RESOLUTION ENHANCEMENT OF ULTRASONIC SIGNALS BY UP-SAMPLED SPARSE DECONVOLUTION

Ewen Carcreff[1,2], Sébastien Bourguignon[1], Jérôme Idier[1], Laurent Simon[2]

[1] LUNAM Université, École Centrale de Nantes / IRCCyN UMR CNRS 6597, 1 rue de la Noë, 44321 Nantes, France
[2] LUNAM Université, Université du Maine / LAUM UMR CNRS 6613, avenue Olivier Messiaen, 72085, Le Mans, France

ABSTRACT

This paper deals with the estimation of the arrival times of overlapping ultrasonic echoes. We focus on approaches based on discrete sparse deconvolution. Such methods are limited by the time resolution imposed by the model discretization, which is usually considered at the data sampling rate. In order to get closer to the continuous-time model, we propose to increase the time precision by introducing an up-sampling factor in the discrete model. The problem is then recast as a Multiple Input Single Output (MISO) deconvolution problem. Then, we propose to revisit standard sparse deconvolution algorithms for MISO systems. Specific and efficient algorithmic implementation is derived in such setting. Algorithms are evaluated on synthetic data, showing improvements in robustness toward discretization errors and competitive computational time compared to the standard approaches.

Index Terms— deconvolution, sparse approximation, MISO systems, ultrasonic data.

1. INTRODUCTION

Estimation of arrival times and amplitudes of superimposed echoes from noisy observations arises in many applications such as RADAR, seismic exploration, ultrasonic nondestructive testing (NDT) or medical imaging. In NDT for example, a known waveform is sent through a material, and reflection occurs at each impedance change. The precise estimation of the echo parameters then leads to the localization and the characterization of the geometrical properties (including flaws) of the inspected object. Consider the signal model:

$$y(t) = \sum_i a_i h(t - t_i) + e(t), \qquad (1)$$

where $h(t)$ is the *a priori* known waveform and $e(t)$ stands for additive noise. The purpose is then to estimate the parameters t_i and a_i from sampled data $\boldsymbol{y} = [y(nT_S)]_{n=1,...,N}$, where T_S is the sampling period. This can be a hard task when the

echoes overlap, creating constructive or destructive interferences. It is particularly critical for ultrasonic data, where the waveform $h(t)$ has generally a strongly oscillating shape.

Many approaches aim to identify a_i and t_i in Eq. (1) as continuous parameters. Cross-correlation methods [1] are computationally simple but show poor performance when echoes overlap. The finite rate of innovation theory [2] offers exact reconstruction provided that the sampling rate is high enough, although in a rather different context than ours – in particular, specific *sampling kernels* $h(t)$ are considered. Related subspace-based methods have also been widely used in this context, even though they are better suited to multiple snapshot data in order to yield robust covariance estimates [3]. For ultrasonic NDT, parametric methods were proposed where (a_i, t_i) are jointly estimated together with shape parameters for each echo, by minimizing a least-squares distance [4]. This is hence a nonlinear approach that can be very sensitive to model errors and local minima.

On the other hand, extensive research has been carried out on deconvolution methods. Indeed, Eq. (1) formulates a continuous convolution :

$$y(t) = (h*x)(t) + e(t) = \int_{-\infty}^{+\infty} h(\tau)\, x(t-\tau)\mathrm{d}\tau + e(t), \ (2)$$

where $x(t)$ is a spike train with time positions t_i and amplitudes a_i. Many deconvolution methods then consider a discretized version of the right-hand term in Eq. (2), which yields a discrete and linear inverse problem $\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{e}$. Regularization is then addressed by introducing a sparsity constraint on the sequence $\boldsymbol{x}$. By exploiting linearity with appropriate regularization techniques, such approaches have shown satisfactory results in the presence of strong overlapping and noise [5]. However, the time resolution is obviously limited by the discretization precision, that usually corresponds to the data sampling frequency.

