A novel algorithm for the identification of dirac impulses from filtered noisy measurements

Sylvain Meignen, Quentin Legros, Yoann Altmann, Steve Mclaughlin

A Novel Algorithm for the Identification of Dirac Impulses from Filtered Noisy Measurements

Sylvain Meignen [1], Quentin Legros [2], Yoann Altmann [3], and Steve McLaughlin[4] [5]
[1]Jean Kuntzmann Laboratory- Bâtiment IMAG,
Université Grenoble Alpes
700 Avenue Centrale
Campus de Saint Martin d'Hères
38401 Domaine Universitaire de Saint-Martin-d'Hères, France
[2,3,4]Heriot-Watt University,
School of Engineering and Physical Sciences (EPS)
Edinburgh, EH14 4AS, Scotland, UK

Abstract

In this paper we address the recovery of a finite stream of Dirac pulses from noisy lowpass-filtered samples in the discrete-time setting. While this problem has been successfully addressed for the noiseless case using the concept of signals with finite rate of innovation, such techniques are not efficient in the presence of noise. In the FRI framework, the determination of the location of Dirac pulses is based on the singular value decomposition of a matrix whose rank in the noise-free case equals the number of Dirac pulses and the signal can be related to the non zero singular values. However, in noisy situations this matrix becomes full rank and the singular value decomposition is subject to *subspace swap*, meaning some singular values associated with noise become larger than some related to the signal. This phenomenon has been recognized as the reason for performance breakdown in the method. The goal of this paper is to propose a novel algorithm that limits the alteration of these singular values in the presence of noise, thus significantly improving the estimation of Dirac pulses.

Keywords: Optimal Sampling, Fourier Analysis, Finite Rate of Innovation, Sparse deconvolution.

[1]Email:sylvain.meignen@univ-grenoble-alpes.fr
[2]Email:qll1@hw.ac.uk
[3]Email:Y.Altmann@hw.ac.uk
[4]Email:S.McLaughlin@hw.ac.uk
[5]This work was partly supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1, the MOD University Defence Research Collaboration (UDRC) in Signal Processing and the UK Royal Academy of Engineering under the Research Fellowship Scheme (RF201617/16/31). This work has also been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir.

1. Introduction

In this paper, we examine the problem of estimating the parameters of a signal composed of a sum of Dirac pulses (DPs) from noisy lowpass filtered samples. This can be viewed as an idealized super-resolution problem which consists in trying to recover high-resolution information from coarse scale measurements. There is a vast literature on this subject and on applications ranging over optical imaging [1], astronomy [2], medical imaging [3] to microscopy [4]. Assuming the lowpass filtered samples are associated with a continuous-time setting, many different methods have been designed including those based on total variation of measures [5] [6] [7] [8] [9] or l_1-minimization [10].

However, the discrete-time setting is more appropriate to practical situations. In that framework, various non-linear super-resolution schemes have been developed including greedy [11], root finding [12] [13], matrix pencils [14], Bayesian methods [15] and compressed sensing approaches [16]. Except for root finding techniques, the just mentioned methods are somewhat limited in that they assume the DPs are located on the grid, which is not relevant in many practical situations. Some authors [6] have recently tried to bridge the gap between approaches based on total variation of measures and its discrete-time counterpart, often called LASSO method in the literature, and have shown that the problem of the recovery of DPs is essentially different: while the support of DPs can be exactly recovered in the continuous-time framework such is not the case in the discrete-time setting. Furthermore, these recovery processes are proved to be unstable even at low noise level.

In the present paper, we focus on a particular class of root finding techniques initially proposed in [17] [12] enabling the exact computation of the parameters of DPs using a small number of lowpass filtered samples, the optimal sampling rate being attained for some specific filters like the periodic Dirichlet or Gaussian kernels, or those reproducing polynomials or exponential functions [18]. These types of methods, often referred to as *optimal sampling techniques*, are based, in the absence of noise, on the so-called *annihilating filter method* to estimate DPs parameters. Essentially, this technique computes a filter whose Z-transform zeros enable direct computation of DPs location. Since this approach is not robust to noise, such a filter can be alternatively computed via the *singular value decomposition* (SVD) of a matrix built from the *discrete Fourier transform* (DFT) of the signal [12]. However, to be efficient in noisy situations, the order of the filter computed with such an SVD-based technique has to be much larger than the number of DPs, and this over-modeling gives rise to spurious filter zeros which can be incorrectly identified as signal poles [19]. Furthermore, and as we will see, the benefit of increasing the order of the filter will prove to be highly dependent on the sampling kernel. Once the location of DPs are estimated, the weights are most of the time computed using a least-square fitting approach.

The principle of the DPs parameters estimation proposed in [19] was based on the remark that, in SVD-based techniques, the K (the number of DPs in the stream) largest singular values (SVs) are slightly impacted by noise. This is however only true to a certain extent and, under low SNR condition, a phenomenon called *subspace swap* occurs: some SVs associated with noise become larger than others related to the signal. In [20], by assuming the sampling kernel reproduces some exponential functions, a condition was given to determine the noise level at which the subspace swap occurs, but no remedies were proposed.

Our objective in this paper is to define a new algorithm to denoise the matrix used in the SVD-based approach, which will help improve the stability of SVs, and thus the estimation of DPs location. For ensuring wide applicability, few hypotheses are made on the sampling kernel except that it is lowpass. The benefits of our algorithm will be illustrated on the problem of the estimation of the parameters of close DPs in which case performance typically degrades.

In Section 2, we first formulate the problem of estimating the parameters of a stream of DPs from noisy lowpass filtered samples , give an overview of the techniques used depending on the context, and introduce some useful notation. In Section 3, we recall approaches based on the concept of signals with finite rate of innovation, focusing on SVD-based techniques. Having illustrated the limitations of the latter technique, we define a new algorithm for the retrieval of DPs parameters based on an improved estimation of SVs, in Sections 4 and 5. The method consists of two different steps: 1) a data denoising step (Section 4) and 2) the determination of DPs locations using a matrix of denoised data in SVD-based techniques (Section 5). In Section 6, the selection of the parameters of the proposed method is discussed and a new algorithm to assess the number of DPs present in a stream introduced. Performance analysis of the DPs retrieval technique and a comparison with existing methods is then presented in Section 7. In this regard, we first illustrate the improvement brought by the proposed denoising procedure over the most commonly used technique in the studied context, namely Cadzow denoising [12], and then focus on the comparison of the proposed algorithm for DPs location estimation with some classic approaches like matching pursuit [21], matrix pencil [22] and MUSIC algorithm [23]. Finally, we illustrate the performance of our new method on more complex examples, in particular when the filter used to compute the samples is not symmetric and when the stream contains pulses with negative weights.

2. Problem Statement

Let us consider a signal on $[0, 1]$ composed of a stream of K DPs, located at t_k, with associated weights c_k, for $k \in \{1, \cdots, K\}$, i.e.,

$$f(t) = \sum_{k=1}^{K} c_k \delta_{t_k}, \tag{1}$$

filtered by some low-pass kernel ϕ to obtain

$$m(t) = \sum_{k=1}^{K} c_k \phi(t - t_k) := \Phi(f)(t). \tag{2}$$

We assume the $\{t_k\}_{k=1,\cdots,K}$ are distinct values on $[0, 1]$. Finally, the signal is corrupted by white Gaussian noise η to obtain the following noisy signal

$$m_\eta(t) = m(t) + \eta(t). \tag{3}$$

Defining $\mathbf{c} = \{c_k\}_{k=1,\cdots,K}$ and $\mathbf{l} = \{l_k\}_{k=1,\cdots,K}$, one seeks to retrieve $\mathbf{c}$ and $\mathbf{l}$ from $m_\eta(t)$. In the discrete-time setting, one assumes m_η is uniformly sampled over $[0, 1]$ with rate $\frac{1}{M}$, i.e.

$$m_\eta(\frac{q}{M}) = m(\frac{q}{M}) + \eta(\frac{q}{M}),\ 0 \leq q \leq M - 1, \tag{4}$$

and the problem is then to recover the DPs parameters from $m_\eta(\frac{q}{M})$.

The literature on these problems is vast. In the continuous-time framework, a classical formulation is to try and find some function g minimizing:

$$\min_{g \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|m_\eta - \Phi(g)\|^2 + \lambda |g|(\mathbb{T}), \tag{5}$$

where $|g|(\mathbb{T})$ is the total variation of the measure g and $\mathbb{T}$ the torus $\mathbb{R}\ /\mathbb{Z}$ [24]. Such a problem has recently received a lot of attention in the literature [25] [26] [27], but is very difficult to handle because it is infinite dimensional and its resolution highly depends on Φ. It is however shown in [5] that, when ϕ is C^2, the support of f is governed by a specific solution to the dual problem of (5), and that, when the signal-to-noise ratio is high enough, the exact number of DPs can be found by solving (5) and their locations converge to the true locations when the noise level goes to zero. It is also shown in [5] that the errors on locations and amplitudes decay linearly with the noise level, meaning the exact recovery of the stream of DPs is not achievable in noisy situations. We must also mention that a very close formalism is used in [7] to show how total variation approaches can be used for super-resolution.

In the discrete-time setting, when the DPs are assumed to be on the grid used to collect the noisy samples, and when l_1 minimization replaces the total variation of measures, one ends up with the so-called LASSO method. Extending the work by Fuchs [28], it is shown in [24] that the minimization problem recovers up to twice as many DPs as the input measures, because DPs can get duplicated on immediate nearest neighbors on the grid, which motivated the definition of the notion of *extended support* to study the stability of the recovery process [29]. The study in [24] is particularly interesting in that it bridges the gap between continuous and discrete time formulations in the noiseless case, but work still needs to be done to fully understand the behavior of DPs recovery with these techniques at high noise level.

In the same discrete-time framework, a slightly different formalism is proposed in [30], where the authors assume the DPs are located on the grid and the weights are positive, and then minimize:

$$\min_{g} \|m_\eta - \Phi(g)\|_1 \text{ s.t. } g \geq 0, \tag{6}$$

meaning the weights in the stream of DPs are positive. This problem can be recast as a linear program since m_η is real valued and $\Phi(g)$ rewritten as $\mathbf{Dx}$ with $\mathbf{D}$ a real matrix and $\mathbf{x}$ are weights. In that paper, the spectrum of ϕ is supposed to be contained in $[-M_0, M_0]$, and, since the maximum number of DPs in the stream is M, the *super-resolution factor* (SRF) is defined as: $M/(2M_0+1)$, and corresponds to the ratio between the scale we wish to see the details and at which we see the data. It is shown in [30] that when $\|\mathbf{x}\|_0 < M_0$, the resolution of (6) leads the recovery of f, but only in the noise-free case. When there is noise, the sparsity is not sufficient as our ability to estimate $\mathbf{x}$ from m_η fundamentally depends on how regular the position of the DPs are. In the noisy case, it is also proved that the recovery depends on SRF but slightly on the shape of the filter's spectrum.

