Robust Spike Train Recovery from Noisy Data by Structured Low Rank Approximation

Laurent Condat
GIPSA-lab, University of Grenoble, Grenoble, France
Contact: see http://www.gipsa-lab.grenoble-inp.fr/∼laurent.condat/

Akira Hirabayashi
Yamaguchi University, Yamaguchi, Japan

***Abstract*—We consider the recovery of a finite stream of Dirac pulses at nonuniform locations, from noisy lowpass-filtered samples. We show that maximum-likelihood estimation of the unknown parameters amounts to solve a difficult, even believed NP-hard, matrix problem of structured low rank approximation. We propose a new heuristic iterative optimization algorithm to solve it. Although it comes, in absence of convexity, with no convergence proof, it converges in practice to a local solution, and even to the global solution of the problem, when the noise level is not too high. Thus, our method improves upon the classical Cadzow denoising method, for same implementation ease and speed.**

I. Introduction and Problem Formulation

Reconstruction of signals lying in linear spaces, including bandlimited signals and splines, has long been the dominant paradigm in sampling theory, rooted in Shannon's work. Recently, analog reconstruction from discrete samples has been enlarged to a broader class of signals, with so-called finite rate of innovation, i.e. ruled by parsimonious models [1]–[3]. This theory predates and parallels the emerging framework of sparse recovery and compressed sensing [4]. The most studied problem in this context, on which we focus in this paper, is the recovery of a finite stream of Dirac pulses, a.k.a. a spike train, from uniform, noisy, lowpass-filtered samples [1], [5]–[8].

More precisely, the sought-after unknown signal s consists of K Dirac pulses in the finite interval $[0,\tau[$, where the real $\tau > 0$ and the integer $K \geq 1$ are known; that is

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

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

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

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

where $\varphi(t)$ is the sampling function and the $\varepsilon_n \sim \mathcal{N}(0,\sigma^2)$ are independent random realizations of Gaussian noise. Note that other noise models could be considered as well, by changing the cost function in eqns. (5), (7), (9) below.

The questions of the choice of the function φ and of the number N of measurements allowing perfect reconstruction, in absence of noise, has been addressed in the literature [6], [7], [9]. In a nutshell, the condition $N \geq 2K + 1$, which we hereafter assume to be true, is necessary and sufficient, provided that φ satisfies some constraints in Fourier domain. Additionally, we assume, without loss of generality and only to simplify the notations, that N is odd, of the form $N = 2M + 1$. Since our emphasis here is on appropriately handling the presence of noise and not on being the most general, we adopt the simplest choice of the Dirichlet sampling function [6], which amounts to periodizing the signal s on the real line before sampling it with the sinc kernel:

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

The extension of the setting to the reconstruction of pulses with real shape, instead of the ideal Dirac distribution, is of obvious practical interest in ultrawideband communications [2] or to detect impulsive signals in biomedical applications [6]. This generalization, or equivalently the choice of another sampling function φ, can be done without difficulty, as shown in [6], and will not be addressed here.

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

II. Maximum Likelihood Parameter Estimation

A natural approach to solve parametric estimation problems is maximum likelihood (ML) estimation; it consists in selecting the model which is the most likely to explain the observed noisy data. In our case, as we have assumed Gaussian noise, this corresponds to solving the nonlinear least-squares problem [10]:

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

Now, applying the discrete Fourier transform to the vector of samples $\{v_n\}_{n=0}^{N-1}$ yields the Fourier coefficients defined by

$\hat{v}_m = \sum_{n=0}^{N-1} v_n e^{-j2\pi mn/N}, \quad \forall m = -M, \dots, M$. We define the Fourier coefficients $\{\hat{\varepsilon}_m\}_{m=-M}^{M}$ similarly. Then, it is easy to show that

$$\hat{v}_m = \sum_{k=1}^{K} a_k e^{-j2\pi m t_k/\tau} + \hat{\varepsilon}_m, \quad \forall m = -M, \dots, M. \quad (6)$$

Since the inverse discrete Fourier transform is unitary, up to a constant, the problem (5) can be rewritten as [10]:

$$\underset{\substack{\{t'_k\}_{k=1}^K \in [0,\tau[^K \\ \{a'_k\}_{k=1}^K \in \mathbb{R}^K}}{\text{minimize}} \sum_{m=-M}^{M} \left| \hat{v}_m - \sum_{k=1}^{K} a'_k e^{-j2\pi m t'_k/\tau} \right|^2. \quad (7)$$

Thus, (7) takes the form of a spectral estimation problem, which consists in retrieving the parameters of a sum of complex exponentials from noisy samples [11]. However, solving (7) is very difficult task, as the function to minimize is oscillating, with many local minima [12]. Numerous methods have been proposed to find a local minimum of the cost function in (7). They mostly proceed by iteratively refining an initial estimate of the solution, which has to be already of good quality. Also, when $N \gg K$ and the locations t_k are not too close to each other, classical spectral estimation techniques like MUSIC and ESPRIT can be used; they are fast but statistically suboptimal. The main advantage of the proposed approach is that it gets rid of such limitations, without any simplifying assumption.

