

SUPER-RESOLUTION OF POSITIVE SPIKES BY TOEPLITZ LOW-RANK APPROXIMATION

Laurent Condat

Univ. Grenoble Alpes, GIPSA-Lab,
F-38000 Grenoble, France
laurent.condat@gipsa-lab.grenoble-inp.fr

Akira Hirabayashi

College of Information Science and Engineering,
Ritsumeikan University,
Shiga, 525-8577, Japan
akirahr@media.ritsumei.ac.jp

ABSTRACT

Super-resolution consists in recovering the fine details of a signal from low-resolution measurements. Here we consider the estimation of Dirac pulses with positive amplitudes at arbitrary locations, from noisy lowpass-filtered samples. Maximum-likelihood estimation of the unknown parameters amounts to a difficult nonconvex matrix problem of structured low rank approximation. To solve it, we propose a new heuristic iterative algorithm, yielding state-of-the-art results.

Index Terms— Dirac pulses, sparse spike deconvolution, super-resolution, structured low rank approximation

1. INTRODUCTION AND PROBLEM FORMULATION

Recently, analog signal reconstruction from discrete measurements has been extended, from the classical paradigm rooted in Shannon's work, to a large class of signals ruled by parsimonious nonlinear models [1, 2]. Here, we focus on the emblematic problem, sometimes referred to as super-resolution [3] of estimating a set of Dirac pulses, a.k.a. spike train or Radon measure, from noisy bandlimited samples or, equivalently, from low frequency Fourier coefficients [1, 3, 4]. This problem has a wide range of applications, including ultra-wideband communications and ultrasound imaging [5]. Let τ be a positive real and let the integer $K \geq 1$ be the known number of spikes. We introduce the torus $\mathbb{T} = \mathbb{R}/\tau\mathbb{Z}$, which is the interval $[0, \tau[$ with 0 and τ identified, and the Dirac mass distribution $\delta(t)$. The unknown signal to estimate is

$$s^*(t) = \sum_{k=1}^K a_k^* \delta(t - t_k^*), \quad \forall t \in \mathbb{T}, \quad (1)$$

where the $(t_k^*)_{k=1}^K$ are the distinct locations in \mathbb{T} and the $(a_k^*)_{k=1}^K$ are the *positive* amplitudes (the star indicates an unknown quantity). The goal is to obtain estimates of these $2K$ parameters from the available data, which consists of noisy bandlimited measurements $(y_n)_{n=0}^{N-1}$, of the form

$y_n = \int_0^\tau s^*(t) \varphi(\frac{n\tau}{N} - t) dt + \varepsilon_n$, where the sampling function $\varphi(t) = (\sin(N\pi t/\tau)) / (N \sin(\pi t/\tau))$ is the Dirichlet kernel and the $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$ are independent random realizations of Gaussian noise. We suppose that N is odd and we set $M = (N - 1)/2$. We also suppose that $M \geq K$, a necessary and sufficient condition to make perfect reconstruction of s^* possible from the data in absence of noise using Prony's method, see Sect. 2.

The novelty of the present work, with respect to the previous work of the authors [4] and related works, is that the amplitudes a_k^* are supposed *positive*, a constraint which is physically grounded, if the pulse amplitudes correspond to an intensity of light or current over a zero-energy floor. Consequently, an estimation method must return estimates of the amplitudes, which are positive as well.

If we apply the discrete Fourier transform (DFT) to the sequence $(y_n)_{n=0}^{N-1}$, we obtain N consecutive lowpass Fourier coefficients of s^* . One can show [4] that they take the form

$$\hat{y}_m = \sum_{k=1}^K a_k^* e^{-j2\pi m t_k^*/\tau} + \hat{\varepsilon}_m, \quad \forall m = -M, \dots, M. \quad (2)$$

Thus, maximum likelihood (ML) estimation of the parameters amounts to solving the nonlinear least-squares problem [4]

$$\underset{\substack{(t_k^*)_{k=1}^K \in \mathbb{T}^K \\ (a_k^*)_{k=1}^K \in]0, +\infty[^K}}{\text{minimize}} \quad \frac{1}{2} \sum_{m=-M}^M \left| \hat{y}_m - \sum_{k=1}^K a_k^* e^{-j2\pi m t_k^*/\tau} \right|^2. \quad (3)$$

So, we are faced with a spectral estimation problem [6]. Solving it exactly is intractable, since the cost function has a multimodal shape with a combinatorial number of local minima and a narrow trough around the global minimum [7]. When $N \gg K$ and the locations t_k^* are not too close to each other, classical spectral estimation techniques like MUSIC or ESPRIT can be used [6]; they are fast but statistically suboptimal: contrary to ML estimation, they do not achieve the Cramér-Rao bounds when the noise level tends to zero [4]. In addition, there is no guarantee that the returned estimates of the amplitudes are positive.

