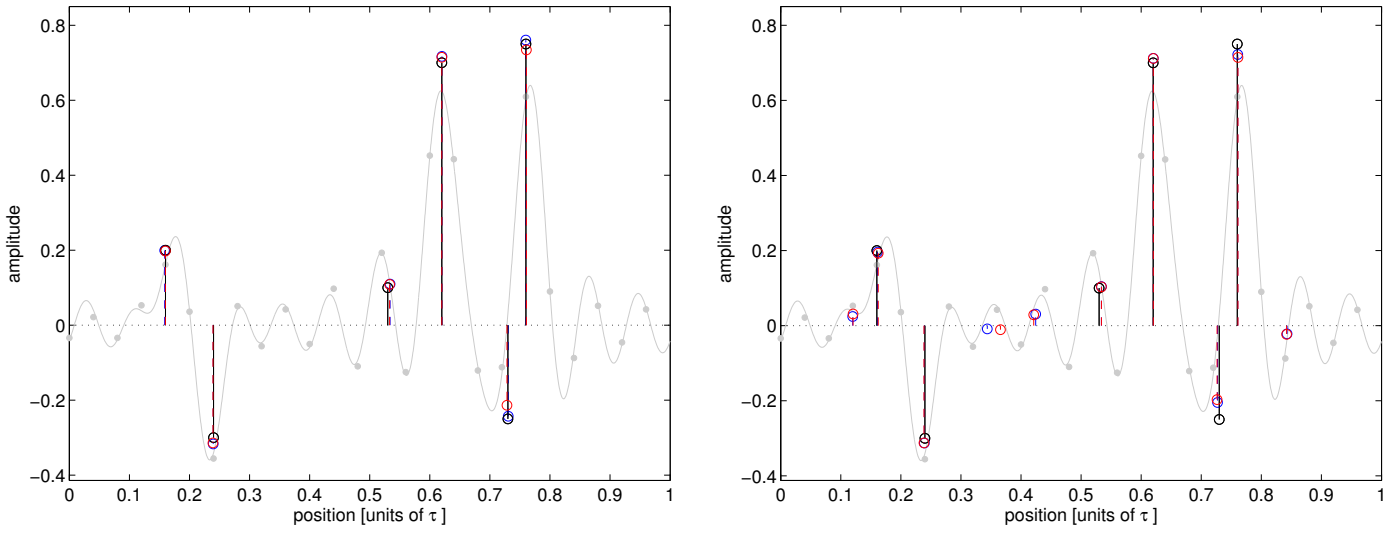
(a) (b)

Fig. 4. The signal consists of $K=2$ Dirac pulses, with $N=11$ noisy measurements, and the SNR is 15dB. (a) Same setting as in Fig. 3; (b) $(t_1,t_2)=(0.42,0.5)$ and $(a_1,a_2)=(1,0.5)$. In black: true pulses. In blue and red: reconstructed locations and amplitudes of the pulses with Cadzow denoising and the proposed algorithm, respectively, for 500 different noise realizations. The difference is small visually, but the proposed algorithm yields smaller estimation errors, with points clouds slightly less dispersed. The computation time for one reconstruction, with 50 iterations, was 14ms.

ratio[5] (SNR) values. For a SNR above 12dB, the MSPE of the proposed method achieves the CRB, which implies that the proposed method yields the optimal ML estimate. For a lower SNR, the performances are degraded, which means that the proposed method only achieves a local minimum of the SLRA problem (14), and not the global solution corresponding to the ML estimate. Although Cadzow denoising performs well, it does not achieve the CRB, even asymptotically, contrary to what is claimed in [3]. The proposed approach outperforms Cadzow denoising consistently, with errors on the locations about 10% lower in average. Interestingly, when solving with the proposed algorithm the SLRA problem (14), but with the Frobenius norm instead of the weighted Frobenius norm (which amounts to replace $\mathbf{W}\circ(\mathbf{T}_P^{(l)}-\mathbf{T}_P)$ by $(\mathbf{T}_P^{(l)}-\mathbf{T}_P)$ in the algorithm), the results are almost exactly the same as with Cadzow denoising; this indicates that the gain provided by the proposed method is explained by the presence in the problem formulation of the appropriate cost function.