The signal defined in Eq. (1) can alternatively be viewed as a signal with finite rate of innovation (FRI), for which specific techniques were developed to retrieve signal parameters [17][12]. These techniques can be implemented in the time domain [31] [20][18], in particular when ϕ reproduces polynomials or some exponential functions. Alternatively, they can be implemented in the Fourier domain [12], namely, one considers the Fourier transform of m_η, assuming the kernel ϕ belongs to $L^1(0,1)$ and is 1-periodic, and then writes:

$$\mathcal{F}(m_\eta)(\xi) = \sum_{k=1}^{K} c_k e^{-i2\pi t_k \xi} \mathcal{F}(\phi)(\xi) + \mathcal{F}(\eta)(\xi), \tag{7}$$

where $\mathcal{F}(f)$ denotes the Fourier transform of f. For instance, such a formulation is often used when ϕ is the periodic Dirichlet kernel. In such a case, considering $n \in \{0, \cdots, M-1\}$, we end up with: $\mathcal{F}(m_\eta)(n) = \sum_{k=1}^{K} c_k e^{-i2\pi t_k n} \mathcal{F}(\phi)(n) + \mathcal{F}(\eta)(n)$. To obtain similar equation as (7) is still possible when ϕ is known only on the grid by assuming t_k belongs to the grid, i.e. $t_k = \frac{l_k}{M}$ for some $l_k \in \{0, \cdots, M-1\}$, and then considering the *discrete Fourier transform* (DFT) of Eq. (4), one obtains the M-periodic sequence:

$$\hat{m}_\eta[n] = \sum_{k=1}^{K} c_k e^{-2i\pi \frac{l_k n}{M}} \hat{\phi}[n] + \hat{\eta}[n], \ n \in \{0, \cdots, M-1\}, \tag{8}$$

in which $\hat{g}[n] := \sum_{q=0}^{M-1} g(\frac{q}{M}) e^{2i\pi \frac{qn}{M}}$. The framework we propose to study in this paper is precisely this one because it requires very few hypotheses on ϕ except that the latter is known only at grid points. The price to pay is to assume the

DPs are on the grid. One should bear in mind, that similar equation to (8) can be obtained without assuming the DPs are located on the grid, but then more must be known on the filter ϕ. So, assuming ϕ is lowpass, such that $\mathcal{F}(\phi)$ is supported on $\{-M_0, \dots, M_0\}$, our goal is to retrieve the DPs parameters from:

$$\hat{m}_\eta[n] = \sum_{k=1}^{K} c_k e^{-2i\pi \frac{l_k n}{M}} \hat{\phi}[n] + \hat{\eta}[n], \ n \in \{-M_0, \cdots, M_0\}. \tag{9}$$

Note finally that, for the sake of simplicity, we also define $m_\eta[q] := m_\eta(\frac{q}{M})$, $m[q] := m(\frac{q}{M})$ and $\phi[q] := \phi(\frac{q}{M})$.

3. Approaches Using the Concept of Signals with Finite Rate of Innovation in the Fourier Domain and Limitations

3.1. Approaches Based on Finite Rate of Innovation for Estimating DPs location

Our goal in this section is to recall how to retrieve the DPs parameters from equations of type (9) exploiting the fact the studied signals are with *finite rate of innovation* (FRI) [17] [12]. Let us first rewrite (9) as $\sum_{k=1}^{K} c_k e^{-2i\pi \frac{l_k n}{M}} = \frac{\hat{m}_\eta[n]}{\hat{\phi}[n]} - \frac{\hat{\eta}[n]}{\hat{\phi}[n]}$ $n \in \{-M_0, \cdots, M_0\}$, and then define $\hat{y}[n] := \sum_{k=1}^{K} c_k e^{-2i\pi \frac{l_k n}{M}}$ and $\hat{y}_\eta[n] := \frac{\hat{m}_\eta[n]}{\hat{\phi}[n]}$. In the absence of noise, $\hat{y}$ can be exactly computed, and then the so-called *annihilating filter method* used to recover the $\{l_k\}_{k=1,\cdots,K}$ from $\hat{y}$. Indeed, $\hat{y}$ is annihilated by a kernel h, such that $(\hat{y} * h)[n] = 0, \forall n$, where

$$\begin{aligned}(\hat{y} * h)[n] &= \sum_{j\in\mathbb{Z}} h[j]\hat{y}[n-j] = \sum_{j\in\mathbb{Z}} h[j] \sum_{k=1}^{K} c_k e^{-i2\pi \frac{l_k(n-j)}{M}} \\ &= \sum_{k=1}^{K} c_k e^{-i\frac{2\pi l_k n}{M}} \underbrace{\sum_{j\in\mathbb{Z}} h[j] e^{i2\pi \frac{l_k j}{M}}}_{H(e^{-i\frac{2\pi l_k}{M}})},\end{aligned} \tag{10}$$

and $H(z)$ is the Z-transform of h. From Eq. (10), it is clear that if $\{e^{-i\frac{2\pi l_k}{M}}\}_{k=1,\cdots,K}$ are the roots of H, then h annihilates $\hat{y}$. Conversely, since, for any m_0, the matrix defined by $(e^{-i2\pi \frac{l_k n}{M}})_{n=m_0,\cdots,m_0+K-1,k=1,\cdots,K}$ is a Vandermonde matrix (n being the row index), it is invertible provided the l_ks are distinct. In this situation, if h annihilates $\hat{y}$, then $H(e^{-i\frac{2\pi l_k}{M}}) = 0$, for $k = 1, \cdots, K$, since c_k is non-zero. Finding the coefficients of h, assuming $h[0] = 1$ and the support of h is $\{0, \cdots, K\}$, amounts to solving the following Yule-Walker system

$$\begin{pmatrix} \hat{y}[0] & \hat{y}[-1] & ... & \hat{y}[-K+1] \\ \hat{y}[1] & \hat{y}[0] & ... & \hat{y}[-K+2] \\ \vdots & \vdots & ... & \vdots \\ \hat{y}[K-1] & \hat{y}[K-2] & ... & \hat{y}[0] \end{pmatrix} \begin{pmatrix} h[1] \\ h[2] \\ \vdots \\ h[K] \end{pmatrix} = - \begin{pmatrix} \hat{y}[1] \\ \hat{y}[2] \\ \vdots \\ \hat{y}[K] \end{pmatrix}.$$

The filter coefficients $\{h[p]\}_{p=1,\cdots,K}$ are thus determined by only $2K+1$ values of $\hat{y}[n]$. Note that since h is entirely defined by the roots of its Z-transform, the above system necessarily has a unique solution.

This approach cannot however recover the location of the DPs when noise is added to the filtered signal since, in that case, $\hat{y}[n]$ is only approximated by $\hat{y}_\eta[n]$ (the quality of estimation degrades as n increases since $\hat{y}_\eta[n] = \hat{y}[n] + \frac{\eta[n]}{\hat{\phi}[n]}$). In such instances, it was suggested in [12] to increase the sampling rate and, assuming the number K of pulses is known, that a *total least-square approximation* (TLSA) should replace the Yule-Walker system described above. First, this approach consists of considering the following rectangular matrix

$$\mathbf{B}_{\eta,T} = \begin{pmatrix} \hat{y}_\eta[-T+K] & \hat{y}_\eta[-T+K-1] & \dots & \hat{y}_\eta[-T] \\ \hat{y}_\eta[-T+K+1] & \hat{y}_\eta[-T+K] & \dots & \hat{y}_\eta[-T+1] \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_\eta[T] & \hat{y}_\eta[T-1] & \dots & \hat{y}_\eta[T-K] \end{pmatrix},$$

for some $K \leq T \leq M_0$. Then, one searches for the unitary vector h minimizing $\|\mathbf{B}_{\eta,T} h\|_2$. This is done by computing the *singular value decomposition* (SVD) of $\mathbf{B}_{\eta,T}$, i.e. $\mathbf{B}_{\eta,T} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^*$ where $\mathbf{U}$ and $\mathbf{W}$ are unitary matrices and $\mathbf{\Sigma}$ the diagonal matrix of SVs (ranked in decreasing order according to their moduli). Indeed, the vector h corresponds to the last column of $\mathbf{W}$ [12], and an estimate $\{\tilde{l}_k\}$ of the locations $\{l_k\}$ are found from the roots of the Z-transform of h the same way as with the annihilating filter technique.

Note that the definition of $\mathbf{B}_{\eta,T}$ requires the knowledge of the number K of DPs present in the stream, and this issue will be discussed later on. The motivation for using this matrix is based on the remark that, in the absence of noise, the rank of $\mathbf{B}_{0,T}$ is K. Thus, DPs location are directly determined from the roots of the Z-transform of any vector in the kernel of $\mathbf{B}_{0,T}$. To prove that the rank of $\mathbf{B}_{0,T}$ is actually K one remarks that

$$\mathbf{B}_{0,T} = \begin{pmatrix} \hat{y}[-T+K] & \hat{y}[-T+K-1] & \dots & \hat{y}[-T] \\ \hat{y}[-T+K+1] & \hat{y}[-T+K] & \dots & \hat{y}[-T+1] \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}[T] & \hat{y}[T-1] & \dots & \hat{y}[T-K] \end{pmatrix}$$

has rank at least K, since it contains a submatrix of rank K (considering the first K columns and rows). Furthermore, it can be decomposed into $\mathbf{B}_{0,T} = \mathbf{VCH}$

with

$$\mathbf{V} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{-\frac{2i\pi l_1}{M}} & e^{-\frac{2i\pi l_2}{M}} & \dots & e^{-\frac{2i\pi l_K}{M}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-\frac{2i\pi l_1(2T-K)}{M}} & e^{-\frac{2i\pi l_2(2T-K)}{M}} & \dots & e^{-\frac{2i\pi l_K(2T-K)}{M}} \end{pmatrix}, \mathbf{C} = \mathrm{diag}(C_1, C_2, \dots, C_K),$$

$$\text{and } \mathbf{H} = \begin{pmatrix} 1 & e^{\frac{2i\pi l_1}{M}} & \dots & e^{\frac{2i\pi l_1 K}{M}} \\ 1 & e^{\frac{2i\pi l_2}{M}} & \dots & e^{\frac{2i\pi l_2 K}{M}} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{\frac{2i\pi l_K}{M}} & \dots & e^{\frac{2i\pi l_K K}{M}} \end{pmatrix},$$

and $C_k = c_k e^{-\frac{2i\pi l_k(-T+K)}{M}}$. Note that since the l_ks are distinct, the rank of each matrix is K, and $\mathbf{B}_{0,T}$ is at most rank K, so that the rank of $\mathbf{B}_{0,T}$ is actually K.