In Sect. 2, we formulate ML estimation of the parameters as structured low rank approximation of a Toeplitz matrix built from the data. The proposed algorithm to solve this problem is presented in Sect. 3. Experimental results illustrate its efficiency in Sect. 4.

2. REFORMULATION OF ML ESTIMATION AS A MATRIX DENOISING PROBLEM

To reformulate ML estimation of the parameters in a form more amenable to numerical computation, we introduce the linear operator Toep , which maps a sequence or vector $(x_m)_{m=-M}^M$ with Hermitian symmetry (i.e. $x_{-m} = x_m^*$) to a square Hermitian Toeplitz matrix:

$$\text{Toep} : (x_{-M}, \dots, x_M) \mapsto \begin{pmatrix} x_0 & \cdots & x_{-M} \\ \vdots & \ddots & \vdots \\ x_M & \cdots & x_0 \end{pmatrix}. \quad (4)$$

Then, a fundamental result is:

Carathéodory's theorem [8]: Let \mathbf{T} be a Hermitian positive semi-definite Toeplitz matrix of size $(M+1) \times (M+1)$ and rank K (we have $1 \leq K \leq M$). Let $\mathbf{x} = (x_m)_{m=-M}^M$ be such that $\mathbf{T} = \text{Toep}(\mathbf{x})$; that is, the x_m are the elements on the diagonals of \mathbf{T} . Then, there exist unique (up to a permutation over the index k) sequences $(t_k)_{k=1}^K \in \mathbb{T}^K$, $(a_k)_{k=1}^K \in]0, +\infty[^K$, such that $x_m = \sum_{k=1}^K a_k e^{-j2\pi m t_k / \tau}$, for every $m = -M \dots, M$.

Thus, Toep defines a one-to-one mapping between a positive sum of complex exponentials and a Toeplitz, positive semi-definite, low rank matrix. Moreover, there is a simple procedure, due to Prony [1, 9], to find the parameters of such a matrix \mathbf{T} :

1) Re-arrange the elements x_m of \mathbf{T} in a Toeplitz matrix \mathbf{T}_K of size $(N-K) \times (K+1)$ (we have $N = 2M+1$) and rank K , as

$$\mathbf{T}_K = \begin{pmatrix} x_{-M+K} & \cdots & x_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ x_M & \cdots & x_{M-K} \end{pmatrix}. \quad (5)$$

2) Compute the right singular vector $\mathbf{h} = [h_0 \cdots h_K]^T$ of \mathbf{T}_K corresponding to the singular value zero (the smallest singular value in practice).

3) Compute the roots $(z_k)_{k=1}^K$ of the polynomial $\sum_{k=0}^K h_k z^k$. Then the locations $(t_k)_{k=1}^K$ are given by $t_k = \frac{\tau}{2\pi} \arg_{[0, 2\pi[}(z_k)$.

4) Given the locations, the vector $\mathbf{a} = [a_1 \cdots a_K]^T$ of the amplitudes is obtained by solving the least-squares linear system

$$\mathbf{U}^H \mathbf{U} \mathbf{a} = \mathbf{U}^H \mathbf{x}, \quad (6)$$

where \cdot^H is the Hermitian transpose, $\mathbf{x} = [x_{-M} \cdots x_M]^T$,

$$\mathbf{U} = \begin{pmatrix} e^{j2\pi M t_1 / \tau} & \cdots & e^{j2\pi M t_K / \tau} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi M t_1 / \tau} & \cdots & e^{-j2\pi M t_K / \tau} \end{pmatrix}. \quad (7)$$

We can note that, in absence of noise, applying this extraction procedure with $\mathbf{x} = \hat{\mathbf{y}}$ yields perfect reconstruction of the parameters $(t_k^*)_{k=1}^K$ and $(a_k^*)_{k=1}^K$. Also, in presence of noise, given estimates of the locations, the ML estimates of the amplitudes are given by (6), with \mathbf{x} replaced by $\hat{\mathbf{y}}$.

