SPARSE UNDERWATER ACOUSTIC IMAGING: A CASE STUDY

Nikolaos Stefanakis[†], Jacques Marchal[⋆], Valentin Emiya[◇], Nancy Bertin[○], Rémi Gribonval[†], Pierre Cervenka[⋆]

[†]INRIA
Centre Inria Rennes
Bretagne Atlantique
35042 Rennes Cedex, France

[⋆] UPMC Univ Paris 06
CNRS UMR 7190
Institut Jean le Rond d'Alembert
78210 Saint-Cyr-l'Ecole, France

[◇] Aix-Marseille Univ
CNRS UMR 7279
LIF-QARMA
13013 Marseille, France

[○] CNRS, IRISA – UMR 6074
Campus de Beaulieu
35042 Rennes Cedex, France

ABSTRACT

Underwater acoustic imaging is traditionally performed with beamforming: beams are formed at emission to insonify limited angular regions; beams are (synthetically) formed at reception to form the image. We propose to exploit a natural sparsity prior to perform 3D underwater imaging using a newly built flexible-configuration sonar device. The computational challenges raised by the high-dimensionality of the problem are highlighted, and we describe a strategy to overcome them. As a proof of concept, the proposed approach is used on real data acquired with the new sonar to obtain an image of an underwater target. We discuss the merits of the obtained image in comparison with standard beamforming, as well as the main challenges lying ahead, and the bottlenecks that will need to be solved before sparse methods can be fully exploited in the context of underwater compressed 3D sonar imaging.

Index Terms— Sonar imaging; underwater acoustics; beamforming; orthogonal matching pursuit; sparse reconstruction; compressed sensing.

1. INTRODUCTION

The use of sparsity prior have proven its potential and efficiency, as well as its versatility in a wide range of signal processing contexts. However, scaling up from synthetic experiments to real world applications remains a challenge in many cases. Representing a signal in a redundant dictionary implies handling huge volumes of data, and a discrete model in which the considered analog data fits. This paper addresses this difficulty in the case of underwater acoustic imaging.

In underwater acoustic scenarii, seafloor imaging and bathymetry are usually performed by sidescan sonars or multibeam echosounders [1]. These classical sonar systems insonify an across-track area at each emission burst: 3D images are built by juxtaposing successive strips, obtained one after the other while the ship is moving ahead. Real 3D-imaging systems are not available, especially due to the large amount of data to acquire and to process.

In this context, the use of a sparse modelling of the problem and of compressed sensing techniques [2, 3, 4] opens interesting doors to new devices that could image directly the 3D space. Such sparse models have been proposed for the related radar imaging problem [5], and for over-the-air sonar imaging [6]. However, existing work essentially tested ideas on simple synthetic data. Dealing with real data implies challenges such as: handling a 3D grid with a higher number of points; detecting targets that are not located on the grid points; detecting complex-shape objects rather than a simple

pattern like a square; using non-ideal transducers for emission and reception, with directivity patterns and calibration issues; and handling phase issues due to the propagation and to the modulation by a carrier frequency.

In this paper, we perform underwater acoustics imaging with a new flexible-configuration sonar device that we built and used to record real data. It leads to a problem statement, described in Section 2, that includes the formulation of the theoretical estimation problem as well as the issues raised by the actual orders of magnitudes of the considered dimensions. In Section 3, we describe the standard beamforming approach as a linear estimator and introduce an alternate linear estimator where emission beams are replaced by elementary emission patterns. In Section 4, we then propose a non-linear estimator based on a sparse model and on a modified version of the Orthogonal Matching Pursuit (OMP) [7] algorithm, in which algorithmic enhancements makes the estimation tractable. As a proof of concept, the proposed approach provides an image of the target comparable to the basic beamforming technology, as shown in Section 5. Though not yet competing with refined beamforming-based multibeam echosounders, the results show the potential of our approach. We discuss this first applicative and algorithmic milestone towards new compressive-sampling-inspired 3D-imaging sonar systems and draw perspectives in Section 6.