In the presence of noise, $\mathbf{B}_{\eta,T}$ is full rank and, denoting $(\lambda_k)_{k=1,\dots,K+1}$ the SVs of $\mathbf{B}_{\eta,T}$, the quality of TLSA is related to the value of λ_{K+1}, since $\min_{h,\ s.t. \|h\|_2=1} \|\mathbf{B}_{\eta,T} h\|_2 = \lambda_{K+1}$, and also to how different $\mathbf{B}_{\eta,T}$ is from $\mathbf{B}_{0,T}$, which can be measured by comparing the first K SVs of these two matrices. In the remainder of the paper, the technique that uses TLSA with $\mathbf{B}_{\eta,T}$ to localize DPs is denoted by FRI_T.

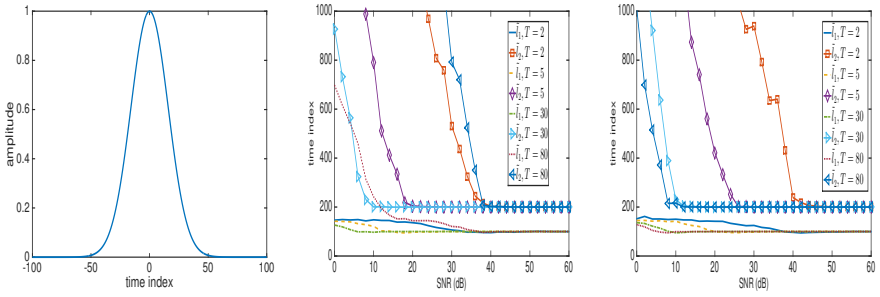
3.2. Limitations

Now, we would like to illustrate the necessity of denoising the matrix $\mathbf{B}_{\eta,T}$ before considering the TLSA approach as is done in FRI_T. To do so, we consider a signal of size $M = 2500$, made of two DPs located at $l_1 = 100$ and $l_2 = 200$, with c_1 and c_2 both equal to 1, the filter ϕ, depicted in Fig. 1 (a), is a Gaussian function, $\phi[q] = e^{-\pi \frac{q^2}{\sigma_\phi^2}}$, for $\sigma_\phi = 40$. The time delay $l_2 - l_1$ is thus equal to 2.5 σ_ϕ, meaning the peaks associated with each DP are well separated in m. For such a σ_ϕ, a reasonable value for the cutoff frequency index M_0 is 92: the criterion used to define M_0 is to consider the first index n such that $\left|\frac{\hat{\phi}[n]}{\hat{\phi}[0]}\right| < 10^{-3}$. Then, some Gaussian white noise is added to the filtered signal to obtain m_η, corresponding to a given input SNR, defined by:

$$SNR(m, m_\eta) = 20 \log_{10} \left(\frac{\|m\|_2}{\|m_\eta - m\|_2} \right). \tag{11}$$

In Fig. 1 (b), we plot the estimated indices $\{\tilde{l}_k\}$ associated with DPs location estimation with FRI_T, with $T \in \{2, 5, 30, 80\}$ with respect to the input SNR. Notice that, whatever the value of T, the detection of the second pulse fails at low SNR. We also remark that considering a larger T does not necessarily result in a better estimation since more noise is then contained in $\mathbf{B}_{\eta,T}$ (see results

obtained for $T = 30$ and $T = 80$ in Fig. 1 (b)). Increasing the sampling rate in that case (i.e. considering more samples $\hat{y}_\eta[n]$) can improve the estimation of DPs location only if the former is not taken to be too large. Furthermore, we notice that even though the filter is symmetric the estimation of the location of the first pulse seems always more accurate. We will show later that this lack of symmetry is mainly due to the presence of additive (Gaussian) noise. Finally, to clearly state that this behavior when T varies is related to the choice of ϕ, we perform the same computation as in Fig. 1 (b), but with the periodic Dirichlet filter, i.e. $\hat{\phi}$ is periodic with period M and such that $\hat{\phi}[n] = 1$ for $n \in \{-M_0, \ldots, M_0\}$ and zero elsewhere on $\{M/2 - 1, \ldots, M/2\}$, meaning the noise is not amplified in $\hat{y}_\eta[n]$ for large n. The results, depicted in Fig. 1 (c), are as reported in the literature: to increase the sampling rate T improves the estimation of DPs location [19]. As just shown, this is not true for a more general lowpass kernel. Furthermore, even when the periodic Dirichlet kernel is used with a large T, the estimation performance significantly degrades around 5 dB, which cannot be considered satisfactory (the results would be even worse if the pulses were moved closer).



(a) (b) (c)

Figure 1: (a): Gaussian filter ($\phi[q]$) for central indices; (b): Estimation of the indices associated with pulses location using FRI_T, with $T = 2$, $T = 5$, $T = 30$ and $T = 80$ (true DPs location are $(l_1, l_2) = (100, 200)$, meaning time delay $l_2 - l_1$ equals 2.5 σ_ϕ, and the results are averaged over 200 realizations); (c): same computation as in (b) except the periodic Dirichlet filter is used instead of the Gaussian filter.

4. Denoising $\hat{y}_\eta$ Using a Two-Threshold Procedure and Piecewise Cubic Hermite Interpolation

In this section, we propose a novel procedure to denoise $\hat{y}_\eta[n]$ for $n = -M_0, \cdots, M_0$. Let us denote $\hat{m}_{\eta,M_0}$ (resp. $\hat{m}_{\eta,M_0}$) the truncation of $\hat{m}_\eta$ (resp. $\hat{m}$) to $|n| \leq M_0$, and then its corresponding representation m_{η,M_0} (resp. m_{M_0}) in the time domain.

The procedure we propose to denoise $\hat{y}_\eta$ is first based on the robust estimation of the standard deviation of the remaining noise in m_{η,M_0}, denoted by σ_{M_0},

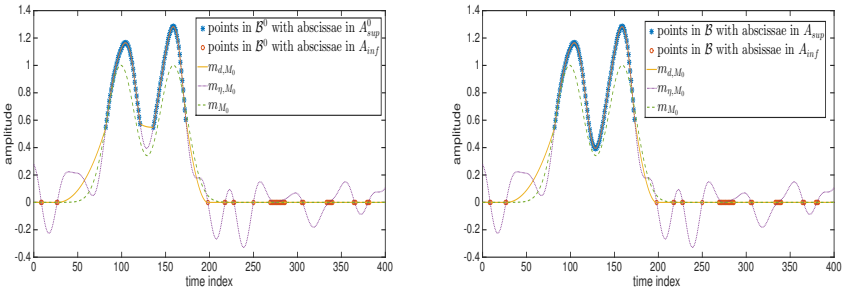
using the *median absolute deviation* criterion [32], i.e.

$$\hat{\sigma}_{M_0} = \text{median}(|m_{\eta,M_0} - \text{median}(m_{\eta,M_0})|)/0.6745.$$

Introducing two sets of coefficients,

$$\begin{aligned} A^0_{\text{sup}} &= \{q,\ |m_{\eta,M_0}[q]| \geq T_1\hat{\sigma}_{M_0}\} \text{ and} \\ A_{\text{inf}} &= \{q,\ |m_{\eta,M_0}[q]| \leq T_2\hat{\sigma}_{M_0}\}, \end{aligned} \tag{12}$$

in which $m_{\eta,M_0}[q] := m_{\eta,M_0}(\frac{q}{M})$, allows us to identify the set of points $\mathcal{B}^0 = (q, m_{\eta,M_0}[q])_{q\in A^0_{\text{sup}}} \bigcup (q,0)_{q\in A_{\text{inf}}}$ which are interpolated using *piecewise cubic monotonic Hermite* interpolation [33]. The signal obtained is denoted by m_{d,M_0}, and we hereafter explain under which conditions on T_1 and T_2 it consists of a denoised version of m_{η,M_0}.



(a) (b)

Figure 2: (a): signal m_{M_0}, m_{η,M_0} and m_{d,M_0} obtained from the points in $\mathcal{B}^0$, input SNR $= -10$ dB (the DPs are located at $(l_1, l_2) = (100, 160)$, the filter ϕ used is displayed in Fig. 1 (a), the time delay between two DPs is 1.75 σ_ϕ, $T_1 = 4$ and $T_2 = 0.2$); (b): same as (a) except that m_{d,M_0} is obtained from $\mathcal{B}$.

Indeed, our denoising strategy is based on the fact that m_{η,M_0}, close to its local maxima is usually much larger than the noise level: most of these points in the set A^0_{sup} are captured by considering a large T_1 (see Fig. 2 (a), $T_1 = 4$). Then, to remove the noise-only part of the signal, we determine abscissae almost surely associated with noise by considering the set A_{inf} with a small T_2 (see Fig. 2 (a), $T_2 = 0.2$). Finally, to set the ordinates of the points with abscissae in A_{inf} to 0 and to use the piecewise cubic monotonic Hermite interpolation ensure the interpolated signal is null between two points with abscissae in A_{inf}. This type of interpolation also guarantees that no oscillations are created between two points with abscissae in A^0_{sup} and a smooth transition between points of $\mathcal{B}^0$ with respective abscissa in A_{inf} and A^0_{sup} (see Fig. 2 (a)).

