



HAL
open science

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

Laurent Condat, Akira Hirabayashi

► **To cite this version:**

Laurent Condat, Akira Hirabayashi. Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements. 2012. hal-00759253v2

HAL Id: hal-00759253

<https://hal.science/hal-00759253v2>

Preprint submitted on 2 Dec 2012 (v2), last revised 20 Dec 2014 (v6)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Laurent Condat and A. Hirabayashi

Dec. 2, 2012

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 reformulated as structured low rank approximation of an appropriate matrix. To solve this difficult, even 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. Thus, the estimation error is smaller than with the classical heuristic method of alternating projections, a.k.a. Cadzow denoising, while sharing its speed and easiness of implementation.

Index Terms

Recovery of Dirac pulses, finite rate of innovation, maximum likelihood estimation, structured low rank approximation, alternating projections, Cadzow denoising, optimization, splitting

I. INTRODUCTION AND PROBLEM FORMULATION

Reconstruction of signals lying in shift-invariant spaces, including bandlimited signals and splines [1], has received long attention in sampling theory [2], [3]. 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 [4]–[7]. Indeed, the estimation of signals ruled by parsimonious models in terms of degrees of freedom has become of utmost importance in signal processing and engineering, as witnessed by intense research around the emerging paradigm of sparsity and compressed sensing.

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], [5], [8]–[11]. 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

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

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 consist of linear uniform noisy measurements $\{v_n\}_{n=0}^{N-1}$ on s , of the form

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

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

L. Condat is with the AGPiG Team of GIPSA-lab, a joint research unit of the CNRS and the Grenoble Institute of Technology, Grenoble, France. Contact: see <http://www.gipsa-lab.fr/~laurent.condat/>

A. Hirabayashi is with the Dept. Computer Science and Engineering, Yamaguchi University, Ube, 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.

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 [10]–[13]. 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 [10], which amounts to periodizing the signal s on the real line before sampling it with the sinc function:

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

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, for instance in ultrawideband communications [6] or to detect impulsive signals in biomedical applications [10]. This extension, or equivalently the choice of another sampling function φ , can be addressed without any difficulty, as described in details in [10]; therefore, to keep the derivations and notations simple, we do not address this general case in this paper.

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; that is, the ML estimate is given by minimizing the anti-log-likelihood of the data, with respect to the model parameters. In our case, since we have supposed the noise to be Gaussian, this corresponds to solving the nonlinear least-squares problem [14]:

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

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

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

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

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

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

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

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

Since the inverse discrete Fourier transform is unitary, up to a constant, the problem (9) can be rewritten as [14]:

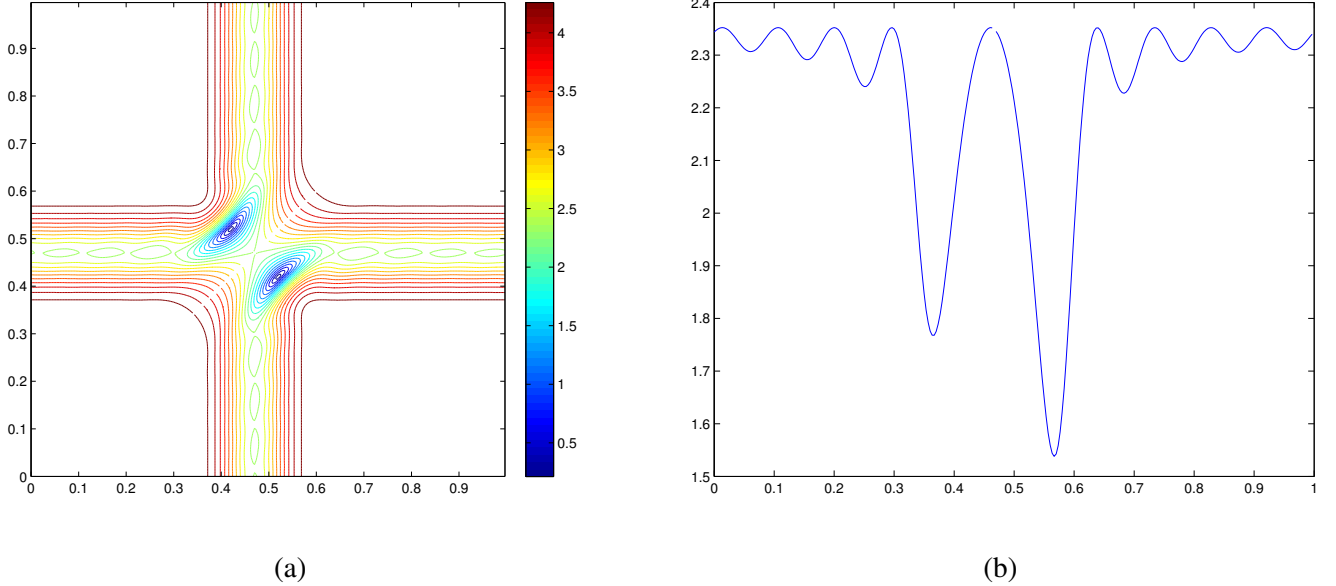


Fig. 1. (a) Contour plot showing the level lines of the cost function in (10), as a function of t'_1 and t'_2 , on the horizontal and vertical axes, respectively. (b) plot of the same cost function as a function of t'_2 , for $t'_1 = 0.47$. For given t'_1 and t'_2 , the values of a'_1 and a'_2 are the optimal values obtained by solving (21). The setting is the same as in Figs. 3 and 4 (a): the true unknown signal consists of $K = 2$ Dirac pulses with $\tau = 1$, $(t_1, t_2) = (0.42, 0.52)$ and $(a_1, a_2) = (1, 1)$; we have $N = 11$ noisy measurements with $\text{SNR}=15\text{dB}$.

Maximum likelihood (ML) estimation problem

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

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 [15]–[17]. This classical problem, which has a wide range of applications, e.g. in communications, radar, sonar, and geophysical seismology [17], has been studied extensively in signal processing [18]–[27]. From the perspective of control theory, spectral estimation was studied as well, with emphasis on the asymptotic optimality as $N \rightarrow +\infty$, see e.g. [28]; we are not in this context here, as the number N of available Fourier coefficients of the signal is fixed. Also, when $N \gg K$ and the locations t_k are not too close to each other, classical spectral estimation techniques like MUSIC [29], [30] and ESPRIT [20], [31], or greedy strategies [32], can be used; they are fast but statistically suboptimal. In this work, we investigate the ML estimation problem in its whole generality, without any simplifying assumption.

The optimal statistical properties of ML estimation for spectral estimation are well known [19], [21]–[23], [33]–[35]. 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 [36]–[40], see the example of Fig. 1. Stochastic optimization approaches can be applied [14], [41], [42], 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) [43]–[46]; see also [47], [48] for relationship with the well known expectation maximization (EM) algorithm. 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 that it does not necessitate any initialization.