In Figs. 5–9, we show examples for larger values of K and N. For the proposed algorithm, we used $\mu=0.1$ and

[5]We define the SNR as $20\log_{10}(\|\mathbf{v}-\boldsymbol{\varepsilon}\|/\|\boldsymbol{\varepsilon}\|)=20\log_{10}(\|\hat{\mathbf{v}}-\hat{\boldsymbol{\varepsilon}}\|/\|\hat{\boldsymbol{\varepsilon}}\|)=20\log_{10}(\|\tilde{\hat{\mathbf{v}}}\|/\|\hat{\boldsymbol{\varepsilon}}\|)$.



(a) (b)

Fig. 5. The signal consists of 6 Dirac pulses. We have $N = 25$ noisy measurements with SNR=25dB. In black: true pulses. In blue and red: reconstructed locations and amplitudes of the pulses with Cadzow denoising and the proposed algorithm, respectively, for one particular noise realization. The solid line in light gray is the lowpass-filtered signal $\check{s} * \varphi$ and the gray dots are the noisy measurements v_n. The computation time for one reconstruction, with 50 iterations, was 19ms. (a) Reconstruction with $K = 6$. The lowpass MSE with Cadzow denoising and the proposed algorithm is 1.39e-3 and 1.22e-3, respectively. (b) Reconstruction with $K = 10$, to illustrate the effects of overfitting: the methods have found imaginary pulses in noise.

$\gamma = 0.51\mu$. Actually, the values for μ and σ yielding fastest convergence depended on the experiment and on the noise level, and we could not come to a rule for tuning these parameters. Qualitatively, Cadzow denoising and the proposed algorithm behave similarly. Yet, the values of lowpass MSE indicate a slightly better reconstruction with the proposed method.

VI. Conclusion

We proposed a new heuristic optimization algorithm for structured low rank approximation. Its combination with Prony's extraction procedure yields an efficient method for the reconstruction of Dirac pulses. The maximum-likelihood parameter estimates are obtained, up to some threshold SNR. Thus, the proposed algorithm outperforms the state-of-the-art approach to date, 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. Many theoretical questions remain open, e.g. on the relationship between the proposed formalism and the very recent developments on total variation minimization [13], [14], [90]. We should also investigate the possibility to generalize the method in higher dimensions and apply it to other model identification problems [53].

References

[1] M. Unser, "Sampling—50 Years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[2] 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, Jun. 2002.

[3] 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.

[4] 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.

[5] 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.

[6] T. Strohmer, "Measure what should be measured: Progress and challenges in compressive sensing," *IEEE Signal Processing Lett.*, vol. 19, no. 12, pp. 887–893, Dec. 2012.

[7] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[8] J. Kusuma and V. K. Goyal, "On the accuracy and resolution of powersum-based sampling methods," *IEEE Trans. Signal Processing*, vol. 57, no. 1, pp. 182–193, 2009.

[9] 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.

[10] 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.

[11] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1741–1757, 2007.

[12] 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.

[13] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," 2012, preprint arXiv:1211.0290.

[14] J.-M. Azaïs, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," 2013, preprint arXiv:1301.5873.

[15] 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.

[16] A. Hirabayashi, Y. Hironaga, and L. Condat, "Sampling signals with finite rate of innovation and recovery by maximum likelihood estimation," *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences*, vol. E96-A, no. 10, Oct. 2013, in press.

[17] Special issue on spectral estimation, *Proc. IEEE*, vol. 70, no. 9, Sep. 1982.

[18] P. Stoica, "List of references on spectral line analysis," *Signal Processing*, vol. 31, no. 3, pp. 329–340, Apr. 1993.