We then remark that A^0_{sup} is made of discrete intervals, between which there should be some points in A_{inf} when the DPs are sufficiently far appart. However, between two close DPs, the magnitude of m_{η,M_0} can be below $T_1\hat{\sigma}_{M_0}$, but still above $T_2\hat{\sigma}_{M_0}$. In such cases, the two DPs are associated with two discrete

intervals in A^0_{sup}, with no points in A_{inf} between them, and to interpolate the points in $\mathcal{B}^0$ in such instances does not lead to a good estimate of m_{M_0} (see Figure 2 (a) between the two DPs). To cope with this situation, we slightly change the definition of A^0_{sup} into:

$$\begin{aligned} A_{\text{sup}} \quad = \quad & \bigcup \{[k_1, k_2],\ s.t. \\ & |m_{\eta,M_0}[k_1]| \geq T_1 \hat{\sigma}_{M_0},\ |m_{\eta,M_0}[k_1 - 1]| < T_1 \hat{\sigma}_{M_0} \\ & |m_{\eta,M_0}[k_2]| \geq T_1 \hat{\sigma}_{M_0}\ |m_{\eta,M_0}[k_2 + 1]| < T_1 \hat{\sigma}_{M_0} \\ & \text{and } \forall k_1 < k < k_2,\ |m_{\eta,M_0}[k]| \geq T_2 \hat{\sigma}_{M_0}\} \,. \end{aligned} \tag{13}$$

Finally, when one uses a relatively small T_1, and when the noise level is high, some discrete intervals in A_{sup} may correspond to noise, and should therefore be removed. To take this issue into account, we also put a lower bound on the length of the intervals kept in A_{sup} as follows. Let $A_{\text{sup},i}$ be the ith interval in A_{sup}, and consider the set $E_i = \{|m_{\eta,M_0}[k]| \geq T_1 \hat{\sigma}_{M_0},\ k \in A_{\text{sup},i}\}$. We then keep in A_{sup} only the intervals $A_{\text{sup},i}$ such that $\#E_i > L_{\min}$, where $\#X$ stands for the cardinal of X, the choice for $L_{\min}$ being discussed later in the paper. One then defines $\mathcal{B} := (q, m_{\eta,M_0}[q])_{q \in A_{\text{sup}}} \bigcup (q, 0)_{q \in A_{\text{inf}}}$, and interpolating the points in $\mathcal{B}$ using piecewise cubic Hermite interpolation, we obtain a signal which is denoted also by m_{d,M_0} for the sake of simplicity.

To illustrate the impact of choosing A_{sup} rather that A^0_{sup} in the interpolation procedure, we consider the same signal made of two DPs studied in Fig. 2 (a) except that m_{d,M_0} is this time constructed from $\mathcal{B}$. The result displayed in Fig. 2 (b) shows that the estimation of m_{M_0} is improved between the two DPs. $L_{\min}$ is set to 10 but the result is not sensitive to that parameter for that particular illustration. In Section 6.1, we will assess the sensitivity of this denoising procedure to T_1, T_2 and T. From now on, $\hat{y}_d$ denotes the denoised version of $\hat{y}_\eta$ corresponding to $\hat{y}_d[n] = \frac{\hat{m}_{d,M_0}[n]}{\hat{\phi}[n]}$, with m_{d,M_0} computed from $\mathcal{B}$ and the denoising technique is denoted by TTPC (for Two Threshold Piecewise Cubic interpolation).

5. DPs Parameters Estimation

Having denoised $\hat{y}_\eta$, we define a new matrix to replace $\mathbf{B}_{\eta,T}$ in the TLSA framework described in Section 3. The entries of the matrix are selected from the denoised sequence $\hat{y}_d$ defined Section 4. Once DPs location are estimated, weights estimation is performed using *least-square fitting* LSF ; we recall hereafter the two procedures for DPs parameters estimation.

First, we define a new matrix

$$\mathbf{B}_{d,T} = \begin{pmatrix} \hat{y}_d[-T+K] & \hat{y}_d[-T+K-1] & \dots & \hat{y}_d[-T] \\ \hat{y}_d[-T+K+1] & \hat{y}_d[-T+K] & \dots & \hat{y}_d[-T+1] \\ \hat{y}_d[-T+K+2] & \hat{y}_d[-T+K+1] & \dots & \hat{y}_d[-T+2] \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_d[T] & \hat{y}_d[T-1] & \dots & \hat{y}_d[T-K] \end{pmatrix},$$

where $K \leq T \leq M_0$, which corresponds to matrix $\mathbf{B}_{\eta,T}$ in which $\hat{y}_\eta$ is replaced by $\hat{y}_d$. Then, we replace $\mathbf{B}_{\eta,T}$ by $\mathbf{B}_{d,T}$ in the TLSA framework to compute an estimate of DPs location. This new method to estimate DPs location is denoted by $FRI_{d,T}$ in the sequel.

Once the locations $\mathbf{l} = \{l_k\}_{k=1,\cdots,K}$ are estimated by $\tilde{\mathbf{l}} = \{\tilde{l}_k\}_{k=1,\cdots,K}$, the simplest way to estimate the weights is by computing a *least square fitting* (LSF) solution $\tilde{\mathbf{c}}$ solving the following problem

$$\tilde{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^K}{\operatorname{argmin}} \|m_\eta - \sum_{k=1}^{K} c_k \phi[. - \tilde{l}_k]\|_2^2, \tag{14}$$

corresponding to the approximation error:

$$E_{\tilde{\mathbf{l}}} = \|m_\eta - \sum_{k=1}^{K} \tilde{c}_k \phi[. - \tilde{l}_k]\|_2^2. \tag{15}$$

This technique to compute the locations and weights is referred to as $FRI_{d,T} - MLE$ in the sequel.

6. Parameter Tuning

6.1. Sensitivity of TTPC to T_1, T_2 and T

In this subsection, we first explain how to tune the parameters T_1, T_2 and T by studying the sensitivity of the denoising procedure, called TTPC and detailed in Section 4, to these parameters in different configurations, namely when the input SNR and the number of DPs in the stream vary. The parameter $L_{\min}$ is set to 10, and its influence will be discussed later.

More precisely, for different signals we are going to study the sensitivity to T_1 and T_2 of the average of $\frac{\|\hat{y}-\hat{y}_d\|_{2,[-T,T]}}{\sqrt{2T+1}}$, in which the subscript $2, [-T, T]$ means we consider the l_2 norm and restrict ourselves to indices n in $[-T, T]$, the normalization by $\sqrt{2T+1}$ being added to obtain the average error on each coefficient $\hat{y}[n]$ (the l_2 error increases necessarily with the number of taps taken into account).

We carry out this simulation on three types of signal, a two DP signal with DPs located at $(l_1, l_2) = (100, 200)$, another with DPs located at $(l_1, l_2) =$

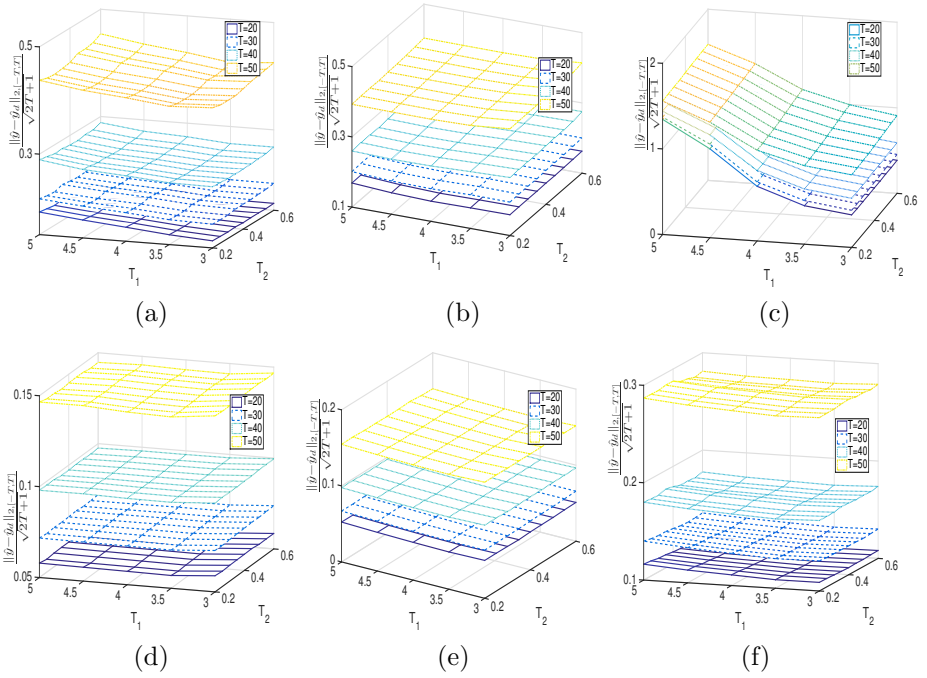


Figure 3: (a) and (d) : Computation of $\frac{\|\hat{y}-\hat{y}_d\|_{2,[-T,T]}}{\sqrt{2T+1}}$ when the signal is made of two DPs with $(l_1, l_2) = (100, 200)$ (i.e., the time delay between DPs is $2.5\sigma_\phi$, (a): SNR = -10 dB (d): SNR = 0 dB; (b) and (e): same as (a) and (d) when $(l_1, l_2) = (100, 120)$, time delay between DPs is $0.5\sigma_\phi$; (c), (d): same as (a) and (f) except that the DPs are located at $(100, 150, 300, 400)$.

$(100, 120)$, and a last one with DPs located at $(l_1, l_2, l_3, l_4) = (100, 150, 300, 400)$ (in each case the weights associated with each DP are equal). The filter we use is still the Gaussian filter of Fig. 1 (a). Numerical results reported in Fig. 3 suggest that, whatever the SNR and in the two DPs case, $\frac{\|\hat{y}-\hat{y}_d\|_{2,[-T,T]}}{\sqrt{2T+1}}$ is very stable whatever T_1 and T_2 in the tested ranges and the time delay between the two pulses (the results are reported in Fig. 3 (a) and (b) (SNR = -10 dB) , and (d) and (e) (SNR = 0 dB)). When the stream contains more DPs, by taking a large T_1 and only when the noise level is high, there is a chance that some pulses are not detected, resulting in a larger error (see Fig. 3 (c)).

Furthermore, by computing the error for various T we notice that while with $T = 20$ or 30 the denoising performance are relatively similar, these degrade when considering a larger T (the discrepancy between the case $T = 20$ and $T = 30$ is always much smaller than that between $T = 30$ and $T = 40$). This suggests to use the smallest value for T compatible with the number of DPs to be detected, but we are going to see a bit later that the location of the DPs also matters for the selection of an appropriate T. So, for the algorithm to work

	$T=20$	$T=30$	$T=40$	$T=50$
$(l_1,l_2)=(100,200), SNR=-10dB$	0.67	0.80	1.09	1.77
$(l_1,l_2)=(100,200), SNR=0dB$	0.21	0.25	0.34	0.56
$(l_1,l_2)=(100,120), SNR=-10dB$	0.85	1.04	1.40	2.23
$(l_1,l_2)=(100,120), SNR=0dB$	0.27	0.32	0.44	0.69
$(l_1,l_2,l_3,l_4)=(100,150,300,400), SNR=-10dB$	0.96	1.15	1.59	2.48
$(l_1,l_2,l_3,l_4)=(100,150,300,400), SNR=0dB$	0.31	0.37	0.48	0.81

Table 1: computation of $\frac{\|\hat{y}-\hat{y}_\eta\|_{2,[-T,T]}}{\sqrt{2T+1}}$, for the studied signal of Figure 3, for the same SNR and values of T as in that figure

well in all situations, a value of T_1 around 3.5 appears to be a good trade-off. Note finally that, in all these simulations, the results appear to be only slightly sensitive to T_2.