So, by virtue of Carathéodory's theorem, the ML estimation problem (3) can be rewritten as

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{C}^N}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{x} - \hat{\mathbf{y}}\|_2^2 \quad \text{s.t.} \quad \text{rank}(\text{Toep}(\mathbf{x})) \leq K \quad \text{and} \\ & \text{Toep}(\mathbf{x}) \succcurlyeq 0, \end{aligned} \quad (8)$$

where $\succcurlyeq 0$ denotes positive semi-definiteness. Since there is a one-to-one correspondence between \mathbf{x} and $\mathbf{T} = \text{Toep}(\mathbf{x})$, we can rewrite (8) as the *structured low rank approximation* (SLRA) problem

$$\begin{aligned} & \underset{\mathbf{T} \in \mathcal{H}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{T} - \text{Toep}(\hat{\mathbf{y}})\|_w^2 \quad \text{s.t.} \quad \mathbf{T} \text{ is Toeplitz} \\ & \text{and rank}(\mathbf{T}) \leq K \quad \text{and} \quad \mathbf{T} \succcurlyeq 0, \end{aligned} \quad (9)$$

where \mathcal{H} is the real Hilbert space of Hermitian matrices of size $(M+1) \times (M+1)$ and the squared weighted norm $\|\cdot\|_w^2$ is such that $\|\mathbf{x} - \hat{\mathbf{y}}\|_2^2 = \|\text{Toep}(\mathbf{x}) - \text{Toep}(\hat{\mathbf{y}})\|_w^2$; that is, $\|\mathbf{A}\|_w^2 = \sum_{i=1}^{M+1} \sum_{j=1}^{M+1} w_{i,j} |a_{i,j}|^2$, where $w_{i,j} = 1/(M+1 - |i-j|)$. So, (9) can be viewed as a *matrix denoising problem*: the matrix $\text{Toep}(\hat{\mathbf{y}})$, or equivalently the data $\hat{\mathbf{y}}$, is denoised by finding the closest matrix consistent with the model structure, from which the parameter estimates are extracted by Prony's procedure. The SLRA formulation (9) has advantages over the parametric form (3); first, there is no initialization issue, as the noisy matrix $\text{Toep}(\hat{\mathbf{y}})$ is a natural initial estimate of the solution. Second, for a low noise level, an algorithm converging to a local solution will actually find the global solution, as we observe in practice.

SLRA problems have many applications, but due to the nonconvexity of the rank constraint, they are very hard [10, 11]. Several methods have been proposed in the literature to obtain a local solution of a SLRA problem, see references in [10] and [4]. To our knowledge, the only efficient publicly available software package for SLRA is the one currently in development by I. Markovsky [12]. However, it only handles real data, whereas the matrices in (9) are complex-valued. In the next section, we propose a new algorithm, which is simple to implement and efficient. In the remainder of this section, we present two alternatives proposed in the literature to approximately solve the estimation problem.