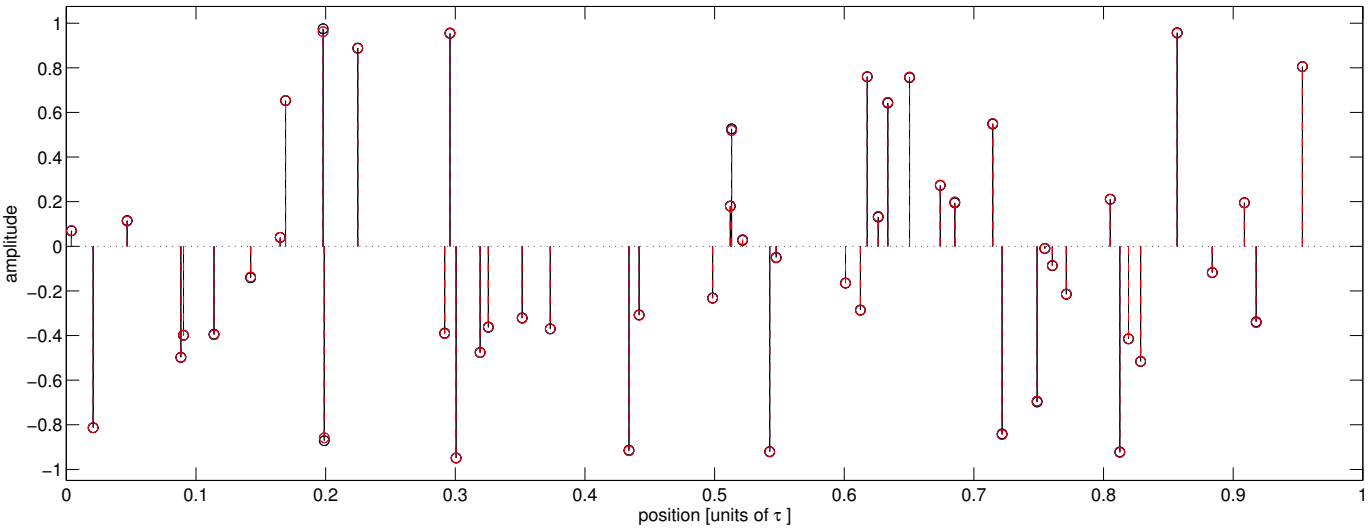


Fig. 6. The signal consists of $K = 50$ Dirac pulses, with $N = 1001$ noisy measurements and SNR=35dB. In black: true pulses. In blue and red: reconstruction with Cadzow denoising and the proposed algorithm, respectively, with 50 iterations. The blue pulses are not visible because they are hidden behind the red pulses. The computation time of each reconstruction process was 11s. Note the presence of two very close pulses with same amplitude's sign around position 0.51, two very close pulses with opposite amplitude's sign around position 0.2, and one pulse with very small amplitude around position 0.75. All the pulses have been correctly identified. The lowpass MSE with Cadzow denoising and the proposed algorithm is 6.5e-3 vs 6.1e-3, respectively.

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

[20] Y. T. Chan, J. M. M. Lavoie, and J. B. Plant, "A parameter estimation approach to estimation of frequencies of sinusoids," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, no. 2, pp. 214–219, Apr. 1981.

[21] D. W. Tufts and R. Kumaresan, "Frequency estimation of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, Sep. 1982.

[22] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 34, no. 5, pp. 1081–1089, 1986.

[23] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 3, pp. 378–391, Mar. 1989.

[24] M. D. Macleod, "Fast nearly ML estimation of the parameters of real or complex tones or resolved multiple tones," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 141–148, Jan. 1998.

[25] B. G. Quinn and E. J. Hannan, *The estimation and tracking of frequency*. Cambridge University Press, 2001.

[26] H. C. So, K. W. Chan, Y. T. Chan, and K. C. Ho, "Linear prediction approach for efficient frequency estimation of multiple real sinusoids: Algorithms and analysis," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2290–2305, Jul. 2005.

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

[28] 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.