To check that the algorithm actually performs some kind of denoising, we also compute $\frac{\|\hat{y}-\hat{y}_\eta\|_{2,[-T,T]}}{\sqrt{2T+1}}$, for the same values of T and SNRs as those of Fig. 3. The results reported in Tab. 1 and compared with those displayed in Fig. 3, confirm that TTPC actually denoises $\hat{y}_\eta$.

6.2. On the Number of DPs

Having studied the influence of parameters T_1, T_2 and T on TTPC, we propose a novel technique to determine the number of DPs in a stream. First, based on the previous study, reasonable values for T_1 and T_2 are 3.5 and 0.2 respectively. With these values we now explain why the denoising sequence $\hat{y}_d$ can be profitably used to determine the number of DPs present in a stream. Consider that there are at most $K_m \geq K$ DPs in the signal of interest, and then define:

$$\mathbf{B}_{\eta,T}^{K_m} = \begin{pmatrix} \hat{y}_\eta[-T+K_m] & \hat{y}_\eta[-T+K_m-1] & \dots & \hat{y}_\eta[-T] \\ \hat{y}_\eta[-T+K_m+1] & \hat{y}_\eta[-T+K_m] & \dots & \hat{y}_\eta[-T+1] \\ \hat{y}_\eta[-T+K_m+2] & \hat{y}_\eta[-T+K_m+1] & \dots & \hat{y}_\eta[-T+2] \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_\eta[T] & \hat{y}_\eta[T-1] & \dots & \hat{y}_\eta[T-K_m] \end{pmatrix},$$

$\mathbf{B}_{d,T}^{K_m}$ being defined similarly replacing $\hat{y}_\eta$ by $\hat{y}_d$. As remarked in [17], the rank of the matrix $\mathbf{B}_{\eta,T}^{K_m}$ is K in the noise-free case, and K_m+1 when the signal is noisy. One expects the first K SVs to be related to the signal while the last K_m-K+1 to the noise. So, what matters is to analyze how the SNR affects the first K SVs and also how low the $(K+1)$th SV remains as the SNR decreases. In this regard, we notice numerically that the first K SVs are stable, regardless of the noise level, for matrix $\mathbf{B}_{d,T}^{K_m}$, but not for matrix $\mathbf{B}_{\eta,T}^{K_m}$, in which the Kth SV is considerably increased. Furthermore, the $(K+1)$th SV remains small in $\mathbf{B}_{d,T}^{K_m}$ but not in $\mathbf{B}_{\eta,T}^{K_m}$. The denoising procedure we propose thus enables to keep a clear separation of the signal and noise parts in matrix $\mathbf{B}_{d,T}^{K_m}$.

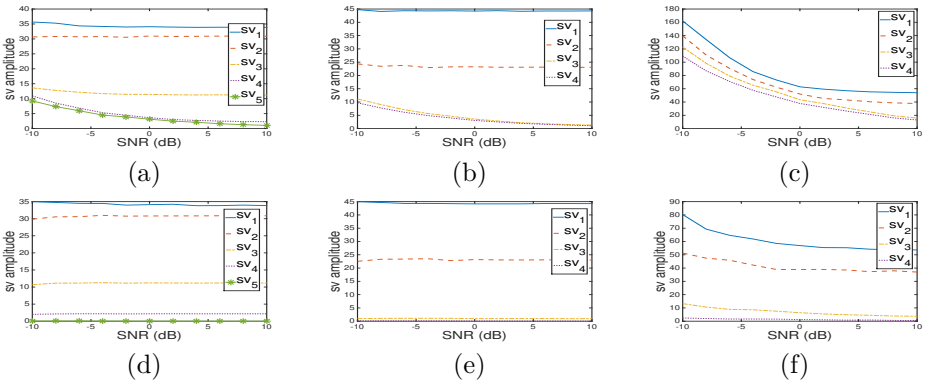


Figure 4: (a) and (d): Amplitude of the first five SVs of $B^{10}_{\eta,30}$ and $B^{10}_{d,30}$ respectively, with respect to input SNR, signal consists of four DPs located at $(l_1, l_2, l_3, l_4) = (100, 150, 300, 400)$ with the same amplitude (the smallest time delay between pulses is 1.25 σ_ϕ, results averaged over 50 realizations); (b) and (e): Amplitude of the first four SVs of $B^{10}_{\eta,30}$ and $B^{10}_{d,30}$ respectively, with respect to input SNR, signal consists of four DPs located at $(l_1, l_2, l_3, l_4) = (100, 120, 300)$ with the same amplitude (the smallest time delay between pulses is 0.5 σ_ϕ, results averaged over 50 realizations); (c) and (f): same as (b) and (e) but SVs correspond to $B^{10}_{\eta,70}$ and $B^{10}_{d,70}$.

To illustrate this, we consider two different situations: the first one is a signal containing four DPs with the same amplitudes and located at $(l_1, l_2, l_3, l_4) = (100, 150, 300, 400)$ (the time delay between the two closest DPs is 1.25 σ_ϕ, the filter ϕ being still the Gaussian filter used in previous simulations). The second signal contains three DPs located at $(l_1, l_2, l_3) = (100, 120, 300)$, with the same amplitudes as for the first signal (the time delay between the two closest DPs is 0.5 σ_ϕ). We assume $K_m = 10$ and then note that $K_m \leq T \leq M_0$. T is also constrained by the minimal time delay between two DPs: in the first case, the SRF, defined in Section 2 and equal to $\frac{M}{2T+1}$ has to be above 50 meaning $T \geq 24.5$, and in the second case T has to be larger than 62. So a relevant value for T in the first case is 30, and the SVs of $\mathbf{B}^{10}_{\eta,30}$ and $\mathbf{B}^{10}_{d,30}$ are reported in Fig. 4 (a) and (d) and behave as explained in the previous paragraph. Now, if we switch to the second case, to consider a larger T than the one given by SRF, leads to some noise remaining in $\hat{y}_d$. To illustrate this, we plot in Fig. 4 (b) and (e) (resp. Fig. 4 (c) and (f)) the first four SVs of $\mathbf{B}^{10}_{\eta,30}$ and $\mathbf{B}^{10}_{d,30}$ (resp. $\mathbf{B}^{10}_{\eta,70}$ and $\mathbf{B}^{10}_{d,70}$): with $T = 30$ the denoising is more efficient than with $T = 70$ but the amplitude of the third SV is much lower, making the choice for T more complicated than with the first signal. Finally note that $L_{\min}$ is set to 10, and a discussion on this parameter follows in the next subsection.

We now exploit the stability of SVs associated with matrix $\mathbf{B}^{K_m}_{d,T}$ to build a procedure that automatically computes the number of DPs. Remarking that SVD can be interpreted in terms of energy contained in subspaces spanned by the singular vectors, for each SNR, we determine the number of DPs by

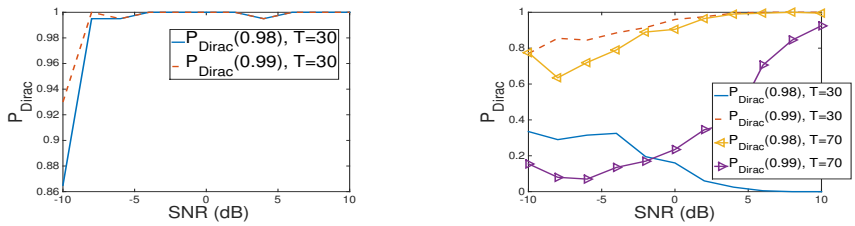


Figure 5: (a): Estimation of the proportion of correct number of DPs for the 4 DP signal via formula (18); (b): same as (a) but for the 3 DP signal. The results are averaged over 200 Monte Carlo realizations.

considering first the amount of the energy contained in the first k SVs through:

$$S(k) = \frac{\sum_{i=1}^{k} |sv_i|}{\sum_{i=1}^{K_m} |sv_i|}, \tag{16}$$

where sv_i is the ith SV, and then estimate the number of DPs through:

$$\hat{K} = \inf_k \{S(k) > T_3\}, \tag{17}$$

where T_3 is some threshold. Finally, we measure the quality of the estimate $\hat{K}$ by computing the proportion of correct DPs number determination for MC realizations of the noise, i.e. by defining:

$$P_{Dirac}(T_3) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (\hat{K}^{[i]} = K), \tag{18}$$

where $\hat{K}^{[i]}$ is the estimation of the number of DPs for the ith realization of the noise. In Fig. 5, we display (18) for the two signals studied in Fig. 4 and when the input SNR varies. This shows that provided T_3 is appropriately chosen (close to 1), the proposed procedure enables the determination of the number of DPs when these are not too close (the four DPs example). When some DPs are very close (the three DPs example), if T is chosen large then some SVs related to noise may not be negligeable and to choose a threshold T_3 close to 1 may result in the computation of a wrong number of DPs. So for our technique to work well in the case of close DPs one had rather take T much lower than the one given by SRF, and then take T_3 sufficiently large.

6.3. Sensitivity to $L_{\min}$

Once the thresholds T_1, T_2 and T are correctly set and the number of DPs determined as explained in the previous two subsections, we investigate the

sensitivity of $FRI_{d,T}$ to parameter $L_{\min}$. For that purpose, we consider a two DP signal where these are moved farther apart corresponding to a time delay between the two pulses either equal to 0.5 σ_ϕ or 1 σ_ϕ.

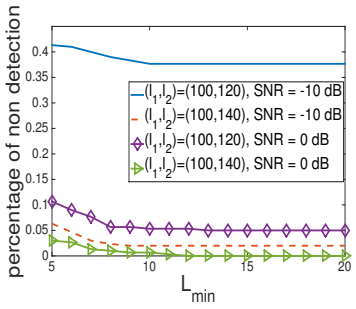


Figure 6: Percentage of non detection with respect to $L_{\min}$ for three different signals containing 2 DPs. The results are averaged over 200 Monte Carlo realizations.

We compute the number of non-detections with respect to $L_{\min}$ when T_1 and T_2 are set to 3.5 and 0.2, and T set to 30. To do so we say that the algorithm achieves DPs detection when $|\tilde{l}_k - l_k| < 10$ for all k (this also means the time delay between pulses has to be larger than 20). Using this definition, we compute the proportion of non detection when $L_{\min}$ varies for the two studied signals. The results displayed in Fig. 6 show that a large $L_{\min}$ should clearly be favored to limit non detection when the DPs are close and the noise level is high but its value matters much less in any other studied situations. Furthermore, when the time delay between the pulses equals 0.5 σ_ϕ and when the SNR equals -10 dB, the non detection rises up to 40 % and, in any other cases, the percentage of non detection remains low. It also transpires that to choose $L_{\min}$ equal to 10 is appropriate in all cases (this a posteriori justifies the choice made up to now).