In this paper, we show that it is practically possible to increase the time resolution in sparse deconvolution algorithms by an up-sampling approach. This is relevant to estimate times of flight, which are continuous values. We propose to revisit well-known sparse approximation algorithms in this context, based on greedy strategies [6, 7, 8], and on ℓ_0 [9] and ℓ_1-norm penalization [10, 11]. Section 2 establishes the

discrete up-sampled convolution model, which is recast as a MISO system. Based on the resulting structure of matrix $\mathbf{H}$, Section 3 studies implementation issues of sparse approximation for up-sampled deconvolution and, more generally, for the estimation of sparse inputs in MISO systems. From synthetic data, a comparison between standard and up-sampled deconvolution is conducted in Section 4. Algorithms are also compared in terms of computational efficiency and performance through Monte-Carlo simulations. Conclusions are finally given in Section 5.

2. UP-SAMPLED CONVOLUTION AS A MISO SYSTEM

Consider the continuous-time convolution model (2), where available data is sampled at period T_S: we note $y_n = y(nT_S)$. Up to our knowledge, all the works in the field of deconvolution consider a *discrete convolution* model, that reads:

$$y_n = \sum_{m=0}^{M-1} h_m \, x_{n-m} + e_n. \tag{3}$$

That is, the right-hand term in Eq. (2) is sampled at the data sampling rate : $h_n = h(nT_S)$ and $x_n = x(nT_S)$. Note that the error term e_n should now also include model errors due to inexact discretization. Let column vectors $\boldsymbol{y}$, $\boldsymbol{h}$, $\boldsymbol{x}$ and $\boldsymbol{e}$ collect the samples of y_n, h_n, x_n and e_n, respectively. Eq. (3) then reads $\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{e}$ where $\mathbf{H}$ is a convolution matrix, whose n-th line is a delayed version of the reversed sequence $[h_{M-1}, \ldots, h_0]$ with $n - 1$ zeros inserted at the beginning. The Toeplitz structure of $\mathbf{H}$ can be exploited to perform efficient computations with Fast Fourier Transform (FFT) algorithms [12]. Note that such definition of $\mathbf{H}$ corresponds to the post-windowing boundary assumption, for which $\boldsymbol{x} = [x_{-M+2}, \ldots, x_0, x_1, \ldots, x_N]^T$, where superscript ^{T} denotes the transposition.

However, in many applications, the data sampling rate is limited and such discretization may not be appropriate. This is particularly true for sparse deconvolution, since the searched sequence is not band-limited. Hence, it may be of interest to consider that h and x in Eq. (2) are discretized at rate T_S/K with K integer. The discrete model becomes:

$$y_n = \sum_{p=0}^{P-1} \overline{h}_p \overline{x}_{nK-p} + e_n \tag{4}$$

with $\overline{h}_p = h(pT_S/K)$, $\overline{x}_p = x(pT_S/K)$ and $P = KM$. Let column vectors $\overline{\boldsymbol{h}}$ and $\overline{\boldsymbol{x}}$ collect the samples $\overline{h}_n$ and $\overline{x}_n$. In matrix form, model (4) reads $\boldsymbol{y} = \overline{\mathbf{H}}\,\overline{\boldsymbol{x}} + \boldsymbol{e}$ where each line of $\overline{\mathbf{H}}$ is formed by the reversed sequence $[\overline{h}_{P-1} \ldots \overline{h}_0]$, with $(n-1)K$ zeros inserted at the beginning of line n. $\overline{\mathbf{H}}$ is now an $N \times KN$ matrix and is no more Toeplitz. One can show, however, that model (4) also reads as the sum of K discrete convolutions:

$$y_n = \sum_{k=1}^{K} \left(\sum_{m=0}^{M-1} h_m^k \, x_{n-m}^k \right) + e_n, \tag{5}$$