III. Prony's Annihilation Property: Reformulation as Matrix Approximation Problem

Let us assume temporarily that there is no noise, i.e. $\hat{\varepsilon}_m = 0$ in (6). Then, the sequence of Fourier coefficients $\{\hat{v}_m\}_{m=-M}^{M}$ can be *annihilated*, a known property which dates back to Prony's work in the eighteenth century [13]. That is, its convolution with the sequence $\{h_k\}_{k=0}^{K}$ is identically zero: $\sum_{k=0}^{K} h_k \hat{v}_{m-k} = 0, \quad \forall m = -M+K, \dots, M$, where the annihilating filter h is defined, up to a constant, by $\sum_{k=0}^{K} h_k z^k = \prod_{k=1}^{K} (z - e^{j2\pi t_k/\tau})$. In matrix form, the annihilation property is

$$\underbrace{\begin{pmatrix} \hat{v}_{-M+K} & \cdots & \hat{v}_{-M} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \hat{v}_M & \cdots & \hat{v}_{M-K} \end{pmatrix}}_{\mathbf{T}_K} \begin{pmatrix} h_0 \\ \vdots \\ h_K \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (8)$$

Let us choose an integer P in $K, \dots, M$ and define the Toeplitz—i.e. with constant values along its diagonals—matrix $\mathbf{T}_P$, of size $N-P \times P+1$, obtained by arranging the values $\{\hat{v}_m\}_{m=-M}^{M}$ in its first row and column; $\mathbf{T}_K$ is depicted in (8). Then, the existence of an annihilating filter of size $K+1$ for the sequence $\{\hat{v}_m\}_{m=-M}^{M}$ is completely equivalent to the property that $\mathbf{T}_P$ has rank at most K.

Hence, turning back to the case when noise is present in the data, we can rewrite (7) as the following *structured low rank approximation* (SLRA) matrix problem:

$$\text{Find } \widetilde{\mathbf{T}}_P \in \underset{\mathbf{T}' \in \mathbb{C}^{N-P \times P+1}}{\arg\min} \|\mathbf{T}' - \mathbf{T}_P\|_w^2 \quad \text{s. t.} \quad \mathbf{T}' \text{ is Toeplitz} \quad \text{and} \quad \text{rank}(\mathbf{T}') \leq K, \quad (9)$$

where the weighted Frobenius norm of a matrix $\mathbf{A} = \{a_{i,j}\} \in \mathbb{C}^{N-P \times P+1}$ is defined by $\|\mathbf{A}\|_w^2 = \sum_{i=1}^{N-P} \sum_{j=1}^{P+1} w_{i,j} |a_{i,j}|^2$ and $w_{i,j}$ is the inverse of the size of the diagonal going through the position (i,j), see formula in [14, eq. (16)].