7. Numerical Results

7.1. Evaluation of TTPC

Our goal in this section is to compare the novel strategy for the denoising of $\hat{y}_\eta$, called TTPC and detailed in Section 4, to the most commonly used technique in the TLSA framework and known as Cadzow denoising [12] [34], the principle of which we recall hereafter.

A strategy to denoise $\hat{y}_\eta$, and known as Cadzow denoising [12] [34], is to consider the following square matrix:

$$\tilde{\mathbf{B}}_{\eta,T} = \begin{pmatrix} \hat{y}_\eta[0] & \hat{y}_\eta[-1] & \dots & \hat{y}_\eta[-T] \\ \hat{y}_\eta[1] & \hat{y}_\eta[0] & \dots & \hat{y}_\eta[-T+1] \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_\eta[T] & \hat{y}_\eta[T-1] & \dots & \hat{y}_\eta[0] \end{pmatrix},$$

for some $K \leq T \leq M_0$. First one remarks that in the absence of noise the rank of $\tilde{\mathbf{B}}_{\eta,T}$ is K and $T+1$ in the noisy case. Then one computes the SVD of $\tilde{\mathbf{B}}_{\eta,T} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{W}}^*$, and set to 0 the smallest $T-K+1$ SVs in $\tilde{\mathbf{\Sigma}}$ to obtain $\tilde{\mathbf{\Sigma}}'$, which in turn enables the definition of a new matrix $\tilde{\mathbf{B}}'_{\eta,T} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}'\tilde{\mathbf{W}}^*$, of rank K but no longer Toeplitz. To retrieve the Toeplitz structure of the initial matrix $\tilde{\mathbf{B}}_{\eta,T}$, one replaces the coefficients on each diagonal of $\tilde{\mathbf{B}}_{\eta,T}$ by the average of the coefficients on this diagonal, to obtain matrix $\tilde{\mathbf{B}}^{[1]}_{\eta,T}$. One then iterates this procedure until the $(K+1)$th SV is smaller that the Kth by some prerequisite factor. Note that a rectangular matrix could be used instead of $\tilde{\mathbf{B}}_{\eta,T}$, but the denoising results are reported to be better on a square matrix [12].

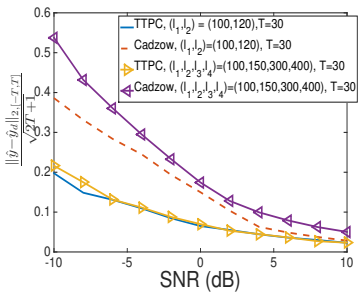


Figure 7: Comparison between Cadzow denoising and TTPC for a two DPs signal located at $(100, 120)$ and a four DPs located at $(100, 150, 300, 400)$. For Cadzow denoising the number of iteration is set to 50. The results are average over 200 Monte Carlo realizations.

To illustrate the improvement brought by TTPC over Cadzow denoising, we consider two different signals: the first one is made of two DPs located at $(100, 120)$ and the second one of DPs located at $(100, 150, 300, 400)$ (the weights being all equal). In Fig. 7, we plot $\frac{\|\hat{y}-\hat{y}_d\|_{2,[-T,T]}}{\sqrt{2T+1}}$ when $\hat{y}_d$ either corresponds to Cadzow or TTPC denoising. In each case, T is set to 30, because it is compatible with SRF for the second signal and a good tradeoff for the first signal. In each studied case, the benefit of using TTPC rather than Cadzow denoising is undeniable.

7.2. Comparison of $FRI_{d,T}$ with Other DPs Location Estimators

In this section, we introduce commonly used techniques to estimate the locations of DPs in a stream, for the sake of comparison with the approach called $FRI_{d,T}$. Classical alternative approaches involves *matrix pencil* algorithm, which we implement following [22], MUSIC algorithm [23], or matching pursuit techniques [21].

The basic idea of matching pursuit (MP) [21] is to approximate the signal m_η by a linear combination of functions selected in the set

$$\mathcal{S} = \{\phi[. - n], n \in \{0, \cdots, M-1\}\}, \tag{19}$$

where ϕ is the filter introduced in the previous section. If $\mathbf{D}$ denotes the $M \times M$ matrix whose columns are the translated versions of the function ϕ, for $n \in \{0, \cdots, M-1\}$, MP aims at solving the following problem

$$\min_x \|\mathbf{m}_\eta - \mathbf{D}\mathbf{x}\|_2, \text{ subject to } \|\mathbf{x}\|_0 \leq K, \tag{20}$$

where $\mathbf{m}_\eta = \{m_\eta[q]\}_q$, and $\|\mathbf{x}\|_0$ is the number of non zero coefficients in $\mathbf{x}$. The algorithm is based on an iterative procedure to construct an approximation of m and it consists of choosing, at each iteration, the function in the so-called dictionary $\mathbf{D}$ whose inner product with the remainder (i.e. the difference between $\mathbf{m}_\eta$ and its approximation at the previous iteration) is maximal in terms of its modulus [21]. In that context, the indices of the non-zero components in $\mathbf{x}$ correspond to DPs location while the values of the corresponding components correspond to their weights. To accelerate the convergence to a solution, a possibility is to move orthogonally to the remainder at each step, which corresponds to *orthogonal matching pursuit* (OMP) [35]. When the weights associated with DPs are positive, it is possible to account for such constraints in the matching pursuit framework using *non negative matching pursuit* (NNMP) [36].

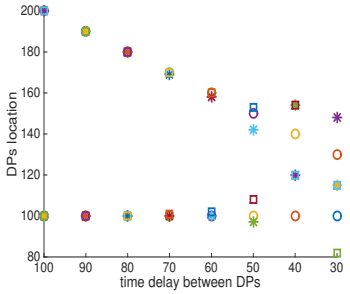


Figure 8: the circles represent the true DPs location, the stars (resp. squares) correspond to the location given by MP (resp. NNMP)

We are first going to show that MP, OMP, and NNMP techniques are irrelevant to the problem of the determination of close DPs. Indeed, consider a signal f made of two DPs with varying time delay between them, and then solve the noiseless problem:

$$\min_x \|\mathbf{m} - \mathbf{D}\mathbf{x}\|_2, \text{ subject to } \|\mathbf{x}\|_0 \leq 2. \tag{21}$$

The location of the DPs computed by means of (21) and displayed in Fig. 8, show that even in the absence of noise DPs cannot be located that way when the pulses are relatively close (the results being identical with OMP, they are not displayed). When one uses NNMP, the obtained locations are neither satisfactory. So, we drop the comparison with these types of methods in the remainder of this paper.

We are now going to see that the other tested methods based on spectral estimation like $FRI_{d,T}$, matrix pencil or MUSIC methods work much better than matching pursuit. As in Section 6.3, we consider a two DP signal in

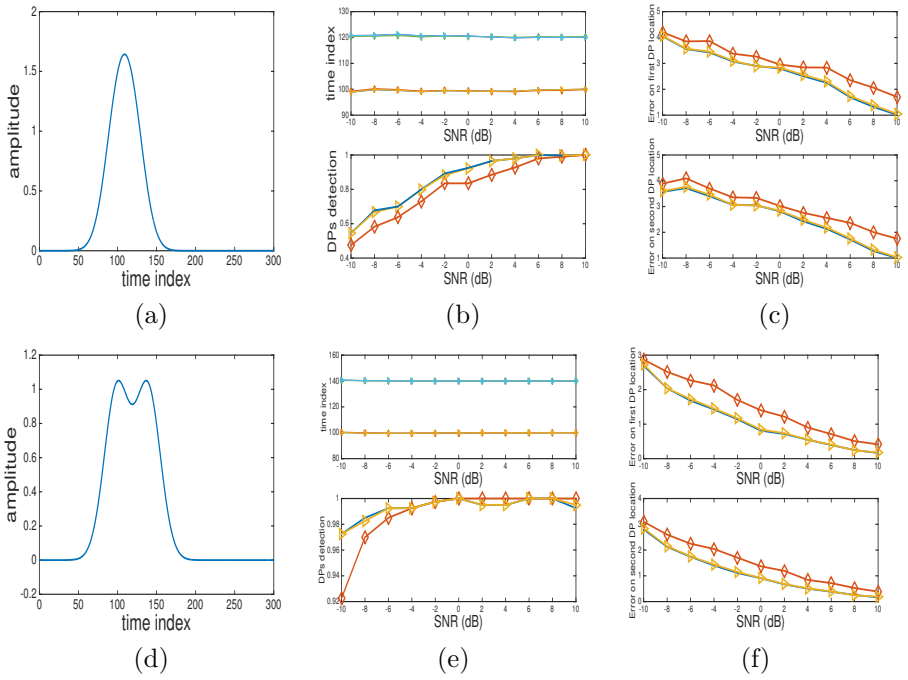


Figure 9: (a): filtered signal when DPs are located in $(l_1, l_2) = (100, 120)$, (b): *top:* estimated positions of DPs using $FRI_{d,T}$ (plain), matrix pencil (diamond), MUSIC (triangle), *bottom:* proportion of good detection using $FRI_{d,T}$ (plain), matrix pencil (diamond), MUSIC (triangle), both for the signal displayed in (a); (c): (22) computed for the two DPs of signal displayed in (a) (methods are associated with the same markers as in (b)). (e) and (f): same as (b) and (c) but for the signal displayed in (d). The results are averaged over 200 Monte Carlo realizations.

which the time delay between the two pulses is either 0.5 σ_ϕ or 1 σ_ϕ: these two situations are displayed in Figure (a), (d). We then estimate DPs location with either $FRI_{d,T}$, matrix pencil or MUSIC algorithms. The value for T is set to 30 for the first method and so, for the sake of a fair comparison, the entries to the other two algorithms are also $\hat{y}_d[n]$, $-T \leq n \leq T$ (if we considered $\hat{y}_\eta$ the results would be much worse). Before estimating the quality of the estimation of DPs location, we say that a given method *achieves DPs detection* if $|\tilde{l}_k - l_k| < 10$ for all k (here $\{\tilde{l}_k\}$ are estimated by any of the three methods). We then assess the quality of the estimates $\{\tilde{l}_k\}$, when the corresponding method achieves DPs detection. For the filtered signal of Fig. 7.2 (a), the average detected locations in case the methods achieve DPs detection are displayed in Fig. 7.2 (b) (top) and the proportion of time the methods achieve DPs detection in Fig. 7.2 (b) (bottom). We then approximate for each DP the expectation of $|\tilde{l}_k - l_k|$ using

Monte Carlo realizations, by

$$Err(\tilde{l}_k, l_k) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} |\tilde{l}_k^{[i]} - l_k|, \tag{22}$$

where $\tilde{l}_k^{[i]}$ is the kth DP location estimated with the ith realization of the noise and N_{MC} is the number of Monte Carlo realizations for which DPs detection has been achieved. For the signal of Fig. 7.2 (a), (22) computed for each DP is displayed in Fig. 7.2 (c). The same analysis is performed on the second row of Figure 7.2, but for the signal of Fig. 7.2 (d).