where $\boldsymbol{h}^k$, $k = 1 \ldots K$ are K *sub-waveforms* with sampling period T_S, such that $h_m^k = h\big((k-1)T_S/K + mT_S\big)$. Similarly, $\boldsymbol{x}^k$ are the corresponding sparse sub-sequences with N points. The matrix form hence reads $\boldsymbol{y} = \sum_{k=1}^{K} \mathbf{H}^k \, \boldsymbol{x}^k + \boldsymbol{e}$, where $\mathbf{H}^k$ are Toeplitz *sub-matrices* obtained by taking every K columns of $\overline{\mathbf{H}}$. In other words, it can be seen as a specific MISO system, as illustrated in Figure 1, where the K filters are obtained by sampling the continuous-time impulse response $h(t)$ at period T_S, with K subsample time shifts $(k-1)T_S/K$, $k = 1 \ldots K$. In the following section, we describe the algorithmic implementations of sparse deconvolution for generic MISO systems.
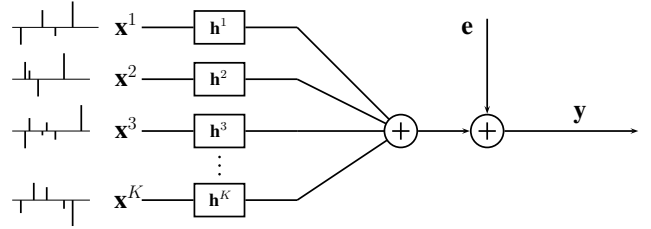

Fig. 1. Diagram of a MISO system.

3. MISO DECONVOLUTION WITH SPARSE APPROXIMATION METHODS

Sparse approximation has become an important field of research in the past fifteen years [11]. It aims at approximating the data $\boldsymbol{y}$ with $\overline{\mathbf{H}}\overline{\boldsymbol{x}}$ where $\overline{\boldsymbol{x}}$ is a sparse sequence[1], that is, $\overline{\boldsymbol{x}}$ has only few non-zero components. We focus on implementation issues for five acknowledged sparse approximation methods applied to MISO deconvolution:

- Three *greedy* algorithms are implemented, namely, by increasing complexity: Matching Pursuit (MP) [6], Orthogonal Matching Pursuit (OMP) [7] and Orthogonal Least Squares (OLS) [8]. Each iteration of such procedures comprises the selection of one component improving the data approximation, and the update of the solution as a combination of the selected components.

- The Single Best Replacement algorithm was recently introduced in [9] and performs local minimization of the penalized least-squares criterion:

$$\left\| \boldsymbol{y} - \overline{\mathbf{H}}\overline{\boldsymbol{x}} \right\|^2 + \mu \|\overline{\boldsymbol{x}}\|_0 \tag{6}$$

  where $\|\overline{\boldsymbol{x}}\|_0$ is the number of non-zero components in $\overline{\boldsymbol{x}}$. This is a combinatorial problem, and local exploration is performed by moves affecting only one component. Each iteration either adds or removes one element in the current support, and the replacement is selected which most decreases criterion (6).

---

[1] Notation $\overline{\mathbf{H}}\overline{\boldsymbol{x}}$ is used here for homogeneity, although most works on sparse approximation do not consider convolution-based operators.

- Last, ℓ_1-norm penalization is considered by minimizing

$$\left\| \boldsymbol{y} - \overline{\mathbf{H}} \overline{\boldsymbol{x}} \right\|^2 + \mu \sum_\ell |\overline{x}_\ell|. \qquad (7)$$

Optimization is performed with the homotopy continuation principle described in [10, 13], which shows formal similarities with greedy methods and even more with SBR since it performs removal moves as well. This algorithm will be referred to as ℓ_1-HC.

The reader is referred to the corresponding references for detailed descriptions of the algorithms. Note that such algorithms were compared for deconvolution purposes in [14], but only within the standard convolution setting of Eq. (3).

The selection steps for MP and OMP mostly amount to computing matrix products $\mathbf{H}^T \cdot$. For MISO systems, using notations of Section 2, such a product is decomposed into K products $\mathbf{H}_k^T \cdot$ with Toeplitz matrices $\mathbf{H}_k$. These are actually cross-correlations, which can be implemented in the frequency domain using two FFTs and one inverse FFT (in dimension N) [12]. In practice, the Discrete Fourier Transforms of all $\boldsymbol{h}_k$ are computed before the algorithm starts. One selection step is then executed by $K+1$ FFTs. Hence, the corresponding cost increases linearly with K. For MP and OMP, the update of the residue is identical to the standard versions. For OMP, it requires the inversion of matrix $\mathbf{H}_\star^T \mathbf{H}_\star$, where subscript $\star$ indexes the active columns of $\mathbf{H}$. In the proposed implementation, the Cholesky factorization of $\mathbf{H}_\star^T \mathbf{H}_\star$ is updated at each iteration at low cost, since one iteration only performs rank-one modifications to such matrix. Doing so, system inversions amount to two triangular system inversions of complexity $\mathcal{O}(i^2)$ where i is the number of active elements.