III. 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 $\{\hat{v}_m\}_{m=-M}^M$ satisfy a linear *annihilating* difference equation [49], a known property which dates back to de Prony's

work in the eighteenth century [50]. That is, the convolution of the sequence of Fourier coefficients with the annihilating filter $\{h_k\}_{k=0}^K$ is identically zero:

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

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

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

In matrix form, the annihilation property is

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

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, named *structured low rank approximation* (SLRA) in the literature [51]–[53]:

Structured low rank approximation (SLRA) problem

$$\begin{aligned} \text{Find } \tilde{\mathbf{T}}_P \in \arg \min_{\mathbf{T}' \in \mathbb{C}^{N-P \times P+1}} \|\mathbf{T}' - \mathbf{T}_P\|_w^2 & \quad (14) \\ \text{s. t. } \mathbf{T}' \text{ is Toeplitz and } \text{rank}(\mathbf{T}') \leq K, & \end{aligned}$$

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

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

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

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

The SLRA problem (14), which consists in projecting a matrix in the intersection of a linear subspace and a nonconvex manifold, is claimed in several papers to be NP-hard [52], see also [54]–[57]. 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 $\tilde{\mathbf{T}}_P$. Moreover, for a low noise level, an algorithm converging to a local solution will find the global solution $\tilde{\mathbf{T}}_P$, as \mathbf{T}_P , $\tilde{\mathbf{T}}_P$ and the true noiseless matrix will be in the same catchment area of the cost function in (14).

As 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 procedure. The whole reconstruction process, also given in [5], is detailed in Fig. 2. In absence of noise,

it yields perfect reconstruction of the parameters. We also note that the estimation of the locations is decoupled from that of the amplitudes.

We must remark that the obtained estimates $\{(t_k, \tilde{a}_k)\}_{k=1}^K$ 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 $\tilde{\mathbf{T}}_P$ are centro-Hermitian, i.e. their entries satisfy

Reconstruction Procedure

Input: The Fourier coefficients $\{\hat{v}_m\}_{m=-M}^M$.

Output: the estimates $\{t_k\}_{k=1}^K$ and $\{\tilde{a}_k\}_{k=1}^K$ of the unknown positions $\{t_k\}_{k=1}^K$ and amplitudes $\{a_k\}_{k=1}^K$.

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:

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

Step 2. Solve the SLRA problem

$$\text{Find } \tilde{\mathbf{T}}_P \in \arg \min_{\mathbf{T}' \in \mathbb{C}^{N-P \times P+1}} \|\mathbf{T}' - \mathbf{T}_P\|_w^2 \quad \text{s. t. } \mathbf{T}' \text{ is Toeplitz and } \text{rank}(\mathbf{T}') \leq K, \quad (18)$$

where the weighted norm is defined in eqns. (15)–(16).

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

$$\tilde{\mathbf{T}}_K = \begin{pmatrix} \tilde{v}_{-M+K} & \cdots & \tilde{v}_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \tilde{v}_M & \cdots & \tilde{v}_{M-K} \end{pmatrix}. \quad (19)$$

Note that both $\tilde{\mathbf{T}}_P$ and $\tilde{\mathbf{T}}_K$ have rank at most K .

Step 4. Compute the right singular vector $\tilde{\mathbf{h}} = [\tilde{h}_0 \cdots \tilde{h}_K]^T$ of $\tilde{\mathbf{T}}_K$ corresponding to the singular value zero: in the SVD $\tilde{\mathbf{T}}_K = \mathbf{L}\mathbf{\Sigma}\mathbf{R}^H$, $\tilde{\mathbf{h}}$ corresponds to the last column of \mathbf{R} . Since, almost surely, $\tilde{\mathbf{T}}_K$ has rank exactly K , $\tilde{\mathbf{h}}$ is unique, up to a constant. 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

$$\tilde{t}_k = \frac{\tau}{2\pi} \arg_{[0, 2\pi[}(\tilde{z}_k), \quad \forall k = 1, \dots, K. \quad (20)$$

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

$$\tilde{\mathbf{U}}^H \tilde{\mathbf{U}} \tilde{\mathbf{a}} = \tilde{\mathbf{U}}^H \hat{\mathbf{v}}, \quad (21)$$

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

$$\tilde{\mathbf{U}} = \begin{pmatrix} e^{j2\pi M \tilde{t}_1 / \tau} & \cdots & e^{j2\pi M \tilde{t}_K / \tau} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi M \tilde{t}_1 / \tau} & \cdots & e^{-j2\pi M \tilde{t}_K / \tau} \end{pmatrix}. \quad (22)$$

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

$\hat{v}_{-m} = \hat{v}_m^*$ and $\tilde{v}_{-m} = \tilde{v}_m^*$, for every $m = -M, \dots, M$, where \cdot^* denotes complex conjugation. Consequently, the polynomial $\sum_{k=0}^K \tilde{h}_k z^k$ is self-inversive [58], 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 end this section by getting onto the state of the art of 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, collaborative filtering, global positioning from local distances, and the now famous Netflix problem, to mention a few; see references in [51], [52], [59]–[64]. 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 [51], [52], [65]–[70].

Several methods of numerical algebra have been proposed to obtain a local solution of a SLRA problem [39], [67], [71]–[76]. For instance, the iterative approach in [39] is based on high-level optimization routines, like 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 computing 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 [59]. 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, a simple alternative for SLRA is to use the heuristic method of *Cadzow denoising* [77], [78], a.k.a. alternating projections. This method is promoted in [5], [10] for the reconstruction of Dirac pulses, and represents the state of the art for this problem. It has also links with singular spectrum analysis (SSA) [79]–[83]. To describe the method, let us denote by \mathcal{M} the set of complex matrices of size $N - P \times P + 1$, endowed with the Frobenius norm. Let \mathcal{T} and \mathcal{R}_K be the closed subsets of \mathcal{M} of Toeplitz matrices and matrices of rank at most K , respectively. We denote by $P_\Omega : \mathcal{H} \rightarrow \Omega$ (a selection of) the closest-point projection onto a closed subset Ω of a metric space \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. Then, with $\mathcal{H} = \mathcal{M}$, “Toeplitzation” $P_{\mathcal{T}}$ simply consists in averaging along the diagonals of the matrix and $P_{\mathcal{R}_K}$ corresponds to SVD truncation, according to the classical Schmidt-Eckart-Young theorem [84] [85, theorem 2.5.3]: if a matrix \mathbf{X} has SVD $\mathbf{X} = \mathbf{L}\mathbf{\Sigma}\mathbf{R}^H$, 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} . We now have all the ingredients to describe the Cadzow denoising iterative algorithm, which takes the simple form:

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

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

This algorithm seems to always converge in practice to a matrix $\tilde{\mathbf{T}}_P \in \mathcal{T} \cap \mathcal{R}_K$, i.e. a Toeplitz matrix of rank at most K . However, contrary to what is sometimes claimed in papers, this convergence has not been proved theoretically. A few convergence results are available in the literature for the alternating projections on nonconvex sets [86]–[89]. To our knowledge, the most advanced results for the Toeplitz and low rank case are the recent ones in [90]–[92]. Moreover, even if Cadzow denoising converges to a Toeplitz matrix $\tilde{\mathbf{T}}_P$ of rank at most K , there is no reason for $\tilde{\mathbf{T}}_P$ to be close, in any sense, to the matrix \mathbf{T}_P from which the algorithm starts. In particular, the experiments in [39], [40], [65] show that the 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). 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.