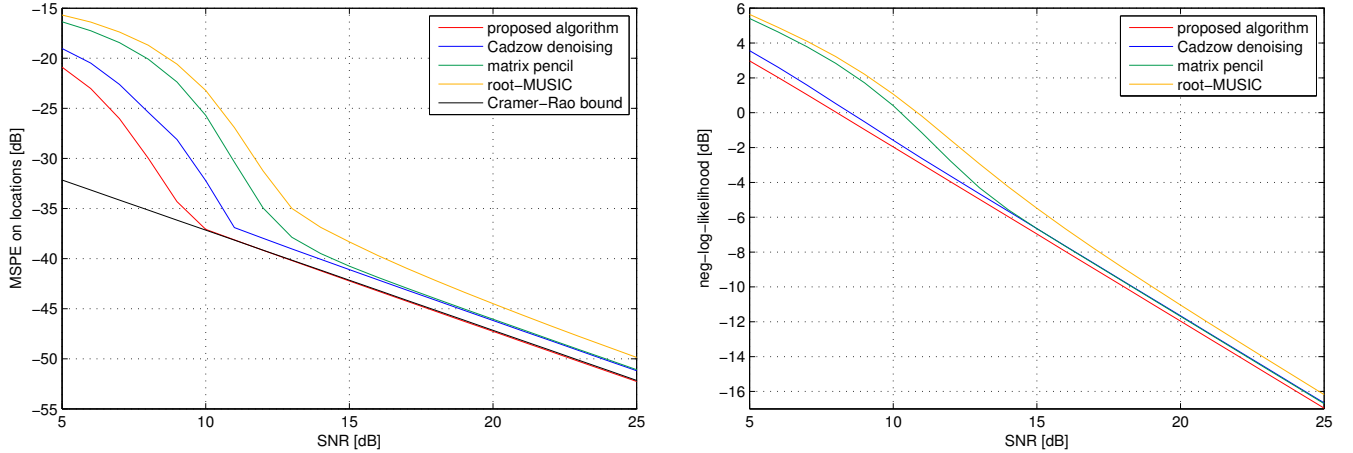


Fig. 1. Plot in log-log scale of the MSPE (left subplot) and the neg-log-likelihood (right subplot) for Experiment 1: the signal s^* consists of $K = 2$ spikes, at locations $(0.42, 0.52)$ ($\tau = 1$), with amplitudes $(1, 1)$. The data consists of $N = 11$ low-frequency samples. The values are averaged over 10^4 noise realizations.

2.1. A Convex Relaxation of the SLRA Problem

An alternative is to replace the nonconvex problem (8) by its convex surrogate

$$\underset{\mathbf{x} \in \mathbb{C}^N}{\text{minimize}} \frac{1}{2} \|\mathbf{x} - \hat{\mathbf{y}}\|_2^2 + \lambda x_0 \quad \text{s.t.} \quad \text{Toep}(\mathbf{x}) \succeq 0, \quad (10)$$

for some regularization parameter $\lambda > 0$. Indeed, from the decomposition of the solution \mathbf{x} as a positive sum of exponentials: $x_m = \sum_{k=1}^{K'} a_k e^{-j2\pi m t_k / \tau}$, $\forall m = -M, \dots, M$, we get that $x_0 = \text{tr}(\text{Toep}(\mathbf{x})) / (M + 1) = \sum_{k=1}^{K'} a_k$ is the *atomic norm* of \mathbf{x} [13]. The main advantage of this formulation is that (10) is a strongly convex semi-definite program, so that many efficient optimization algorithms exist to compute its unique solution. But it is difficult to tune the parameter λ and there is no guarantee that $K' = K$. We leave for future work the performance evaluation of this approach and the difficult theoretical question of how the minimal separation on the locations, the CRB, the SNR, N and K are related, for the solutions of (9) and (10) [14–17].

2.2. Cadzow Denoising

The simple heuristic method of alternating projections, a.k.a. *Cadzow denoising* [18], is sometimes used by practitioners to solve approximation problems with two constraint sets, by alternating projection onto them. It is promoted in [1, 5] for the reconstruction of Dirac pulses. Applied to the problem (9), and in the notations of the next section, $\mathbf{T}^{(0)} = \text{Toep}(\hat{\mathbf{y}})$ and $\mathbf{T}^{(l+1)} = P_{\Omega_2}(P_{\Omega_1}(\mathbf{T}^{(l)}))$, where l is the iteration counter. In practice, this algorithm seems to always converge to a matrix satisfying all the constraints. But it does not yield even a local solution of the problem (9), since there is no cost function minimized by this simple procedure. In Sect. 4, we use

as stopping criterion $\|\mathbf{T}^{(l+1)} - P_{\Omega_1}(\mathbf{T}^{(l)})\|_F < 1e-12$, where $\|\cdot\|_F$ is the Frobenius norm.