[29] A. Prasad, D. Kundu, and A. Mitra, "Sequential estimation of the sum of sinusoidal model parameters," *Journal of Statistical Planning and Inference*, vol. 138, no. 5, pp. 1297–1313, May 2008.

[30] M. A. Lexa, M. E. Davis, and J. S. Thompson, "Reconciling compressive sampling systems for spectrally sparse continuous-time signals," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 155–171, Jan. 2012.

[31] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Trans. Signal Processing*, vol. 60, no. 7, pp. 3809–3822, Jul. 2012.

[32] P. Boufounos, V. Cevher, A. C. Gilbert, Y. Li, and M. J. Strauss, "What's the frequency, Kenneth?: Sublinear Fourier sampling off the grid," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, ser. Lecture Notes in Computer Science, 2012, vol. 7408, pp. 61–72.

[33] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," 2012, preprint arXiv:1207.6053.

[34] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, 2013, to appear.

[35] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high resolution frequencies estimation methods compared to the Cramér-Rao bounds," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 11, pp. 1703–1720, Nov. 1989.

[36] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér-Rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2140–2150, Dec. 1990.

[37] C. R. 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.

[38] 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, Sep. 2008.

[39] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observation," *IEEE Trans. Inform. Theory*, vol. 20, no. 5, pp. 591–598, 1974.

[40] J. A. Rice and M. Rosenblatt, "On frequency estimation," *Biometrika*, vol. 75, no. 3, pp. 477–484, 1988.

[41] P. Lemmerling and S. V. Huffel, "Analysis of the structured total least squares problem for Hankel/Toeplitz matrices," *Numerical Algorithms*, vol. 27, no. 1, pp. 89–114, 2001.

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

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

[44] 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.

[45] 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.

[46] J.-K. Hwang and Y.-C. Chen, "A combined detection-estimation algorithm for the harmonic-retrieval problem," *Signal Processing*, vol. 30, no. 2, pp. 177–197, 1993.

[47] 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.