We end this section by mentioning a possible approach to circumvent the difficulty of SLRA. It consists in

replacing the nonconvex rank by its convex surrogate, the nuclear norm, in the problem formulation. That is, instead of the SLRA problem (14), we solve

$$\text{Find } \tilde{\mathbf{T}}_P \in \arg \min_{\mathbf{T}' \in \mathbb{C}^{N-P \times P+1}} \frac{1}{2} \|\mathbf{T}' - \mathbf{T}_P\|_w^2 + \lambda \|\mathbf{T}'\|_* \quad \text{s. t. } \mathbf{T}' \text{ is Toeplitz}, \quad (23)$$

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. This approach, initiated in [93], has become popular, due to the existence of fast polynomial-time semi-definite programming algorithms and good results in many cases [61]–[64], [94], [95]. However, in our setting, two pulses can be arbitrarily close to each other, so that the measurements can be highly coherent. Consequently, the estimation quality of the nuclear-norm approach is significantly degraded [96], [97], see Fig. 7. Moreover, the choice of the right value of λ to use in a given situation is very difficult.

IV. A NEW ITERATIVE OPTIMIZATION METHOD FOR SLRA

We consider the generic optimization problem:

$$\text{Find } \tilde{x} \in \arg \min_{x \in \mathcal{H}} F(x) \quad \text{s.t. } x \in \Omega_1 \cap \Omega_2, \quad (24)$$

where \mathcal{H} is a real Hilbert space of finite dimension, Ω_1 and Ω_2 are two closed subsets of \mathcal{H} such that $\Omega_1 \cap \Omega_2 \neq \emptyset$, and $F : \mathcal{H} \rightarrow \mathbb{R}$ 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 proposed algorithm to solve (24) 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 $i \geq 0$,

$$\begin{cases} x^{(i+1)} = P_{\Omega_1}(s^{(i)} + \gamma(x^{(i)} - s^{(i)}) - \mu \nabla F(x^{(i)})) \\ s^{(i+1)} = s^{(i)} - x^{(i+1)} + P_{\Omega_2}(2x^{(i+1)} - s^{(i)}) \end{cases}.$$

The following convergence results are corollaries of more general results derived in [98].

Theorem 1. *In (24), suppose that the sets Ω_1 and Ω_2 are convex. In the algorithm, suppose that $2\gamma > \beta\mu$. Then, the sequence $(x^{(i)})_{i \in \mathbb{N}}$ generated by the algorithm converges to some element \tilde{x} solution of the problem (24).*

Theorem 2. *In (24), suppose that the sets Ω_1 and Ω_2 are convex and that $F = 0$. In the algorithm, suppose that $\gamma = \mu = 0$. Then, the algorithm reverts to the Douglas-Rachford splitting algorithm, well known in optimization theory [99]. Consequently, the sequence $(x^{(i)})_{i \in \mathbb{N}}$ generated by the algorithm converges to some element $\tilde{x} \in \Omega_1 \cap \Omega_2$.*

In absence of convexity, these results do not apply, so that we will use the algorithm as a heuristic. The SLRA problem (14) can be recast as an instance of (24) 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 $\langle \mathbf{X}, \mathbf{X}' \rangle = \sum_{i,j} x_{i,j} x'_{i,j} \in \mathbb{R}$; $\Omega_1 = \mathcal{R}_K$ is the closed nonconvex subset of \mathcal{H} of matrices with rank at most K ; $\Omega_2 = \mathcal{T}$ is the linear subspace of \mathcal{H} of Toeplitz matrices; $F(\mathbf{X}) = \frac{1}{2} \|\mathbf{X} - \mathbf{T}_P\|_w^2$, 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 $\beta = \max(\{w_{i,j}\}) = 1$.

Cadzow denoising and the Douglas-Rachford algorithm are two heuristic methods, which aim at finding an element in the intersection of two closed subsets of a metric space, by alternately enforcing closeness to the two constraint sets. 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 [100]–[103]. The Douglas-Rachford iteration generally outperforms the alternating projections for these applications [100], [102], [103]. The motivation for applying the new proposed algorithm, which can be viewed as a Douglas-Rachford iteration with an adequately plugged gradient descent, stems from this practical success. However, there are almost no convergence results for the Douglas-Rachford method in the nonconvex case [104]. We leave for future work the convergence study of the proposed algorithm. Especially, we would like to prove that

if the algorithm converges to some element, then this element is necessarily a local solution to (24), i.e. a local minimizer of the cost function F lying in $\Omega_1 \cap \Omega_2$. As a first step in this direction, it is easy to see that for every fixed point (\tilde{x}, \tilde{s}) of the algorithm, $\tilde{x} \in \Omega_1 \cap \Omega_2$.

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. But, in such cases, running again the algorithm with lower values of μ and γ —for instance, dividing them by three—seemed sufficient to obtain convergence; this is what we did to generate the results of Fig. 3. 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 at most K , which is a local minimizer of the weighted Frobenius cost function. More precisely, we checked that the positions $\{\tilde{t}_k\}_{k=1}^K$ 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 [65, 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 position 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

$$Q(z) = \sum_{m=0}^{2M} (m - M) \hat{v}_{m-M}^* z^m. \quad (25)$$

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 try to fit one Dirac pulse on this residual sequence. We observed empirically that, when \tilde{z}_{k_0} in (20) has unit amplitude, it is always a root of the corresponding polynomial Q .

V. EXPERIMENTAL RESULTS

The reconstruction procedure described in Sect. III was implemented in Matlab, with the choice between Cadzow denoising and the proposed algorithm to solve the SLRA problem. The code is freely available on the webpage of the first author. The results obtained with this code for several experiments are illustrated in Figs. 3–10. We used $P = M$ in all cases, because we observed faster convergence for the two algorithms in that case. In Figs. 4–6, 8–10, $\tau = 1$ and we set $\mu = 0.1$ and $\gamma = 0.51\mu$. Actually, the value of μ yielding fastest convergence depended on the experiment and the noise level, and we could not come to a rule for tuning this parameter.