Efficient implementations of OLS and SBR require extensive access to elements of the Gram matrix $\mathbf{H}^T \mathbf{H}$ – more precisely, to $\mathbf{H}_\star^T \mathbf{H}_\circ$, where subscript $_\circ$ indexes the non-active columns of $\mathbf{H}$ at a given iteration, see for example implementation details given in [9]. For MISO systems, $\mathbf{H}^T \mathbf{H}$ is composed of blocks $\mathbf{H}_k^T \mathbf{H}_\ell$, that are Toeplitz matrices with elements corresponding to the cross-correlation between $\boldsymbol{h}^k$ and $\boldsymbol{h}^\ell$. Hence, the pre-computation of the $K(K+1)/2$ distinct cross-correlation sequences – also in the Fourier domain – between the K impulse responses gives all useful information about the Gram matrix.

ℓ_1-HC minimizes criterion (7) by gradually decreasing the value of parameter μ [10, 13]. At iteration j, all possible values of μ producing a change in the sign of the current solution are computed, among which the next value $\mu^{(j)}$ is selected as the maximal one that satisfies $\mu^{(j)} < \mu^{(j-1)}$. Such computations are indeed similar to those of previously described greedy methods (see for example [13] for explicit equations). More precisely, addition tests require the computation of two matrix products $\mathbf{H}^T \cdot$, that is, $2(K+1)$ FFTs. Removal tests amount to two system inversions with matrix $\mathbf{H}_\star^T \mathbf{H}_\star$, performed by Cholesky factorization as previously explained.

All algorithms require the pre-computation of the products $\mathbf{H}_k^T \boldsymbol{y}$, which are performed using FFT. To sum up, the complexity of each iteration of MP, OMP and ℓ_1-HC is proportional to K. On the contrary, OLS and SBR require the pre-computation of $(K+1)^2$ FFTs, but their core computations remain roughly constant as K increases.

4. SIMULATION RESULTS

4.1. Deconvolution of a complex signal

We consider the data shown in Figure 2, generated from Eq. (1) with 8 echoes, randomly distributed on the continuous time axis. Consequently, none of them falls exactly on any restoration grid. The waveform is a 5 MHz sine wave with a Gaussian envelope [4]. The data is sampled at 25 MHz and corrupted by 10 dB SNR Gaussian noise. Three overlapping problems occur at approximately 1, 4.5 and 8.5 μs.

Deconvolution is performed with standard algorithms (*i.e.,* $K = 1$) and using up-sampling with $K = 6$. Greedy methods are stopped when the norm of the approximation error $\left\| \boldsymbol{y} - \mathbf{H} \boldsymbol{x} \right\|^2$ becomes lower than a given threshold, depending on the noise power (see for example [14]). Similarly, the regularization parameter for SBR and ℓ_1-HC is tuned[2] in order to get solutions with similar approximation errors. All the standard methods fail to correctly locate the echoes on any of the three problems and many spike locations and signs are badly estimated. The up-sampled deconvolution leads to more satisfactory results. In particular, the erroneous behavior of all algorithms at 8.5 μs has been corrected for $K = 6$. OLS and SBR with $K = 6$ also solve the overlapping problem at 1 μs. However, they still fail to correctly locate the two close echoes at 4.5 μs, where estimation results are even slightly worse than with $K = 1$. This is due to the suboptimal nature of the greedy algorithms, that reached a local minimum of the data misfit criterion. The ℓ_1-norm deconvolution achieves correct location of the two close echoes at 4.5 μs and at 8.5 μs. On the other hand, it produces spurious small spikes and double spikes, which are typical artifacts of ℓ_1-norm-based sparse approximation.