We first notice that when the three methods achieves DPs detection, the average estimation of DPs location is similar with the three tested methods (see the top of Figures 7.2 (b) and (e)), and (22) computed for each DP is the same for $FRI_{d,T}$ and MUSIC but slightly higher for matrix pencil technique. Similarly the proportion of time $FRI_{d,T}$ and MUSIC achieve DPs detection are very similar, but smaller with the matrix pencil technique. So, this simple study suggests that regarding the determination of DPs location, the most important step is the denoising step of $\hat{y}_\eta$, namely TTPC, and then $FRI_{d,T}$ or MUSIC behave the same way in terms of the estimation of DPs location.

7.3. Evaluation of the Weights Estimation Procedure in $FRI_{d,T} - MLE$

To investigate the quality of the weights estimator associated associated with method $FRI_{d,T} - MLE$ defined in Section 5, we introduce the *oracle maximum likelihood estimator* (oracle MLE) of $\mathbf{c}$, computed assuming DPs location are known, i.e.,

$$\mathbf{c}_{oMLE} = \underset{\mathbf{c}\in\mathbb{R}^K}{\operatorname{argmin}} \|m_\eta - \sum_{k=1}^{K} c_k \phi[. - l_k]\|_2^2. \tag{23}$$

We denote by E_{oMLE} the minimal energy obtained when solving Eq. (23). We also define the *restricted maximum likelihood estimator* (restricted MLE) of $\mathbf{c}$ as

$$\mathbf{c}_{reMLE} = \underset{\mathbf{c}\in\mathbb{R}^K, \mathbf{q}\in\mathcal{E}_{\tilde{\mathbf{l}},\Delta}}{\operatorname{argmin}} \|m_\eta - \sum_{k=1}^{K} c_k \phi[. - q_k]\|_2^2, \tag{24}$$

where $\mathcal{E}_{\tilde{\mathbf{l}},\Delta} = \left\{\mathbf{q} = (q_k)_{k=1,\cdots,K},\ q_k \in \{\tilde{l}_k - \Delta, \cdots, \tilde{l}_k + \Delta\}\right\}$. The corresponding energy is denoted by E_{reMLE}. To measure the quality of the weights estimation associated with the above mentioned technique, we consider the following normalized error (written for $\tilde{c}_k$):

$$Err_n(\tilde{c}_k, c_k) = Err(\tilde{c}_k, c_k)/c_k. \tag{25}$$

To illustrate the limitations of MLE estimation for the weights, we again consider two signals, one with $(l_1, l_2) = (100, 120)$ and the other $(l_1, l_2) = (100, 140)$

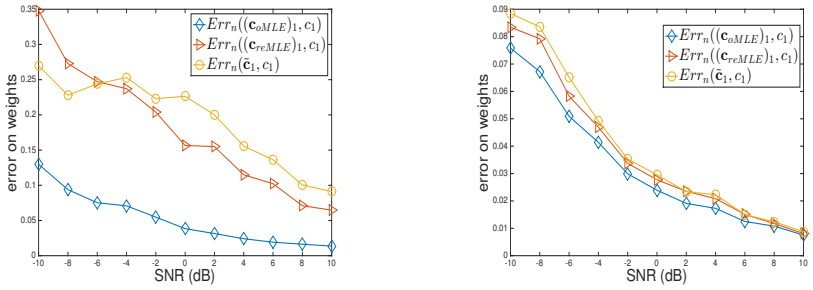


Figure 10: (a): normalized estimation error on c_1 when $\mathbf{c}$ is estimated either by $\mathbf{c}_{oMLE}$, $\mathbf{c}_{reMLE}$, or $\tilde{\mathbf{c}}$ with $(l_1, l_2) = (100, 120)$ ($\Delta = 10$); (b): same as (a) except that the signal correspond to DPs located in $(l_1, l_2) = (100, 140)$. Results are averaged over 200 Monte Carlo realizations.

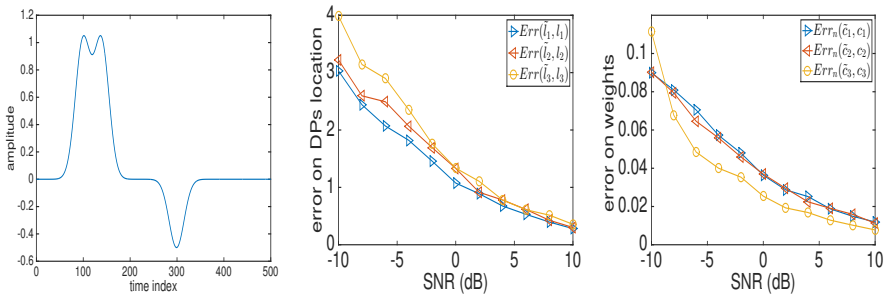
and, in both signals, the DPs are associated with identical weights. Then, we compute $\mathbf{c}_{oMLE}$, $\tilde{\mathbf{c}}$ and $\mathbf{c}_{reMLE}$, the latter two only when $FRI_{d,T}$ achieves DPs detection. We take $\Delta = 10$ in the definition of $\mathcal{E}_{\tilde{\mathbf{l}},\Delta}$ so that we are sure that $E_{reMLE} \leq E_{oMLE}$, because $\mathbf{l}$ is contained in $\mathcal{E}_{\tilde{\mathbf{l}},\Delta}$ (by definition of DPs detection, defined in subsection 7.2). Looking at Fig. 10 (a), we notice that the minimum of E_{reMLE} is not attained for $\mathbf{q} = \mathbf{l}$ which induces a larger error on the weights estimation than with $oMLE$ (we only focus on the behavior of coefficient c_1, that of c_2 being very similar). We then notice that when the time delays between DPs increases, the estimation of the weights given by any of the three methods is similar as reported in Fig. 10 (b)(note that in that case, DPs detection is almost always achieved).

In a MLE framework, the DPs location estimation thus needs to be increasingly more accurate as the time delay between DPs shortens to allow for an accurate computation of the weights. This however is not guaranteed by restricted MLE. To improve the weights estimation we thus believe that it is better to use a technique that ensures a very good estimation of DPs location before computing the weights.

7.4. Illustration of $FRI_{d,T} - MLE$ on a Three Dirac Pulse Signal

Our goal here is to illustrate the behavior of $FRI_{d,T} - MLE$ on a three Dirac pulse signal. The measure of the quality of DP parameters estimation we use is the function Err introduced in (22), and we compute $Err(\tilde{l}_k, l_k)$ and $Err_n(\tilde{c}_k, c_k)$, for $k = 1, \cdots, K$ to assess the quality of estimation of DPs locations and weights. As already noticed, MUSIC applied to the sequence $\hat{y}_d$ behaved similarly to $FRI_{d,T}$ for the estimation of DPs location for a two DP signal, and since this remains true for a three DP signal, we do not display the results obtained with MUSIC.

The studied signal is displayed in Fig. 11 (a), the DPs are located at



(a) (b) (c)

Figure 11: (a): Filtered signal with DPs located at $(l_1, l_2, l_3) = (100, 140, 300)$ with weights respectively equal to 1, 1 and -0.5; (b): Error on DPs location (formula (22)), when estimates of DPs location are computed with $FRI_{d,T}$ (with $T = 30$, $T_1 = 3.5$, $T_2 = 0.2$ and $L_{\min} = 10$); (c): Error on weights measured with $Err_n(\tilde{c}_k, c_k)$, when estimates of DPs weights are computed with $FRI_{d,T} - MLE$. In (b) and (c), the results are averaged over 200 Monte Carlo realizations.

$(l_1, l_2, l_3) = (100, 140, 300)$ with weights respectively equal to 1, 1 and -0.5. We display the results of DPs location estimation in Fig. 11 (b) using $FRI_{d,T}$, and then the estimation of the weights in Fig. 11 (c). This example illustrates that the denoising algorithm is still efficient when some of the weights are negative, and that the quality of estimation does not depend on the sign of the weights.

7.5. Generalization to Asymmetric Skewed Gaussian Filters

In the results presented so far, we have only considered a symmetric filter modeled by the Gaussian function. In this section, we broaden the scope of our study by investigating the behavior of $FRI_{d,T} - MLE$ when the filter ϕ is an asymmetric Gaussian filter with skewness α. Skewed filtered are used in many different domains of applications, as for instance source separation [37] or skewed Kalman filters for time series analysis [38]. For the sake of simplicity, we focus here on skewed Gaussian filters [39] which are defined for $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ by

$$\phi_\alpha(x) = 2\phi(x)\Phi(\alpha x), \tag{26}$$

where:

$$\Phi(\alpha x) = \int_{-\infty}^{x} \phi(t)dt = \frac{1}{2}\left[1 + erf(\frac{x}{\sqrt{2}})\right],$$

with $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$. The parameter α controls the asymmetry of the peak: the peak is right-skewed (resp. left-skewed) when $\alpha > 0$ (resp, $\alpha < 0$). Note that $\phi_0(x) = \phi(x)$. For illustration purposes, Figure 12 shows different skewed Gaussian windows.