The results in Figs. 3–10 show that the two algorithms behave quite similarly; in a given setting, they succeed and fail in estimating the same Dirac pulses, in average. We could not find an experimental setting where the two methods would behave differently and consistently for different noise realizations. This is remarkable, because Cadzow denoising does not minimize any criterion, while our method aims at reaching the optimal maximum-likelihood estimate. However, as is clearly visible in Fig. 3, in average, the precision in the estimated positions is higher with the proposed method; the error was about 10% lower in average in comparison with Cadzow denoising, for a large range of values of the signal-to-noise ratio (SNR). Especially, contrary to what is claimed in the literature, Cadzow denoising does not reach the ML estimate when the noise level tends to zero. The price to pay for the optimality of our method is that a bit more iterations are needed to achieve convergence, for large values of K and N . In the experiment of Fig. 3, the proposed algorithm yields the exact, *global* solution of the SLRA problem (14), which is the ML estimate, for SNR above 12dB. We checked this property empirically by solving (10) with brute force, by exhaustive search over the space of parameters.

We note that the choice of a criterion to quantify the estimation error between the true unknown signal and the estimated signal is far from trivial. Using an analog distance like $\|s - \tilde{s}\|_{L_p}$ does not make sense for streams of Dirac distributions. 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 what is the best way to assign each estimated pulse to a true pulse? Sorting the values in increasing order as $0 \leq t_1 < \dots < t_k < \tau$ is incorrect, because of the intrinsic periodicity of the setting: the boundaries at $t = 0$ and $t = \tau$ are identified. In Fig. 3, we considered the best assignment between the two possible permutations and computed the mean squared periodic error (MSPE). But for more than two pulses, this best assignment is not stable and can be irrelevant, because the insertion of a pulse with small amplitude, which does not correspond to a true pulse, at some position, can drastically modify the MSPE. Moreover, the MSPE on the positions is blind with respect to the amplitudes,

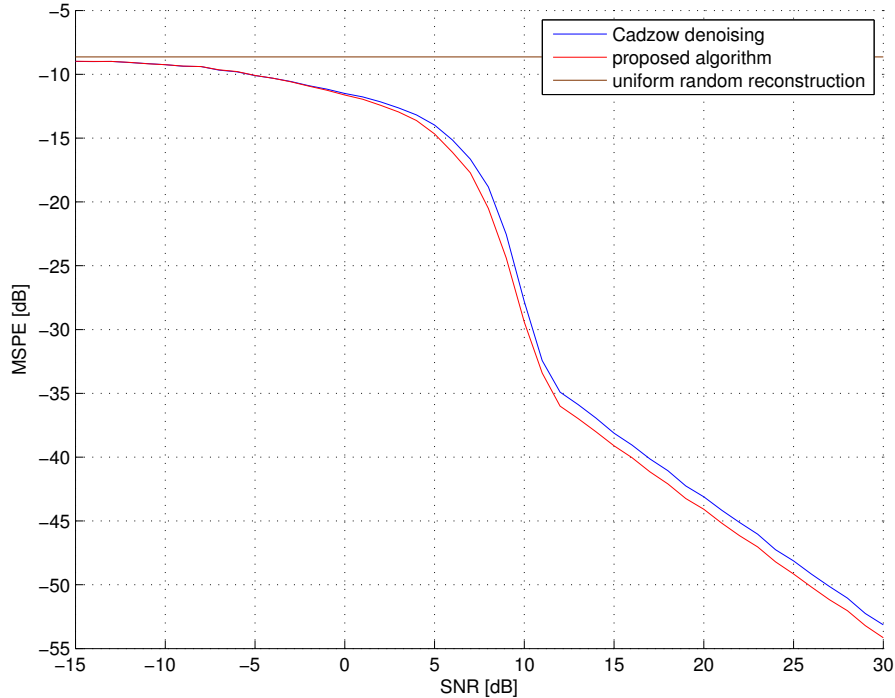


Fig. 3. Plot in log-log scale of the mean squared periodic error (MSPE) $\min((\tilde{t}_1 - t_1)_\tau^2 + (\tilde{t}_2 - t_2)_\tau^2, (\tilde{t}_1 - t_2)_\tau^2 + (\tilde{t}_2 - t_1)_\tau^2)$, where $(x)_\tau = ((x + \frac{\tau}{2}) \bmod \tau) - \frac{\tau}{2}$, for the estimation of the locations of $K = 2$ Dirac pulses, averaged over 10,000 noise realizations for every integer value of the signal-to-noise ratio (SNR). The true parameters were $(t_1, t_2) = (0.42, 0.52)$ and $(a_1, a_2) = (1, 1)$, 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 in average. 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 $[0, \tau]$, with MSPE $\frac{41}{300}$.

while we would like an error on the small pulses to have less weight than on the large pulses. Finally, the ℓ_2 distance between the true noiseless Fourier coefficients and the estimated ones $\{\tilde{v}_m\}_{m=-M}^M$ in (19) can be a good numerical criterion, but it does not have a natural interpretation.

VI. 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 discussed several aspects of the problem and gave an overview of many relevant results spread in the literature. We proposed a new heuristic optimization algorithm, which converges to a local solution of the NP-hard nonconvex problem. The 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, including the computation of Cramér-Rao bounds [5], [9], [105], [106], is currently led by the authors.

REFERENCES

- [1] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.
- [2] —, "Sampling—50 Years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
- [3] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Processing Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.
- [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, Jun. 2002.
- [5] 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.
- [6] 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.