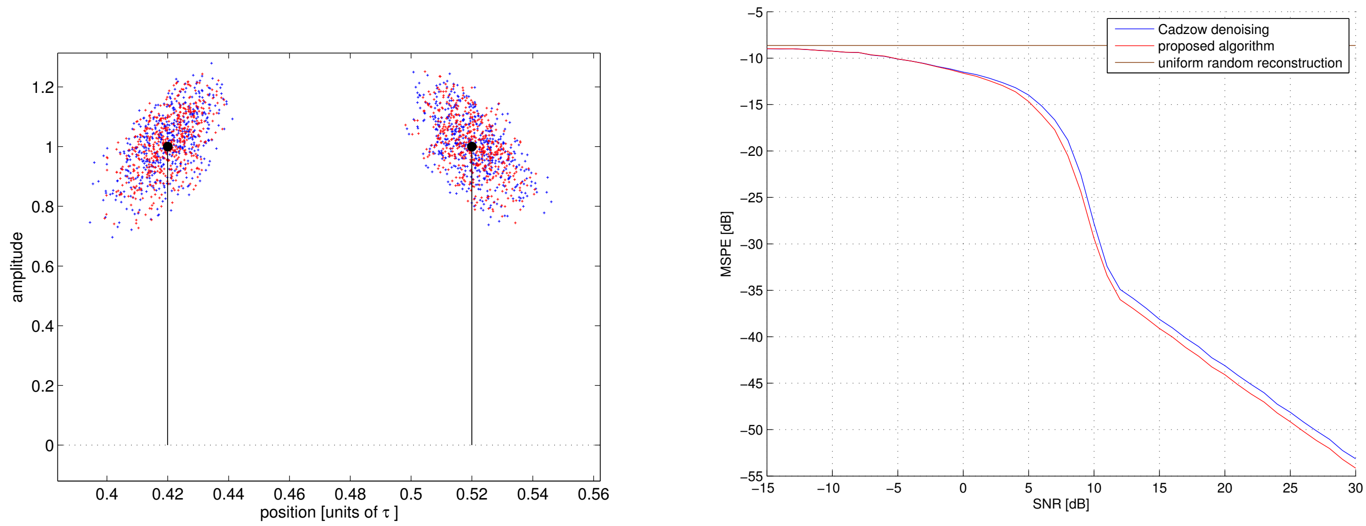
After the SLRA problem (9) has been solved, the procedure to recover the estimates of the parameters is the following [1]. First, reshape the obtained Toeplitz matrix $\widetilde{\mathbf{T}}_P$ to a Toeplitz matrix $\widetilde{\mathbf{T}}_K$ of size $N-K \times K+1$. Second, compute the right singular vector $\tilde{\mathbf{h}} = \{\tilde{h}_k\}_{k=0}^{K}$ of $\widetilde{\mathbf{T}}_K$, corresponding to the singular value 0. Third, compute the roots $\{\tilde{z}_k\}_{k=1}^{K}$ of the polynomial $\sum_{k=0}^{K} \tilde{h}_k z^k$; the estimates $\{\tilde{t}_k\}_{k=1}^{K}$ of the locations are given by $\tilde{t}_k = \frac{\tau}{2\pi} \arg_{[0,2\pi[}(\tilde{z}_k)$. Fourth, the estimates $\{\tilde{a}_k\}_{k=1}^{K}$ of the amplitudes are obtained by solving the linear system $\widetilde{\mathbf{U}}^{\mathrm{H}} \widetilde{\mathbf{U}} \tilde{\mathbf{a}} = \widetilde{\mathbf{U}}^{\mathrm{H}} \hat{\mathbf{v}}$, where $\hat{\mathbf{v}} = [\hat{v}_{-M} \ \cdots \ \hat{v}_M]^{\mathrm{T}}$, $\cdot^{\mathrm{H}}$ denotes the Hermitian transpose, and

$$\widetilde{\mathbf{U}} = \begin{pmatrix} e^{j2\pi M \tilde{t}_1/\tau} & \dots & e^{j2\pi M \tilde{t}_K/\tau} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi M \tilde{t}_1/\tau} & \dots & e^{-j2\pi M \tilde{t}_K/\tau} \end{pmatrix}. \quad (10)$$

We note that this procedure yields the ML estimates solution to (7), only if the roots $\{\tilde{z}_k\}_{k=1}^{K}$ are all on the complex unit circle. This is the case, by centro-Hermitian symmetry of the matrices, except if the noise level is too high; in this case, two roots could merge and then split in a pair $(\tilde{z}_k, \tilde{z}_{k'} = 1/\tilde{z}_k^*)$ on both sides of the unit circle, yielding $\tilde{t}_k = \tilde{t}_{k'}$.

Thus, the proposed process consists in *denoising* the matrix $\mathbf{T}_P$, or equivalently the measurements $\{v_n\}_{n=0}^{N-1}$, by finding the closest data consistent with the model's structure, from which the parameters are estimated by Prony's method. In absence of noise, the parameters are perfectly recovered. However, the SLRA problem (9) at the heart of the procedure, which consists in projecting a matrix on a nonconvex manifold, is believed to be NP-hard [15]. Yet, the main advantage of the SLRA formulation, compared to (7), is that there is no initialization problem: an iterative algorithm to solve (9) proceeds directly, with the noisy matrix $\mathbf{T}_P$ as initial estimate of the solution $\widetilde{\mathbf{T}}_P$. Moreover, for a low noise level, an algorithm converging to a local solution will actually find the global solution $\widetilde{\mathbf{T}}_P$, as we observe in practice.