4.2. Monte-Carlo simulations

We now compare algorithmic performances with Monte-Carlo simulations. The deconvolution algorithms are run for 2000 synthetic data sets, containing 15 echoes, with the same waveform, SNR and total duration as the data used in Figure 2. The signals therefore contain strongly overlapping echoes. Algorithms are tuned as explained in § 4.1.

Since true spikes do not belong to any of the discrete reconstruction grids, estimation errors are computed using a distance between two spike trains inspired by the work

---

[2]Note that the ℓ_1-norm introduces bias on amplitude estimation. Thus, for ℓ_1-HC, amplitudes are corrected before computing the approximation error.
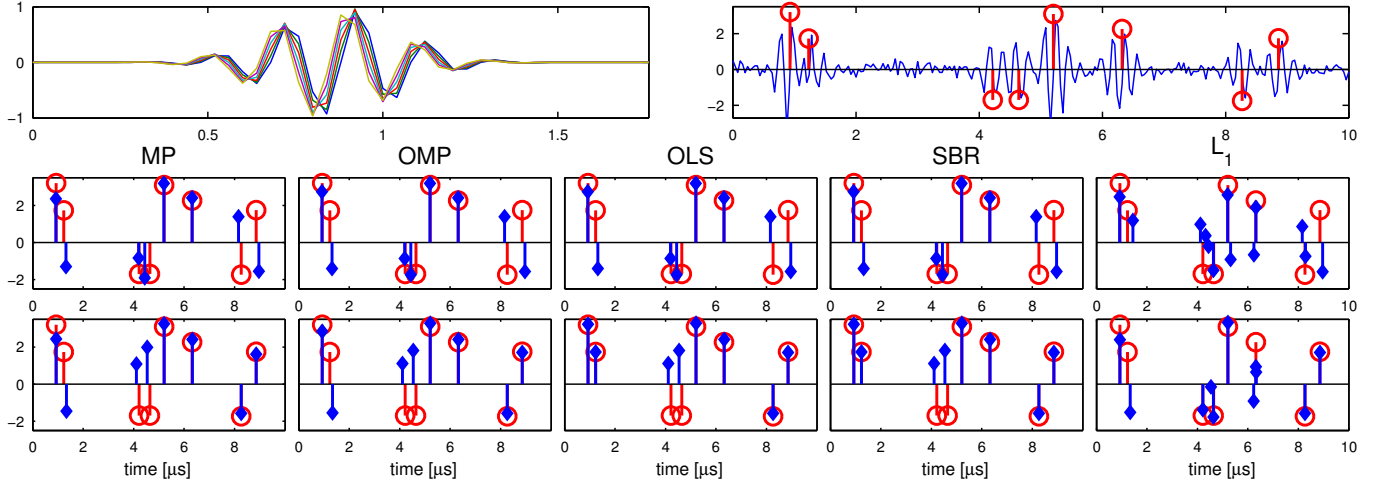
Fig. 2. Deconvolution example from synthetic data. Top left: waveforms used for deconvolution with $K = 6$. Top right: data (–) and true spikes (○). Middle: results with $K = 1$. Bottom: results with $K = 6$. True spikes (○) and estimated spikes (◇).

in [15]: amplitudes are first binarized to ± 1 in order to give the same importance to all detections. Then, the spike trains are convolved with a double-side exponential kernel $e^{-|t|/\tau}$, with $\tau = T_S$, producing slight spike spreading. Finally, the ℓ_2-norm between the two convolved spike trains is computed. Figure 3 (top) shows such estimation errors obtained by the implemented algorithms for different values of the up-sampling factor K. As can be expected, errors decrease with K. For example, using $K = 4$ yields an error reduction of about 25% with respect to $K = 1$, except for ℓ_1-HC. The relative performances of the different algorithms are also in accordance with their complexity, that is, MP has the greater error, followed by OMP, OLS and SBR. The results for ℓ_1-HC appear to be less sensitive to up-sampling, and show the worst performance among all methods for $K \geq 2$. Most of this behavior can be explained by the nature of the spike distance, which strongly penalizes the false detections of small amplitude spikes, inherent to ℓ_1-norm penalization. On the contrary, other simulation showed that ℓ_1-HC yields the smallest errors using a distance without amplitude binarization. Note also that for $K \geq 6$, error reduction becomes negligible. This can be explained by the intrinsic variance on the time delay estimation due to the presence of noise [1].