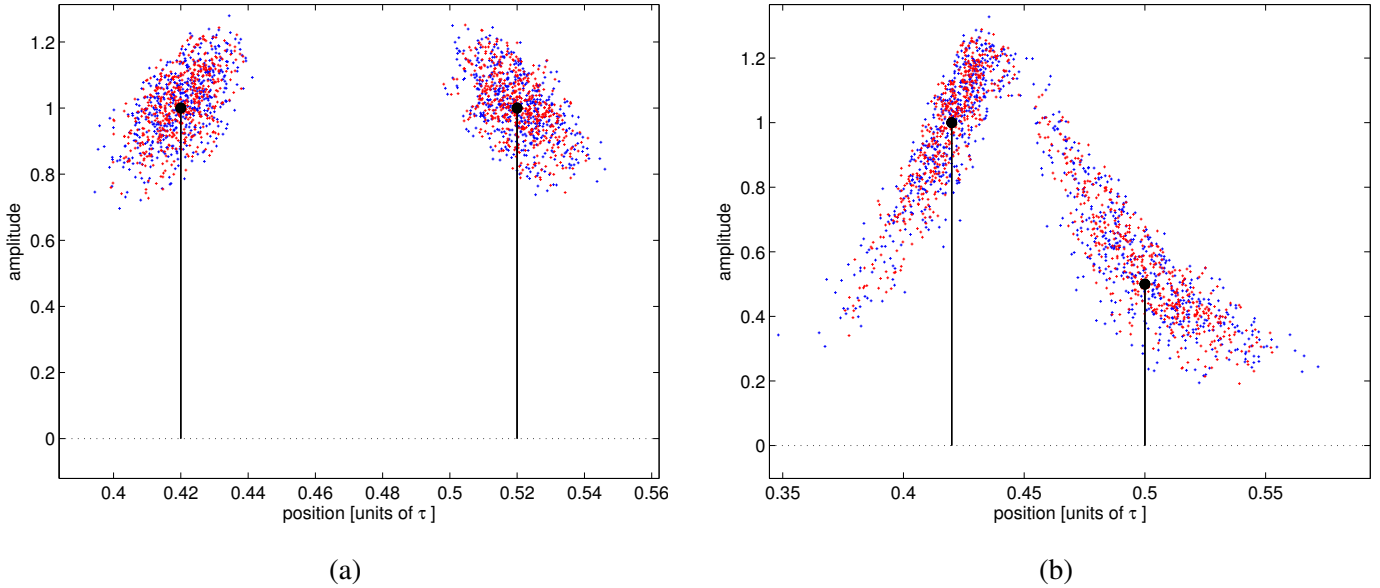


Fig. 4. The signal consists of $K = 2$ Dirac pulses, with $N = 11$ noisy measurements and $\text{SNR}=15\text{dB}$. (a): same setting as in Fig. 3, with $(t_1, t_2) = (0.42, 0.52)$ and $(a_1, a_2) = (1, 1)$. (b): $(t_1, t_2) = (0.42, 0.5)$ and $(a_1, a_2) = (1, 0.5)$. In black: true pulses. In blue and red: reconstructed positions and amplitudes of the pulses with Cadzow denoising and the proposed method, respectively, for 500 different noise realizations. Although hardly visible, the proposed method yields smaller estimation errors, with points clouds slightly less dispersed. Note that there is no reason that the estimation be unbiased. The computation time for one reconstruction, with 50 iterations, was 14ms.

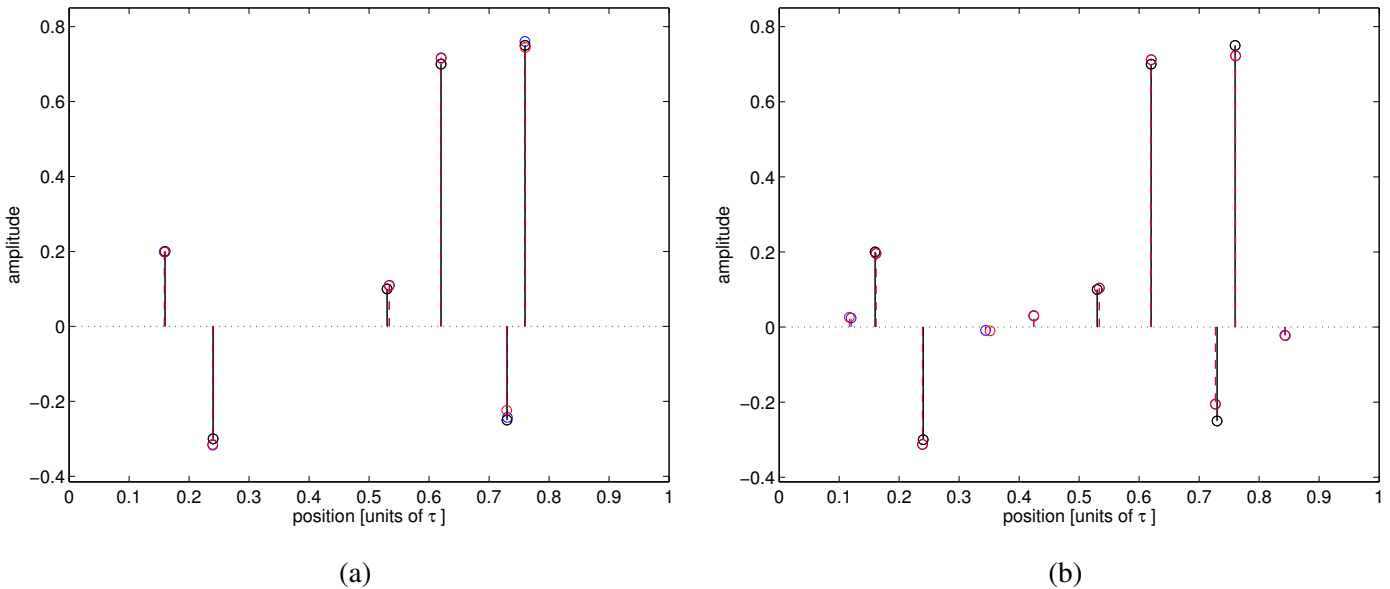


Fig. 5. The signal consists of 6 Dirac pulses. We have $N = 25$ noisy measurements with $\text{SNR}=25\text{dB}$. In black: true pulses. In blue and red: reconstructed positions and amplitudes of the pulses with Cadzow denoising and the proposed method, respectively, for one particular noise realization. The computation time for one reconstruction, with 50 iterations, was 19ms. (a) Reconstruction with $K = 6$. (b) Reconstruction with $K = 10$, to illustrate the effects of overfitting: the methods have found imaginary pulses in noise.

- [7] 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.
- [8] 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.
- [9] 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.
- [10] R. Tur, Y. C. Eldar, and Z. Friedman, “Innovation rate sampling of pulse streams with application to ultrasound imaging,” *IEEE Trans.*