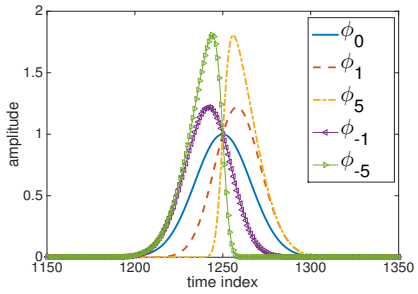


Figure 12: Illustration of Gaussian windows with different skewness

Here we study the impact of the skewness of the filter on the quality of the estimation of DPs parameters, therefore for the sake of comparison we study a signal composed of three Dirac pulses, filtered by differently skewed Gaussian windows, and estimate the weights (all equal) and locations of the DPs using $FRI_{d,T} - MLE$ (with $T = 30$). Before estimating the quality of the estimation of DPs location and weights, we investigate the percentage of non detection with respect $L_{\min}$ when the skewness of the filter varies (T, T_1 and T_2 being fixed as in the previous subsections), and for different SNRs. The criterion we use is the same as previously, i.e. $|\tilde{l}_k - l_k| < 10$ for each k. The results reported in Fig. 13 (b) and (c) show that for medium to high SNRs, there are very few non detection regardless of $L_{\min}$ and whatever the skewness. We also note that the behaviors are similar for skewness symmetric with respect to zero. In the case of a low SNR, the percentage of non detection is decreased when the modulus of the skewness is large by choosing a small $L_{\min}$ (see Fig. 13 (a)). For more symmetric filter the non detection appear to be quite insensitive to $L_{\min}$ at high noise level. The reason why one had rather use a smaller $L_{\min}$ when the modulus of the skewness increases can be explained by the fact that the filters in Fig 12 all have the same integral, and to increase the skewness increases the amplitude of the filter while decreases its essential support, making it more "peaky".

The results on the estimation of the location of the DPs are displayed in Fig. 14 (a)-(c). In these computations we use the optimal L_{min} for each skewness as suggested by the previous study. First, we remark that to change the skewness of the filter generally does not significantly alter the quality of the estimation of DPs location, and then, regarding the weight estimation, the results are also of similar quality.

8. Conclusion

In this paper, we proposed a novel algorithm for the retrieval of the parameters of a stream of Dirac pulses from noisy lowpass filtered samples. Having shown the limitations of the techniques based on the concept of signals with

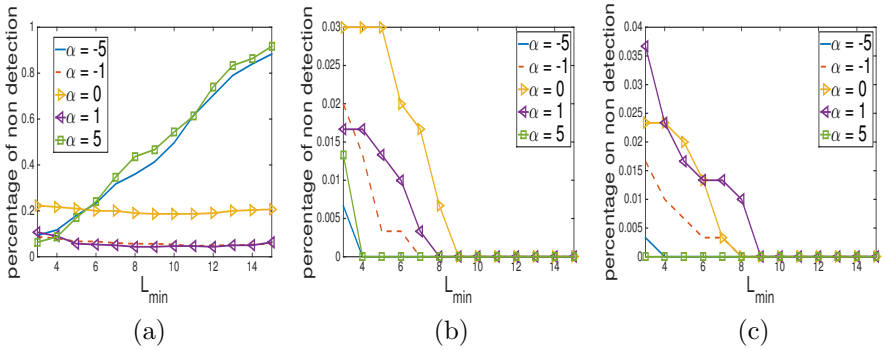


Figure 13: (a): percentage of non detection for a three Dirac pulse signal located at $l_1 = 100$, $l_2 = 140$ and $l_3 = 300$, the time delay between the closest Dirac pulses is 1 σ_ϕ, when the input SNR equal -10 dB, for different skewness for the filter and when $L_{\min}$ varies; (b): same as (a) but when the input SNR is 0 dB, (c): same as (a) but when the input SNR is 10 dB

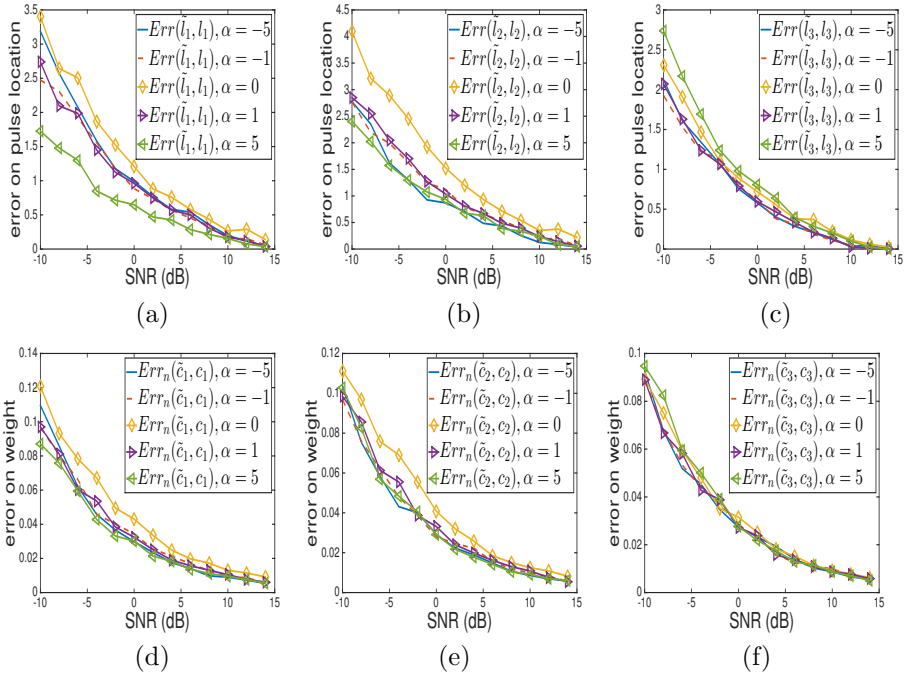


Figure 14: (a): Error on DP location for $l_1 = 100$ when the signal is the same as previously (a three Dirac pulse signal located at $l_1 = 100$, $l_2 = 140$ and $l_3 = 300$, the time delay between the closest Dirac pulses is 1 σ_ϕ), when the filter is either ϕ is either ϕ_{-5}, ϕ_{-1}, ϕ_0, ϕ_1, or ϕ_5; (b): Same as (a) but for the Dirac pulse located at $l_2 = 140$; (c): same as (a) but for the Dirac pulse located at $l_2 = 300$; (c): Error on Dirac weight associated with pulse located in l_1, when the filter is chosen among ϕ_{-5}, ϕ_{-1}, ϕ_0, ϕ_1, or ϕ_5; (e): Same as (d) but for the pulse located in l_2; (f): same as (d) but for the pulse located in l_3.

finite rate of innovation in their original formulation, we developed a new algorithm based on the denoising of the matrix used in the determination of the Dirac pulses location, and then showed that this technique enabled the determination of the number of pulses present in the stream and finally the retrieval of the parameters of close Dirac pulses, even in a very noisy context. A comparison of the proposed denoising procedure with other commonly used procedure like Cadzow denoising showed the benefit of the proposed method, and its fundamental role in the designing of efficient techniques for the retrieval of DPs parameters. In the future, this algorithm could be extended to handle non-Gaussian noise models and more complex filters.

[1] J. Lindberg, "Mathematical concepts of optical superresolution," *Journal of Optics*, vol. 14, no. 8, p. 083001, 2012.

[2] K. G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging," *Astronomy & Astrophysics*, vol. 436, no. 1, pp. 373–378, 2005.

[3] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, 2008.

[4] C. W. Mccutchen, "Superresolution in microscopy and the abbe resolution limit," *JOSA*, vol. 57, no. 10, pp. 1190–1192, 1967.

[5] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[6] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, 2017.

[7] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[8] K. Degraux, G. Peyré, J. Fadili, and L. Jacques, "Sparse support recovery with non-smooth loss functions," in *Advances in Neural Information Processing Systems*, 2016, pp. 4269–4277.

[9] Y. De Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and applications*, vol. 395, no. 1, pp. 336–354, 2012.

[10] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM journal on mathematical analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.

[11] J. Odendaal, E. Barnard, and C. Pistorius, "Two-dimensional superresolution radar imaging using the music algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 42, no. 10, pp. 1386–1391, 1994.

[12] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and C. L., "Sparse sampling of signal innovations," *IEEE Sig. Proc. Mag.*, vol. 37, no. 12, pp. 797–799, 2008.

[13] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Processing*, vol. 14, no. 1, pp. 17–47, 2015.

[14] L. Demanet, D. Needell, and N. Nguyen, "Super-resolution via superset selection and pruning," *arXiv preprint arXiv:1302.6288*, 2013.

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

[16] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

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

[18] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix," *IEEE Transactions on signal processing*, vol. 55, no. 5, pp. 1741–1757, 2007.

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

[20] X. Wei and P. L. Dragotti, "Guaranteed performance in the fri setting," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1661–1665, 2015.

[21] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.

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

[23] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8. Citeseer, 1983, pp. 336–339.

[24] V. Duval and G. Peyré, "Sparse regularization on thin grids i: the lasso," *Inverse Problems*, vol. 33, no. 5, p. 055008, 2017.

[25] K. Bredies and H. K. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2013.

[26] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on pure and applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[27] C. Dossal, M.-L. Chabanol, G. Peyré, and J. Fadili, "Sharp support recovery from noisy random measurements by ?1-minimization," *Applied and Computational Harmonic Analysis*, vol. 33, no. 1, pp. 24–43, 2012.

[28] J.-J. Fuchs, "Sparsity and uniqueness for some specific under-determined linear systems," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5. IEEE, 2005, pp. v–729.

[29] C. Dossal, "A necessary and sufficient condition for exact sparse recovery by ?1 minimization," *Comptes Rendus Mathematique*, vol. 350, no. 1-2, pp. 117–120, 2012.

[30] V. I. Morgenshtern and E. J. Candes, "Super-resolution of positive sources: The discrete setup," *SIAM Journal on Imaging Sciences*, vol. 9, no. 1, pp. 412–444, 2016.

[31] J. A. Urigüen, T. Blu, and P. L. Dragotti, "Fri sampling with arbitrary kernels," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, 2013.

[32] L. Sachs, *Applied Statistics: A Handbook of Techniques*. New York: Springer-Verlag, 1984, vol. p.253.

[33] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," *SIAM Journal on Numerical Analysis*, vol. 17, pp. 238–246, 1980.

[34] J. Cadzow, "Signal enhancement - a composite property mapping algorithm," *IEEE Trans. Accoust. Speech and Sig. Proc.*, vol. 36, no. 1, pp. 49–67, 1988.

[35] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theo.*, vol. 53, no. 12, pp. 4655 – 4666, 2007.

[36] M. Yaghoobi, D. Wu, and M. E. Davies, "Fast non-negative orthogonal matching pursuit," *IEEE Sig. Proc. Let.*, vol. 22, no. 9, pp. 1229 – 1233, 2015.

[37] W. Yu, G. Sommer, K. Danniilidis, and J. Duncan, "Using skew gabor filter in source signal separation and local spectral orientation analysis," *Image and Vision Computing*, vol. 23, pp. 377–392, 2005.

[38] P. Naveau, M. G. Genton, and X. Shen, "A skewed Kalmann filter," *Journal of multivariate analysis*, vol. 94, pp. 382–400, 2005.

[39] A. Azzalini, "A class of distributions which includes the normal ones," *Scand. J. Statist.*, vol. 12, pp. 171–178, 1985.