Central Processing Unit (CPU) times are evaluated with Matlab running on a personal laptop computer with 4 Go RAM and double-Core CPUs clocked at 2.5 GHz. Results are plotted in Figure 3 (bottom). MP and OMP are the fastest and their cost increase linearly with K, which is coherent with the analysis in Section 3. OLS, followed by SBR, are more costly, which is in accordance with their increased complexity. We note that most CPU time required by these algorithms is due to pre-computations, whose cost increase roughly quadratically with K (see Section 3). Note that the cost of ℓ_1-norm deconvolution is the highest one, which is also in accordance with the results in Figures 2 and 3 (top):

ℓ_1-HC estimates show more spikes than other algorithms, hence their computations require more iterations.
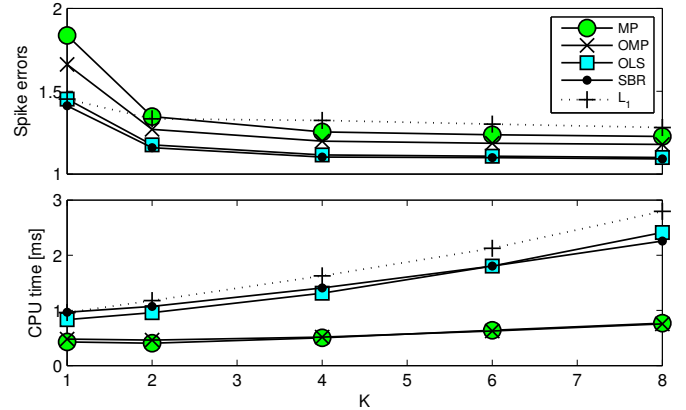

Fig. 3. Estimation errors (top) and computational costs (bottom) for different algorithms versus the up-sampling factor K. Results are averaged on 2000 random realizations.

5. CONCLUSION

An up-sampling approach for sparse deconvolution has been proposed for the estimation of time delays in typical ultrasonic NDT data. A model was introduced based on a finer time discretization of the convolution model than usual approaches. The model was recast has a MISO system, for which well-known sparse approximation methods were studied and computationally optimized for deconvolution. Synthetic simulations revealed the efficiency of up-sampled deconvolution to estimate times of arrival in presence of noise, even for strong overlapping. Computation costs were evaluated that confirmed the implementation efficiency. In particular, increasing the up-sampling factor only produces a reasonable increase of the CPU time.

6. REFERENCES

[1] A. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 527–533, June 1981.

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

[3] A. Bruckstein, Tie-Jun Shan, and T. Kailath, "The resolution of overlapping echoes," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 6, pp. 1357–1367, December 1985.

[4] R. Demirli and J. Saniie, "Model-based estimation of ultrasonic echoes. Part I: Analysis and algorithms," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 48, no. 3, pp. 787–802, May 2001.

[5] J. Idier, *Bayesian Approach to Inverse Problems*, ISTE Ltd and John Wiley & Sons Inc, April 2008.

[6] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[7] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, November 1993, vol. 1, pp. 40–44.

[8] S. Chen, S.A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[9] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli Gaussian Deconvolution to Sparse Signal Restoration," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4572–4584, October 2011.

[10] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, USA, March 2005, vol. 5, pp. 733–736.

[11] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, August 2010.

[12] G.H. Golub and C.F. Van Loan, *Matrix Computations*, vol. 3, Johns Hopkins University Press, Baltimore and London, October 1996.

[13] S. Maria and J.-J. Fuchs, "Application of the global matched filter to stap data: an efficient algorithmic approach," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006, vol. 4.

[14] S. Bourguignon, C. Soussen, H. Carfantan, and J. Idier, "Sparse deconvolution: Comparison of statistical and deterministic approaches," in *IEEE Statistical Signal Processing Workshop*, June 2011, pp. 317–320.

[15] M. C. W. Van Rossum, "A novel spike distance," *Neural Computation*, vol. 13, no. 4, pp. 751–763, April 2001.