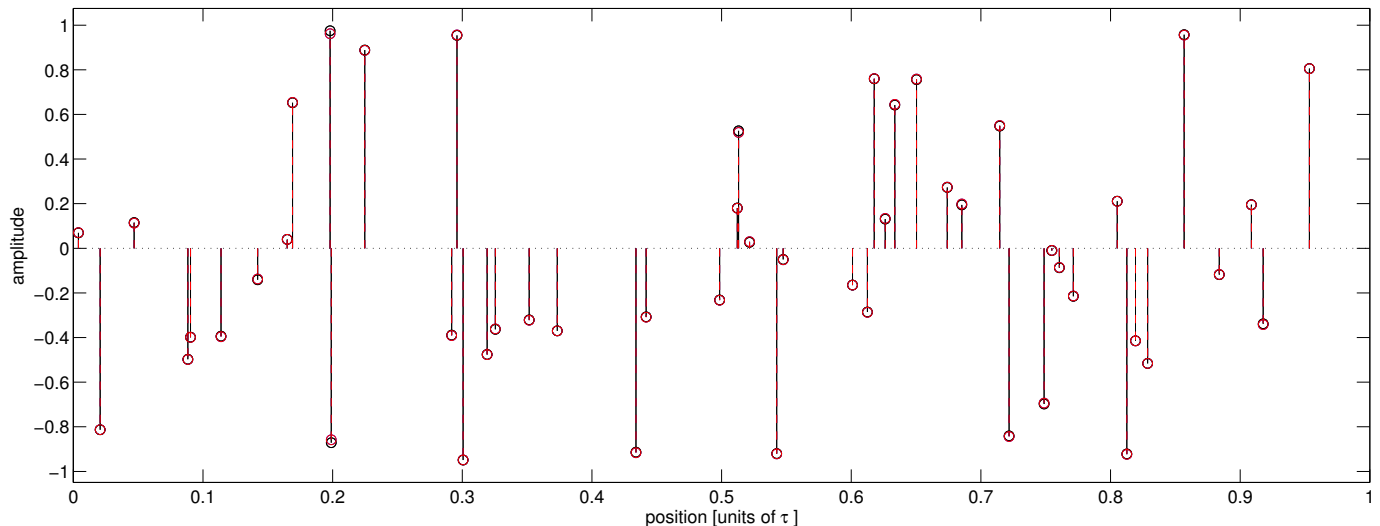


Fig. 6. The signal consists of $K = 50$ Dirac pulses, with $N = 1001$ noisy measurements and $\text{SNR}=35\text{dB}$. In black: true pulses. In blue and red: reconstruction with Cadzow denoising and the proposed method, respectively, with 50 iterations. The blue pulses are not visible because they are 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.

Signal Processing, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.

- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] Special issue on spectral estimation, *Proc. IEEE*, vol. 70, no. 9, Sep. 1982.
- [16] P. Stoica, "List of references on spectral line analysis," *Signal Processing*, vol. 31, no. 3, pp. 329–340, Apr. 1993.
- [17] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, NJ, 2005.
- [18] 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.
- [19] 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.
- [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] 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.
- [22] 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.
- [23] 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.
- [24] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Processing*, vol. 39, no. 8, pp. 1836–1847, Aug. 1991.
- [25] 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.
- [26] B. G. Quinn and E. J. Hannan, *The estimation and tracking of frequency*. Cambridge University Press, 2001.
- [27] 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.
- [28] M. Hou, "Parameter identification of sinusoids," *IEEE Trans. Automat. Contr.*, vol. 52, no. 2, pp. 467–472, Feb. 2012.
- [29] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [30] G. Bienvenu and L. Kopp, "Adaptivity to background noise spatial coherence for high resolution passive methods," in *Proc. of IEEE ICASSP*, vol. 5, Apr. 1980, pp. 307–310.

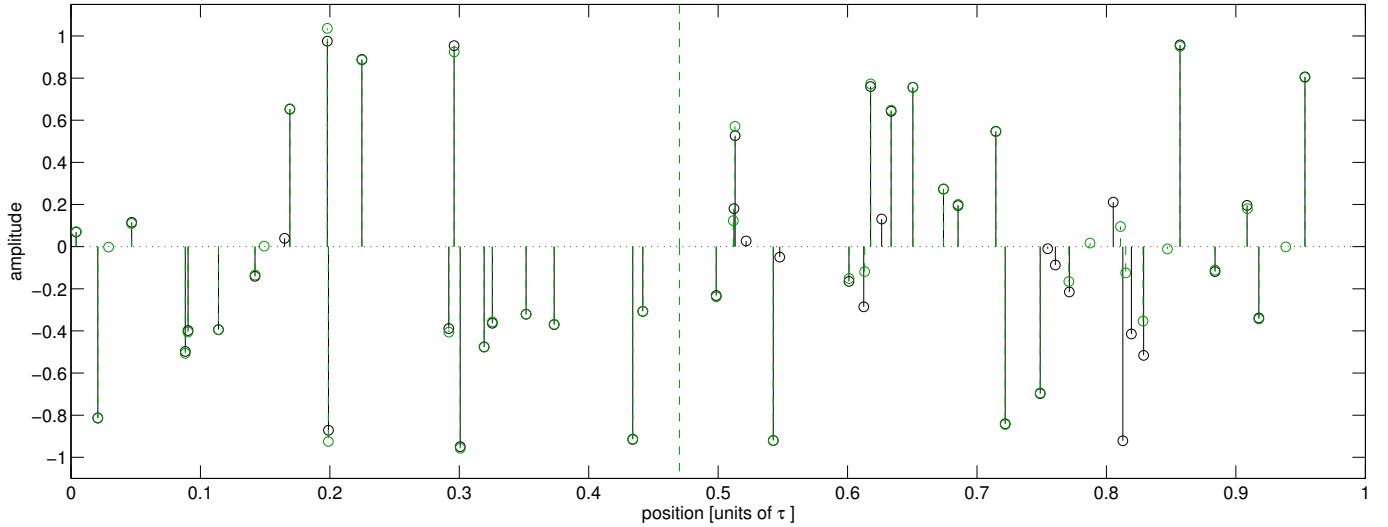


Fig. 7. Same setting as in Fig. 6. In black: true pulses. In green: reconstructed positions and amplitudes of the pulses, by solving the convex problem (23), instead of (14). The proposed algorithm is used, by replacing hard-thresholding $P_{\mathcal{R}_K}$ of the singular values by soft-thresholding; that is, each of the diagonal elements $\{\sigma_{i,i}\}_{i=1}^{P+1}$ of the matrix Σ is replaced by $\max(\sigma_{i,i} - \mu\lambda, 0)$. Convergence of this proximal algorithm to a solution of (23) is guaranteed [98]. The value $\lambda = 0.1$ giving the best visual result was used, with $\mu = 1.9$, $\gamma = 0.51\mu$. Contrary to Cadzow denoising and the proposed approach whose convergence rate is observed linear, convergence in this convex setting is much slower; we used 1,000 iterations to generate this plot, in 260s. 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.