3. A NEW OPTIMIZATION ALGORITHM FOR SLRA

Let us first consider the generic optimization problem:

$$\underset{z \in \mathcal{Z}}{\text{minimize}} F(z) \quad \text{s.t.} \quad z \in \Omega_1 \cap \Omega_2, \quad (11)$$

where \mathcal{Z} is a real Hilbert space, Ω_1 and Ω_2 are two closed subsets of \mathcal{Z} , and $F : \mathcal{Z} \rightarrow \mathbb{R}$ is a convex and differentiable function with β -Lipschitz continuous gradient, for some $\beta > 0$; that is, $\|\nabla F(z) - \nabla F(z')\| \leq \beta \|z - z'\|$, $\forall (z, z') \in \mathcal{Z}^2$. The algorithm proposed by the first author [19] to solve this problem is:

Proposed optimization algorithm. Choose the parameters $\mu > 0$, $\gamma \in]0, 1[$, such that $2\gamma > \beta\mu$, and the initial estimates $z^{(0)}, v^{(0)} \in \mathcal{Z}$. Then iterate, for every $l \geq 0$,

$$\begin{cases} z^{(l+1)} = P_{\Omega_1}(v^{(l)} + \gamma(z^{(l)} - v^{(l)}) - \mu \nabla F(z^{(l)})), \\ v^{(l+1)} = v^{(l)} - z^{(l+1)} + P_{\Omega_2}(2z^{(l+1)} - v^{(l)}). \end{cases}$$

If the sets Ω_1 and Ω_2 are convex, the algorithm can be shown to converge to a solution of (11) under mild assumptions [19].

Then, we recognize that the SLRA problem (9) takes the form of (11), with $\mathcal{Z} = \mathcal{H}$ and the following correspondence:

- Ω_1 is the subset of \mathcal{H} of positive semi-definite matrices of rank $\leq K$. The projection $P_{\Omega_1}(\mathbf{A}) = \mathbf{Q} \text{diag}(\mathbf{d}') \mathbf{Q}^H$ onto this set, with respect to the Frobenius norm, of a matrix $\mathbf{A} = \mathbf{Q} \text{diag}(\mathbf{d}) \mathbf{Q}^H$, is obtained by truncating its eigendecomposition: the eigenvalues in \mathbf{d}' are obtained by setting all except the K largest nonnegative elements (if any) of \mathbf{d} to zero.

- Ω_2 is the linear subspace of \mathcal{H} of Toeplitz matrices. The projection P_{Ω_2} consists in replacing the elements of a same diagonal by their average.
- The cost function is $F(\mathbf{A}) = \frac{1}{2}\|\mathbf{A} - \text{Toep}(\hat{\mathbf{y}})\|_w^2$, so that $\nabla F(\mathbf{A}) = \mathbf{W} \circ (\mathbf{A} - \text{Toep}(\hat{\mathbf{y}}))$, where \circ is the entrywise product and \mathbf{W} is the matrix with the $w_{i,j}$ as elements. The Lipschitz constant of ∇F is $\beta = \max(\{w_{i,j}\}) = 1$.

In absence of convexity of the low-rank constraint set Ω_1 , the algorithm is used as a heuristic, without convergence guarantee. In practice, we set $\mu = 1.3$, $\gamma = 0.51\mu$, $z^{(0)} = v^{(0)} = \text{Toep}(\hat{\mathbf{y}})$. The stopping criterion is $\|z^{(l+1)} - P_{\Omega_2}(z^{(l+1)})\|_F < 1e-12$, and the solution estimate output by the algorithm is $P_{\Omega_2}(z^{(l+1)})$. In this setting, the algorithm was found to almost always converge to a local solution of (9). In very rare cases, when the noise level is high, the algorithm enters a cycle and does not converge. In such case, decreasing the values of μ and γ is sufficient to ensure convergence. This behavior was already observed in [4], but in the present context, the absence of convergence is even more rare, so that the additional constraint $\mathbf{T} \succcurlyeq 0$ in (9) seems to add further robustness to the algorithm.

4. EXPERIMENTAL RESULTS

The proposed algorithm is compared with Cadzow denoising, the function `rootmusic` of Matlab implementing a variant of MUSIC [20], and the matrix pencil method [21]. Since these two last methods are not guaranteed to give spikes with positive amplitudes, once the locations are found, the amplitudes are obtained by replacing the least-squares solution (6) by the non-negative least-squares solution, computed with the `lsqnonneg` command of Matlab; since such amplitudes can be zero, the method can fail to return K distinct spikes with positive amplitudes. To compare the methods, we must choose a way to quantify the estimation error between the true unknown signal s^* and its estimate. When all the amplitudes a_k^* are equal, it is possible to evaluate the error on the locations by finding the best assignment between the true spikes and their estimates. This is the case in the first experiment with $K = 2$ spikes, so that we define the mean squared periodic error (MSPE) $\min((t_1 - t_1^*)_{\tau}^2 + (t_2 - t_2^*)_{\tau}^2, (t_1 - t_2^*)_{\tau}^2 + (t_2 - t_1^*)_{\tau}^2)/2$, where $(x)_{\tau} = ((x + \frac{\tau}{2}) \bmod \tau) - \frac{\tau}{2}$. The advantage of this error criterion is that we can compute the corresponding Cramér–Rao bound (CRB), like in [1]. In the general case, since our target is ML estimation, we can simply compare the value of the neg-log-likelihood, which is the cost function in (3), (8) and (9), achieved by the different methods.