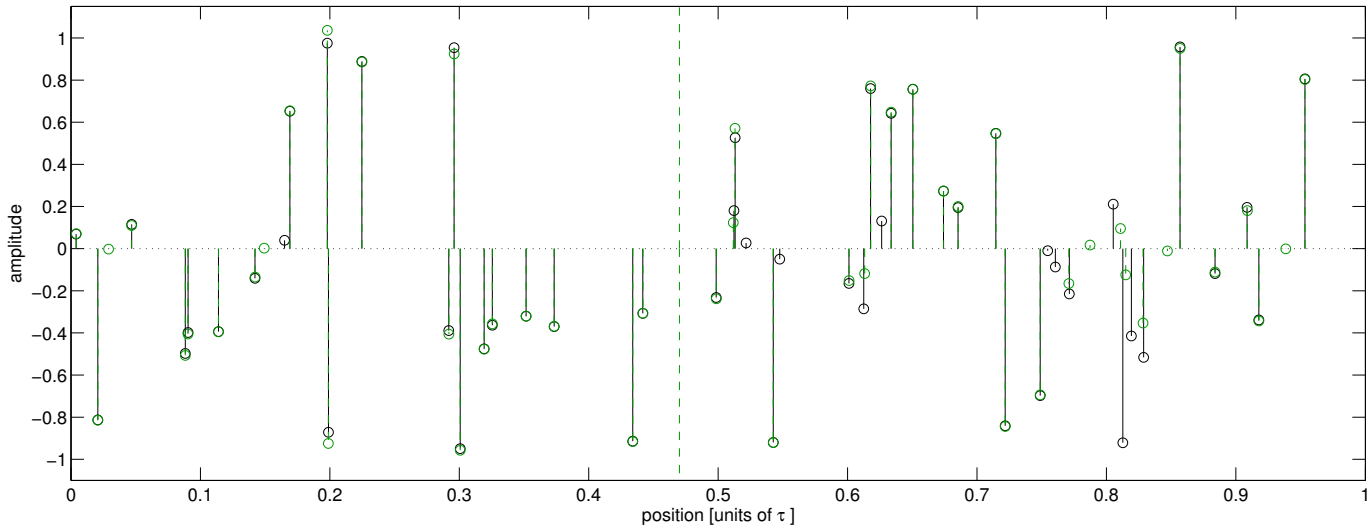


Fig. 7. Same setting as in Fig. 6. In black: true pulses. In green: reconstructed locations and amplitudes of the pulses, by solving the convex problem (25). The value $\lambda = 0.2$ giving the best visual result was used, with $\mu = 1.9$, $\gamma = 0.51\mu$. Note the presence of a vertical line around $t = 0.47$, because a pair of roots outside the unit circle is obtained at Step. 4 of the reconstruction procedure. Thus, the procedure has failed to produce $K = 50$ Dirac pulses at distinct locations.

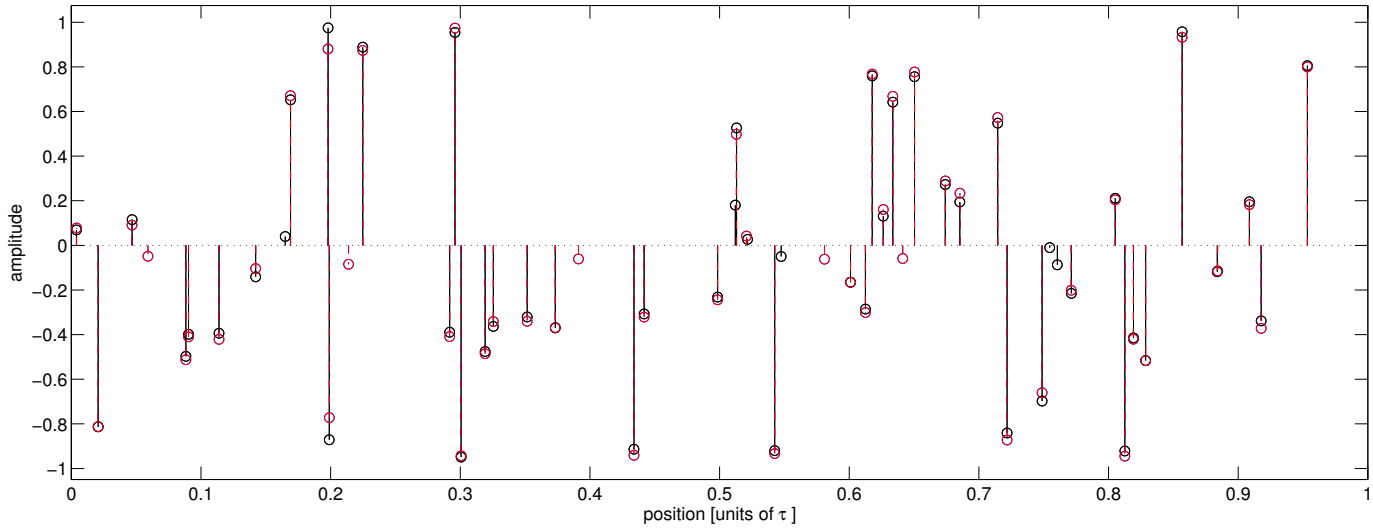


Fig. 8. Same setting as in Fig. 6, but with SNR=15dB. Five small pulses have been incorrectly identified, showing the intrinsic impossibility to distinguish small pulses and noise. The lowpass MSE with Cadzow denoising and the proposed algorithm is 109e-3 vs 108e-3, respectively.

[48] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Processing*, vol. 44, no. 2, pp. 281–295, 1996.

[49] 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.

[50] 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.

[51] G. M. Riche de Prony, "Essai expérimental et analytique : sur les lois de la dilatabilité de uides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool à diérentes températures," *Journal de l'école polytechnique*, vol. 1, no. 22, 1795.

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

[53] ——, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2012.