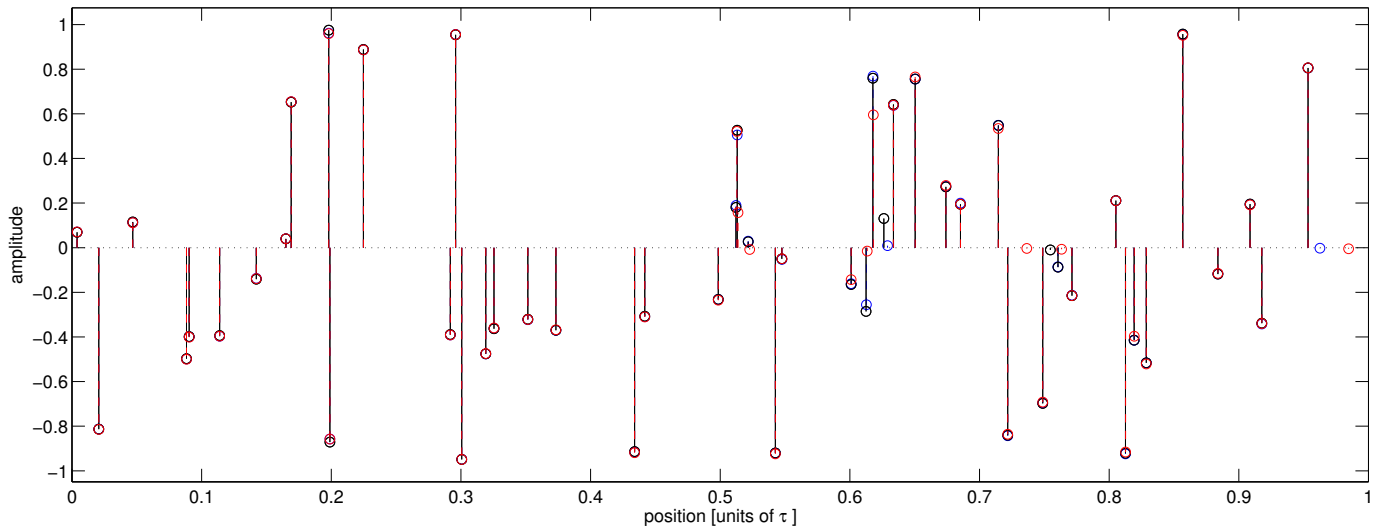


Fig. 8. Same setting as in Fig. 6 but with only 5 iterations, when convergence of the denoising process to a Toeplitz matrix of rank K has not been achieved. Consequently, a few small pulses have not been correctly identified. Note that Cadzow denoising needs less iterations to converge in this setting, hence the result here is slightly better. The computation time for each reconstruction process was 1.2s.

- [31] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [32] 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.
- [33] 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.
- [34] C. R. Rao and L. C. Zhao, "Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals," *IEEE Trans.*

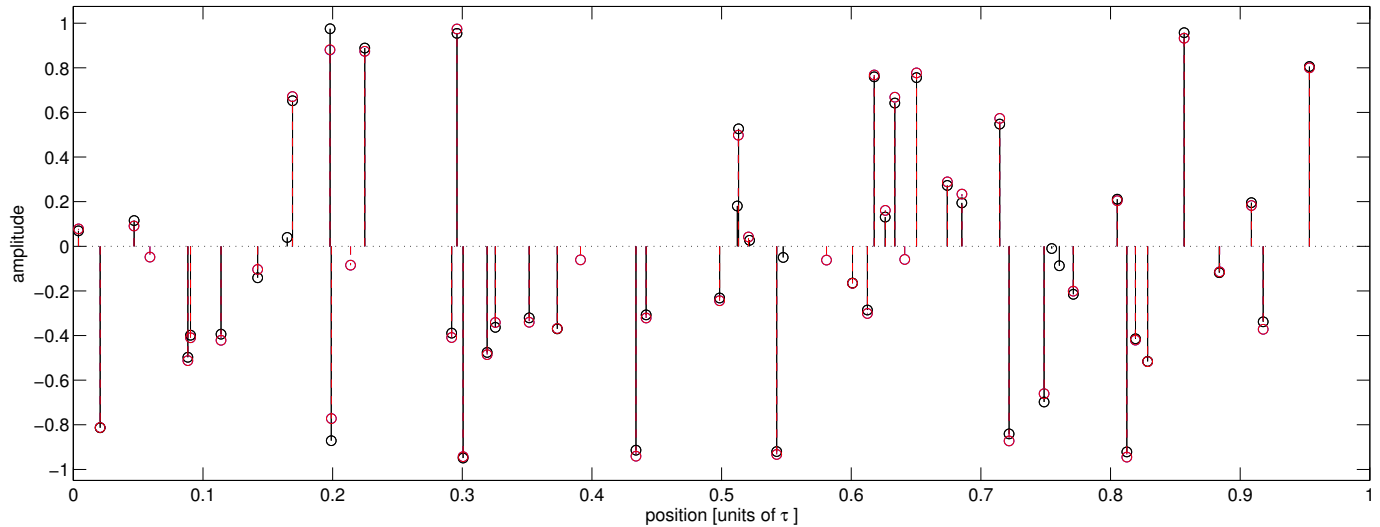


Fig. 9. Same setting as in Fig. 6 but with SNR=15dB. Five small pulses have been incorrectly identified, showing the intrinsic impossibility to distinguish small pulses and noise.

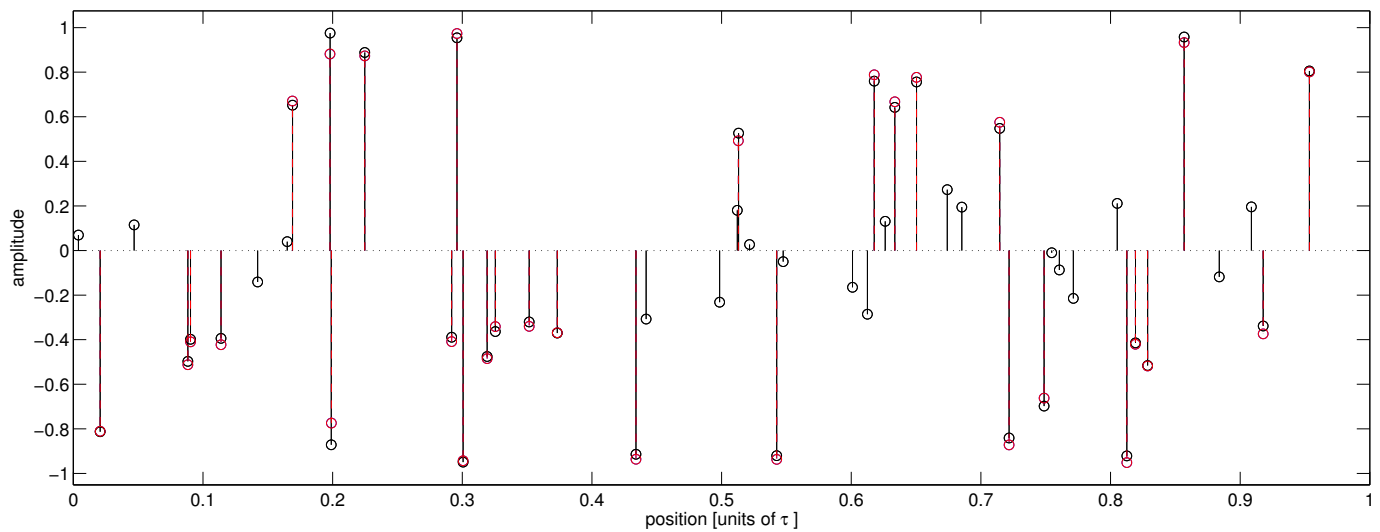


Fig. 10. The signal and the measurements are exactly the same as in Fig. 9, but the reconstruction is performed with $K = 30$, to illustrate the effects of underfitting. Interestingly, the 30 pulses with largest absolute amplitudes have been identified.