The first experiment consists in estimating $K = 2$ spikes of same amplitude; the results are reported in Fig. 1. The proposed method is the only one to achieve the CRB for a signal-to-noise ratio (SNR) (equal to $10 \log_{10}(\|\hat{\mathbf{y}} - \hat{\hat{\mathbf{e}}}\|_2^2 / \|\hat{\hat{\mathbf{e}}}\|_2^2)$)

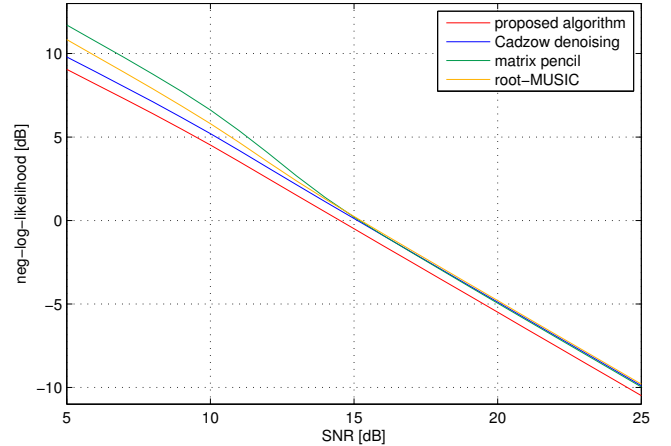


Fig. 2. Plot in log-log scale of the neg-log-likelihood for Experiment 2, see the details in Fig. 3.

larger than 10dB, which shows that it computes the ML estimate in this regime.

In the second experiment, $K = 6$ and $N = 25$; the results are reported in Figs. 2 and 3. Here also, the proposed algorithm outperforms the other methods.

We note that root-MUSIC and the matrix pencil method are noniterative, whereas Cadzow denoising and the proposed algorithm have a cost per iteration dominated by one eigen-decomposition, with complexity $O(M^3)$. Matlab code implementing the different methods and used to generate the figures is available on the webpage of the first author.

5. CONCLUSION

We proposed a new nonconvex formulation, based on structured low rank approximation, for maximum-likelihood estimation of a spike train, with arbitrary locations and positive amplitudes. We also proposed a heuristic optimization algorithm to solve this problem, which is simple to implement, fast, and efficient: it finds the global solution in practice, when the noise level is below some threshold. The perspectives for future work include the performance study in case of missing samples and other pulse shapes than the ideal Dirac spike. We will also investigate the properties of the convex relaxation of the problem.