[54] J. W. Gillard and A. A. Zhigljavsky, "Optimization challenges in the structured low rank approximation problem," *Journal of Global Optimization*, 2012.

[55] 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.

[56] 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.

[57] 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., 2008, vol. 352, pp. 312–321.

[58] I. Markovsky and K. Usevich, "Software for weighted

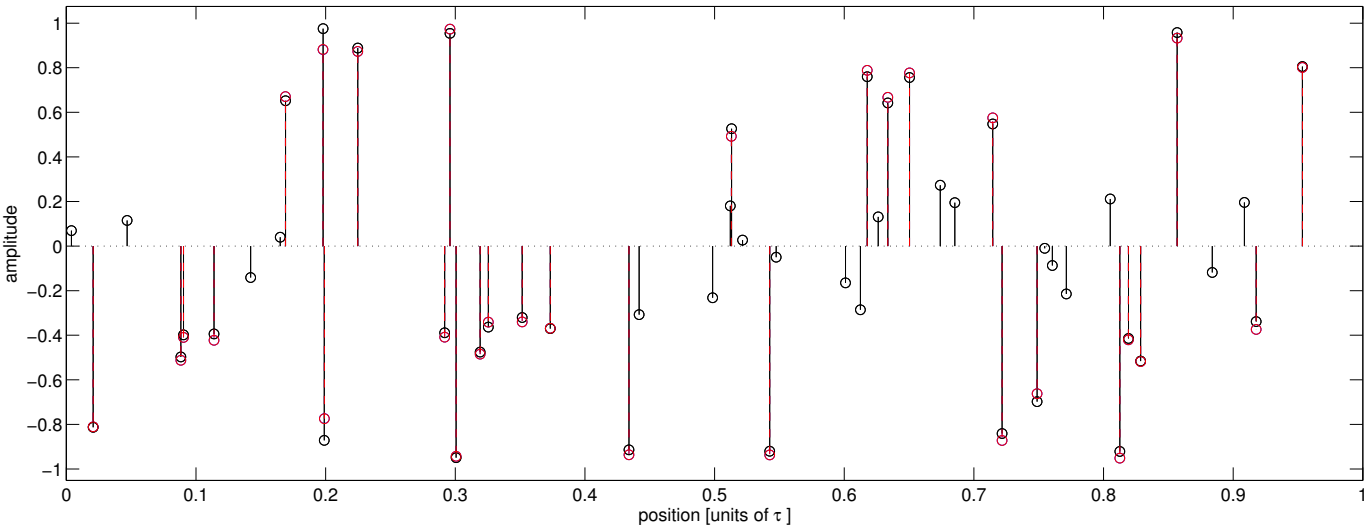


Fig. 9. The signal and the measurements are exactly the same as in Fig. 8, but the reconstruction is performed with $K = 30$, to illustrate the effects of underfitting. Interestingly, the 30 pulses with largest absolute amplitudes have been identified.

structured low-rank approximation," 2012, documention of a software package, downloadable at http://homepages.vub.ac.be/∼imarkovs/slra/software.html.

[59] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications in system identification and realization," 2012, preprint.

[60] B. De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3104–3113, 1994.

[61] J. Rosen, H. Park, and J. Glick, "Total least norm formulation and solution for structured problems," *SIAM J. Matrix Anal. Appl.*, vol. 17, pp. 110–126, 1996.

[62] S. V. Huel, H. Park, and J. B. Rosen, "Formulation and solution of structured total least norm problems for parameter estimation," *IEEE Trans. Signal Processing*, vol. 44, no. 10, Oct. 1996.

[63] S. Van Huffel, I. Markovsky, R. Vaccaro, and T. Söderström, "Total least squares and errors-in-variables modeling," *Signal Processing*, vol. 87, no. 10, pp. 2281–2282, 2007.

[64] S. Van Huffel, C. L. Cheng, N. Mastronardi, C. Paige, and A. Kukush, "Total least squares and errors-in-variables modeling," *Computational Statistics & Data Analysis*, vol. 52, pp. 1076–1079, 2007.