Signal Processing, vol. 41, no. 3, pp. 1461–1464, Mar. 1993.

- [35] 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.
- [36] 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.
- [37] J. A. Rice and M. Rosenblatt, “On frequency estimation,” *Biometrika*, vol. 75, no. 3, pp. 477–484, 1988.
- [38] 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.
- [39] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, “Structured low rank approximation,” *Linear Algebra Appl.*, vol. 366, pp. 157–172, 2003.
- [40] J. Gillard and A. Zhigljavsky, “Analysis of structured low rank approximation as an optimization problem,” *Informatica*, vol. 22, no. 4, pp. 489–505, 2011.
- [41] 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.

- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] S. Chrétien and A. O. Hero, “On EM algorithms and their proximal generalizations,” *ESAIM: Probability and Statistics*, vol. 12, pp. 308–326, 2008.
- [49] 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.
- [50] 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.
- [51] I. Markovsky, “Structured low-rank approximation and its applications,” *Automatica J. IFAC*, vol. 44, no. 4, pp. 891–909, 2008.
- [52] —, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2012.
- [53] J. W. Gillard and A. A. Zhigljavsky, “Optimization challenges in the structured low rank approximation problem,” *Journal of Global Optimization*, 2012.
- [54] B. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Computing*, vol. 25, no. 2, pp. 227–234, 1995.
- [55] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] I. Markovsky and K. Usevich, “Software for weighted structured low-rank approximation,” ECS, Univ. of Southampton, Tech. Rep. 339974, 2012, documentation of a software package, downloadable at <http://users.ecs.soton.ac.uk/im/slra.html>.
- [60] K. Lee and Y. Bresler, “ADMIRA: Atomic decomposition for minimum rank approximation,” *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4402–4416, Sep. 2010.
- [61] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 712–772, 2009.
- [62] E. J. Candès and Y. Plan, “Matrix completion with noise,” *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [63] J.-F. Cai, E. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.
- [64] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, “Hankel matrix rank minimization with applications in system identification and realization,” 2012, preprint.
- [65] 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.
- [66] 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.
- [67] 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.
- [68] 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.
- [69] 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.
- [70] 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.
- [71] 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.
- [72] 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.
- [73] 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.
- [74] M. Schuermans, “Weighted low rank approximation: Algorithms and applications,” Ph.D. dissertation, Katholieke Universiteit Leuven, 2006.
- [75] 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.

- [76] R. Borsdorf, “Structured matrix nearness problems: Theory and algorithms,” Ph.D. dissertation, The University of Manchester, UK, Jun. 2012.
- [77] 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.
- [78] —, “Total least squares, matrix enhancement, and signal processing,” *Digital Signal Processing*, vol. 4, pp. 21–39, 1994.
- [79] J. W. Gillard, “Cadzow’s basic algorithm, alternating projections and singular spectrum analysis,” *Statistics and Its Interface*, vol. 3, no. 3, pp. 333–343, 2010.
- [80] N. Golyandina, “On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods,” 2010, preprint ArXiv:1005.4374.
- [81] T. J. Harris and H. Yuan, “Filtering and frequency interpretations of Singular Spectrum Analysis,” *Physica D: Nonlinear Phenomena*, vol. 239, no. 20–22, pp. 1958–1967, 2010.
- [82] A. Korobeynikov, “Computation- and space-efficient implementation of SSA,” *Statistics and Its Interface*, vol. 3, pp. 357–368, 2010.
- [83] A. Zhigljavsky, “Singular spectrum analysis for time series,” in *International Encyclopedia of Statistical Science*. Springer, 2011, pp. 1335–1337.
- [84] C. Eckart and G. Young, “The approximation of one matrix by another of lower rank,” *Psychometrika*, vol. 1, pp. 211–218, 1936.
- [85] G. H. Golub and C. F. V. Loan, *Matrix computations*, 3rd ed. Johns Hopkins Univ. Press, 1996.
- [86] 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.
- [87] N. E. Hurt, “Signal enhancement and the method of successive projections,” *Acta Appl. Math.*, vol. 23, pp. 145–162, 1991.
- [88] A. S. Lewis and J. Malick, “Alternating projections on manifolds,” *Math. Oper. Res.*, vol. 33, 2008.
- [89] 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.
- [90] D. R. Luke, “Prox-regularity of rank constraint sets and implications for algorithms,” 2011, preprint arXiv:1112.0526.
- [91] F. Andersson and M. Carlsson, “Alternating projections on non-tangential manifolds,” 2011, preprint arXiv:1107.4055.
- [92] F. Andersson, M. Carlsson, and P.-A. Ivert, “A fast alternating projection method for complex frequency estimation,” 2011, preprint arXiv:1107.2028.
- [93] 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.
- [94] Z. Liu and L. Vandenberghe, “Interior-point method for nuclear norm approximation with application to system identification,” *SIAM J. on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.
- [95] B. Bhaskar and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” in *Proc. of the 49th Annual Allerton Conference*, 2011.
- [96] I. Markovsky, “Data modeling using the nuclear norm heuristic,” 2011, technical Report 21936.
- [97] —, “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.
- [98] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” 2011, preprint hal-00609728.
- [99] 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.
- [100] 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.
- [101] 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.
- [102] 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.
- [103] J. Schaad, “Modeling the 8-Queens problem and Sudoku using an algorithm based on projections onto nonconvex sets,” Master’s thesis, The University of British Columbia, Aug. 2010.
- [104] F. J. Aragón Artacho and J. M. Borwein, “Global convergence of a non-convex Douglas-Rachford iteration,” Mar. 2012, preprint arXiv:1203.2392.
- [105] C. R. Rao and L. C. Zhao, “Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals,” *IEEE Trans. Signal Processing*, vol. 41, pp. 1461–1464, Mar. 1993.
- [106] E. Dilaveroglu, “Nonmatrix Cramér-Rao bound expressions for high-resolution frequency estimators,” *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 463–474, 1998.