We now tackle the state-of-the-art to solve SLRA problems, which have a wide range of applications [15]. A few algorithms, able to find a local solution of the SLRA problem (9), have been proposed in the community of numerical algebra [16]–[18]. For instance, the iterative approach in [16] is based on a BFGS quasi-Newton solver. Besides the difficulty of implementation, the algorithm is very costly, as it requires computing many singular value decompositions (SVD) at each iteration. To our knowledge, the only publicly available software package for SLRA is the one currently in development by Ivan Markovsky [19]. However, it only handles real-valued,



(a) (b)

Fig. 1. The signal to estimate from $N = 11$ noisy measurements, consists in $K = 2$ Dirac pulses. The true parameters are $(t_1, t_2) = (0.42, 0.52)$ and $(a_1, a_2) = (1, 1)$, with $\tau = 1$ and $P = M = 5$. (a) In black, the true pulses. In blue and red, the locations and amplitudes reconstructed by Cadzow denoising and the proposed algorithm, respectively, for 500 different noise realizations. The signal-to-noise-ratio (SNR) was 15dB and the computation time for every reconstruction, with 50 iterations, was 14ms. The proposed method yields lower errors, with a points cloud slightly less dispersed. (b) Plot in log-log scale of the mean squared periodic error (MSPE) on the locations $\min\big((\tilde{t}_1 - t_1)_\tau^2 + (\tilde{t}_2 - t_2)_\tau^2, (\tilde{t}_1 - t_2)_\tau^2 + (\tilde{t}_2 - t_1)_\tau^2\big)$, where $(x)_\tau = \big((x + \frac{\tau}{2}) \mod \tau\big) - \frac{\tau}{2}$, averaged over 10,000 noise realizations for every SNR value. An upper bound of the error is given by the naive estimator, which sets the locations randomly and uniformly in $[0, \tau[$.

and not complex-valued, matrices. We note that replacing in the problem the rank by its convex surrogate, the nuclear norm, does not perform well in our setting, where two close pulses yield highly coherent measurements [20]. Thus, practitioners rely on a popular heuristic method, called *Cadzow denoising* [21], which is used in [1], [6] for the recovery of Dirac pulses. This algorithm consists in denoising the matrix $\mathbf{T}_P$ by alternating projections: at each iteration, the matrix is replaced by its closest, in Frobenius norm, matrix of rank at most K, and then the obtained matrix is replaced by its closest Toeplitz matrix. Although Cadzow denoising seems to always converge in practice to a Toeplitz matrix of rank at most K, there exists no global proof of convergence to date, contrary to a common belief [22]. Anyways, the obtained matrix is not a local minimizer of the cost function $\|\cdot - \mathbf{T}_P\|_w^2$ [12], [16]. In the next section, we propose a new algorithm to compute a local solution of the SLRA problem (9), thus improving theoretically upon Cadzow denoising.

IV. A New Optimization Method for SLRA

Let us consider the generic optimization problem:

$$\text{Find } \tilde{x} \in \operatorname*{arg\,min}_{x \in \mathcal{H}} F(x) \quad \text{s.t.} \quad x \in \Omega_1 \cap \Omega_2, \qquad (11)$$

where $\mathcal{H}$ is a real Hilbert space of finite dimension, Ω_1 and Ω_2 are two closed subsets of $\mathcal{H}$, and $F : \mathcal{H} \to \mathbb{R}$ is a differentiable function with Lipschitz-continuous gradient; that is, there exists some $\beta > 0$ such that $\|\nabla F(x') - \nabla F(x)\| \leq \beta \|x - x'\|$, $\forall x, x' \in \mathcal{H}$. Recently [23], the first author proposed a new algorithm to solve (11):

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

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

where P_Ω denotes the closest-point projection onto $\Omega \subset \mathcal{H}$. It has been proved in [23] that if Ω_1 and Ω_2 are convex and $2\gamma > \beta\mu$, the sequence $(x^{(i)})_{i \in \mathbb{N}}$ converges to some element $\tilde{x}$ solution to the problem (11).

In absence of convexity, this result does not apply, so that we will use the method as a heuristic, without guarantee of convergence. The SLRA problem (9) can be recast as an instance of (11) as follows: $\mathcal{H} = \mathbb{C}^{N-P \times P+1}$ is the real Hilbert space of complex-valued matrices of size $N - P \times P + 1$ with centro-Hermitian symmetry, endowed with Frobenius inner product $\langle \mathbf{X}, \mathbf{X}' \rangle = \sum_{i,j} x_{i,j} x'^{*}_{i,j}$; Ω_1 is the closed nonconvex subset of $\mathcal{H}$ of matrices with rank at most K; Ω_2 is the linear subspace of $\mathcal{H}$ of Toeplitz matrices. The operations involved in the algorithm are the following:

- P_{Ω_1} corresponds to SVD truncation, according to the Schmidt-Eckart-Young theorem: if a matrix $\mathbf{X}$ has SVD $\mathbf{X} = \mathbf{L\Sigma R}^{\mathrm{H}}$, then $P_{\Omega_1}(\mathbf{X})$ is obtained by setting to zero the singular values in $\mathbf{\Sigma}$, except the K largest.
- The "Toeplitzation" operation P_{Ω_2} simply consists in averaging along the diagonals of the matrix.
- The cost function is $F(\mathbf{X}) = \frac{1}{2}\|\mathbf{X} - \mathbf{T}_P\|_w^2$, so that $\nabla F(\mathbf{X}) = \mathbf{W} \circ (\mathbf{X} - \mathbf{T}_P)$, where $\circ$ is the entrywise (Hadamard) product and the matrix $\mathbf{W}$ has entries $\{w_{i,j}\}$.

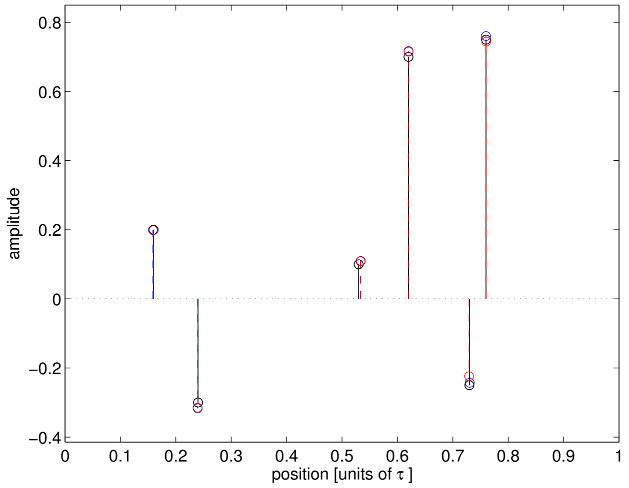


Fig. 2. The signal consists of $K = 6$ Dirac pulses. We have $N = 25$ noisy measurements with SNR=25dB. In black: true pulses. In blue and red: reconstructed positions and amplitudes of the pulses with Cadzow denoising and the proposed algorithm, respectively. The computation time, with 50 iterations, was 19ms in both cases.

The Lipschitz constant of ∇F is $\beta = \max(\{w_{i,j}\}) = 1$.

We observed empirically that the proposed algorithm always converges, for an appropriate choice of μ and γ. Moreover, the matrix obtained at convergence is always Toeplitz, of rank at most K, and a local solution to (9); see more details in [14].

We show in Fig. 1 a comparison with Cadzow denoising for the recovery of $K = 2$ Dirac pulses from $N = 11$ measurements. We observe that the estimation error on the pulses' locations is about 10% lower in average with our method. We recognize that this improvement is small for the simple setting considered here, with ideal Dirac pulses and a sinc sampling kernel. Our ongoing work is to investigate more general scenarios, with pulses having real shape and noise which is not white and Gaussian. We expect the improvement of our method over Cadzow denoising to be more significant in such cases. Yet, we emphasize that both methods have essentially the same complexity and convergence speed, dominated by one SVD per iteration. Another example is given in Fig.2 and experiments with larger size are shown in the extended version of this paper [14].

V. Conclusion

We proposed a new heuristic optimization algorithm to solve structured low rank approximation problems. For the recovery of Dirac pulses, this efficient matrix denoising procedure, combined with Prony's extraction method, yields the maximum-likelihood parameter estimates, up to some threshold SNR. Many theoretical questions related to the performances of the approach are open and currently investigated by the authors. Especially, stability guarantees similar to the ones recently developed for a convex relaxation of the problem [8], [24], are sought after. A Matlab implementation of the proposed method is available on the webpage of the first author.

Acknowledgment

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

References

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

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

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

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

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

[6] 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.

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

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

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

[10] A Hirabayashi, T. Iwami, S. Maeda, and Y. Hironaga, "Reconstruction of the sequence of Diracs from noisy samples via maximum likelihood estimation," in *Proc. of IEEE ICASSP*, 2012, pp. 3805–3808.

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

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

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

[14] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements," preprint hal-00759253, 2012.

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

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

[17] M. Schuermans, *Weighted low rank approximation: Algorithms and applications*, Ph.D. thesis, Katholieke Universiteit Leuven, 2006.

[18] R. Borsdorf, *Structured Matrix Nearness Problems: Theory and Algorithms*, Ph.D. thesis, The University of Manchester, UK, June 2012.

[19] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," Tech. Rep. 339974, ECS, Univ. of Southampton, 2012, documention of a software package, see `https://github.com/slra/slra`.

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

[21] 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.

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

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

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