[65] I. Markovsky, D. M. Sima, and S. V. Huffel, "Total least squares methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 2, pp. 212–217, 2010.

[66] M. A. Hitz, "On computing nearest singular Hankel matrices," in *Proc. of the 2005 international symposium on Symbolic and algebraic computation (ISSAC)*, 2005, pp. 171–176.

[67] P. Lemmerling, N. Mastronardi, and S. V. Huffel, "Fast algorithm for solving the Hankel/Toeplitz structured total least squares problem," *Numerical Algorithms*, vol. 23, pp. 371–392, 2000.

[68] M. Schuermans, P. Lemmerling, and S. V. Huel, "Structured weighted low rank approximation," *Numerical Linear Algebra with Applications*, vol. 11, no. 5–6, pp. 609–618, June–Aug. 2004.

[69] M. Schuermans, "Weighted low rank approximation: Algorithms and applications," Ph.D. dissertation, Katholieke Universiteit Leuven, 2006.

[70] S. R. Khare, H. K. Pillai, and M. N. Belur, "Numerical algorithm for structured low rank approximation problem," in *Proc. of International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Jul. 2010.

[71] R. Borsdorf, "Structured matrix nearness problems: Theory and algorithms," Ph.D. dissertation, The University of Manchester, UK, Jun. 2012.

[72] 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.

[73] ——, "Total least squares, matrix enhancement, and signal processing," *Digital Signal Processing*, vol. 4, pp. 21–39, 1994.

[74] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.

[75] G. H. Golub and C. F. V. Loan, *Matrix computations*, 3rd ed. Johns Hopkins Univ. Press, 1996.

[76] P. L. Combettes and H. J. Trussell, "Method of successive projections for finding a common point of sets in metric spaces," *J. Optim. Theory Appl.*, vol. 67, pp. 487–507, 1990.

[77] N. E. Hurt, "Signal enhancement and the method of successive projections," *Acta Appl. Math.*, vol. 23, pp. 145–162, 1991.

[78] A. S. Lewis and J. Malick, "Alternating projections on manifolds," *Math. Oper. Res.*, vol. 33, 2008.

[79] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang, "Restricted normal cones and the method of alternating projections," 2012, preprint arXiv:1205.0318.

[80] D. R. Luke, "Prox-regularity of rank constraint sets and implications for algorithms," 2011, preprint arXiv:1112.0526.

[81] F. Andersson and M. Carlsson, "Alternating projections on non-tangential manifolds," 2011, preprint arXiv:1107.4055.

[82] F. Andersson, M. Carlsson, and P.-A. Ivert, "A fast alternating projection method for complex frequency estimation," 2011, preprint arXiv:1107.2028.

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

[84] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.

[85] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York: Springer-Verlag, 2010.

[86] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," *J. Opt. Soc. Am. A*, vol. 19, no. 7, pp. 1334–1345, Jul. 2002.

[87] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 188–209, 2005.

[88] V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," *Proc. National Academy Sciences USA*, vol. 104, no. 2, pp. 418–423, Jan. 2007.

[89] F. J. Aragón Artacho, J. M. Borwein, and M. K. Tam, "Recent results on Douglas–Rachford methods for combinatorial optimization problems," 2013, preprint arXiv:1305.2657.

[90] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," 2012, preprint arXiv:1204.0562.

[91] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. of American Control Conference*, Arlington, Virginia, Jun. 2001.

[92] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[93] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identication," *SIAM J. on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.

[94] I. Markovsky, "Data modeling using the nuclear norm heuristic," 2011, technical Report 21936.

[95] ——, "How effective is the nuclear norm heuristic in solving data approximation problems?" in *Proc. of IFAC Symposium on System Identification (SYSID)*, Brussels, Belgium, Jul. 2012.

[96] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," 2013, preprint arXiv:1303.4348.

[97] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.

[98] T. Routtenberg and J. Tabrikian, "Non-Bayesian periodic Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 61, no. 4, pp. 1019–1032, 2013.