6. REFERENCES

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

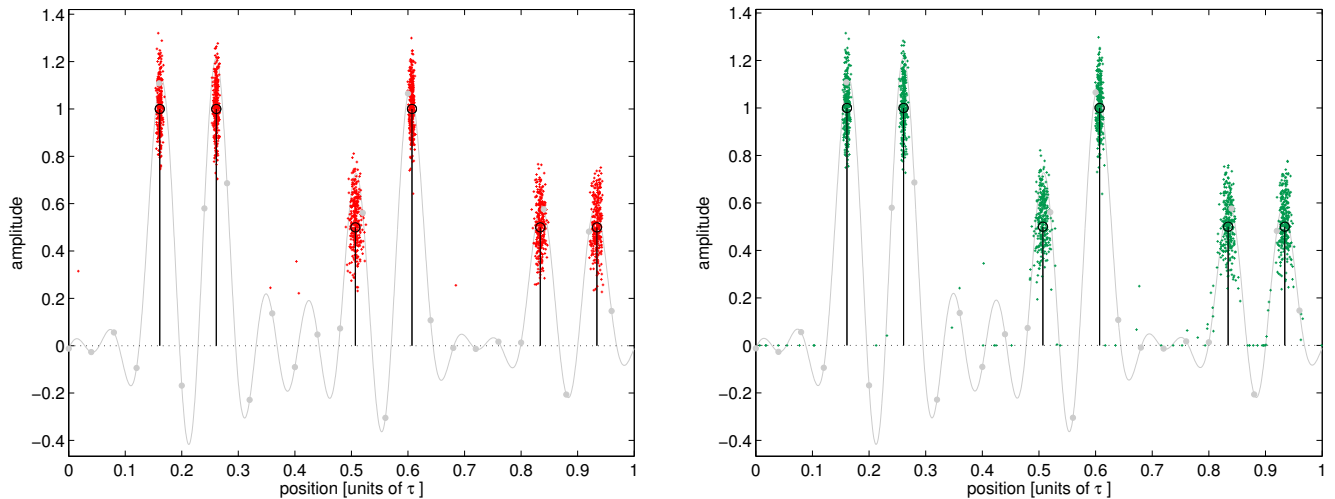


Fig. 3. Experiment 2: the signal s^* consists of $K = 6$ spikes, at locations $(0.161, 0.261, 0.507, 0.607, 0.834, 0.934)$ ($\tau = 1$), with amplitudes $(1, 1, 0.5, 1, 0.5, 0.5)$, represented in black. The data consists of $N = 25$ low-frequency samples (gray dots) corrupted by noise with $\text{SNR} = 12\text{dB}$. The red dots (left subplot) and the green dots (right subplot) show the spike estimates obtained with the proposed method and the matrix pencil method, respectively, for 300 different noise realizations. Note that the proposed method is guaranteed to provide $K = 6$ different spikes with positive amplitudes, which is not the case of the matrix pencil method, see the many dots with amplitude zero.

- and G. Kutyniok, Eds., pp. 148–209. Cambridge University Press, 2012.
- [3] E. J. Candès and C. Fernandez-Granda, “Super-resolution from noisy data,” *Journal of Fourier Analysis and Applications*, vol. 13, no. 6, pp. 1229–1254, Dec. 2013.
- [4] 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.
- [5] R. Tur, Y. C. Eldar, and Z. Friedman, “Innovation rate sampling of pulse streams with application to ultrasound imaging,” *IEEE Trans. Signal Processing*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.
- [6] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, NJ, 2005.
- [7] J. Gillard and A. Zhigljavsky, “Analysis of structured low rank approximation as an optimization problem,” *Informatica*, vol. 22, no. 4, pp. 489–505, 2011.
- [8] C. Carathéodory, “Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen,” *Rendiconti del Circolo Matematico di Palermo*, vol. 3, pp. 193–217, 1911.
- [9] D. Potts and M. Tasche, “Parameter estimation for nonincreasing exponential sums by Prony-like methods,” *Linear Algebra and its Applications*, vol. 439, no. 4, pp. 1024–1039, 2013.
- [10] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*, Springer, 2012.
- [11] J. W. Gillard and A. A. Zhigljavsky, “Optimization challenges in the structured low rank approximation problem,” *Journal of Global Optimization*, vol. 53, pp. 733–751, 2013.
- [12] I. Markovsky and K. Usevich, “Software for weighted structured low-rank approximation,” *J. Comput. Appl. Math.*, vol. 256, pp. 278–292, 2014.
- [13] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Trans. Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
- [14] D. L. Donoho, “Superresolution via sparsity constraints,” *SIAM J. Math. Anal.*, vol. 23, pp. 1309–1331, Sept. 1992.
- [15] L. Demanet and N. Nguyen, “The recoverability limit for super-resolution via sparsity,” preprint arXiv:1502.01385, 2015.
- [16] V. I. Morgenshtern and E. J. Candès, “Super-resolution of positive sources: the discrete setup,” preprint arXiv:1504.00717, 2015.
- [17] X. Wei and P. L. Dragotti, “Guaranteed performance in the FRI setting,” *IEEE Signal Processing Letters*, 2015, to appear.
- [18] 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.
- [19] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.
- [20] A. Barabell, “Improving the resolution performance of eigenstructure-based direction-finding algorithms,” in *Proc. of IEEE ICASSP*, Apr. 1983, pp. 336–339.
- [21] Y. Hua and T. K. Sarkar, “Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.