2. PROBLEM STATEMENT

2.1. Specifications of the new sonar prototype

The proposed sonar prototype is made of two 26 cm long linear arrays of 64 transducers each. The antennas work at a carrier frequency $F_e = 480\ kHz$ and with a bandwidth $F_B = 160\ kHz$. Each elementary transducer can operate either as a receiver or as an emitter; nevertheless, it is not possible to interchange the transducers utility during the imaging process. A computer drives both the emission of $N_E = 64$ amplified channels, and the reception of $N_R = 64$ others channels. The system can emit several (N_P) successive transmission sequences called *pings*. At each ping, the N_E transmitted signals can be different from each other in order to form an *emission beam* or in order to achieve a complex radiation pattern.

These specifications allow to carry out experiments in a tank or in the sea with a range up to a few tens of meters. Due to its flexibility, sidescan sonar, multibeam systems or original configurations (such as acoustical camera) can be simulated.

2.2. Description of the experiment

The underwater imaging experiments take place in a tank of dimensions $2.9\ m \times 0.9\ m \times 0.5\ m$ filled with fresh water. The sound speed c is constant and equal to 1484 m/s; thus, the wavelength for
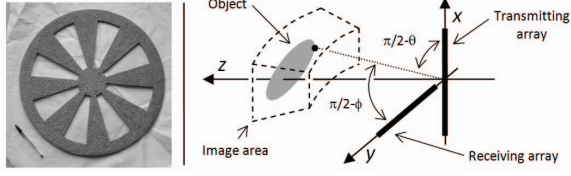
Fig. 1. Picture of the target (left) and its relative position to the coordinate system and to the antenna (right).

$F_e = 480\,kHz$ is $\lambda \approx 3\,mm$. The target to be imaged is a round wheel with outer diameter of $52\,cm$ made of plywood. The wheel is partially covered with sand and is lying on the surface of the water in the tank. For the experiment performed in this paper, the transmitting array is placed along the x-axis and the receiving array along the y-axis (Fig 1). The elementary signal that drives the emitters is a frequency-modulated window of short duration $T = 20\,\mu s$ with 10 periods of the carrier frequency F_e:

$$e(t) = e^{(j2\pi F_e t)} e^{(-\pi^2 F_B^2 t^2/2)} \mathrm{rect}(t/T). \tag{1}$$

2.3. Direct model formulation

The 3D acoustic scene to be sensed is discretized into M small volumes called *voxels* with centers at locations $\mathbf{p}_k$, $k = 1, ..., M$. The N_E emitters are indexed by i and the N_R receivers are indexed by j. Positions of the emitters (resp. receivers) are denoted by vectors $\mathbf{p}_i$ (resp. $\mathbf{p}_j$). The signal emitted by transducer i at ping p is denoted $\mathbf{e}_{i,p}(t)$. The signal recorded at the receiver j at ping p is given by

$$m_{j,p}(t) = \sum_{k=1}^{N_k} b_k \sum_{i=1}^{N_E} e_{ip}(t - \tau_{ik} - \tau_{kj}) \tag{2}$$

where $\tau_{ik} \triangleq \frac{\|\mathbf{p}_k - \mathbf{p}_i\|_2}{c}$ and $\tau_{kj} \triangleq \frac{\|\mathbf{p}_k - \mathbf{p}_j\|_2}{c}$ are the forward and backward propagation delays from the elements of the two antennas to the k-th voxel and b_k is the *backscattering coefficient* of the voxel. Note that only the times of flight are taken into account into the model, and not the spreading attenuation term. The signal $m_{j,p}(t)$ recorded at the jth receiver at ping p is eventually sampled as $m_{j,p}[n] = m_{j,p}(nT_s)$, with $F_s = 1/T_s = 2\,MHz$ the sampling frequency.

2.4. Inverse problem formulation

The 3D acoustic imaging problem is the estimation of the relative energy backscattered from each voxel k, given the $N_P \times N_R$ recorded signals: in other words, given the discrete time-series $m_{j,p}[n]$ measured at all receivers for all pings, we wish to estimate the backscattering coefficients b_k. This is an inverse problem that can be expressed in more compact form:

$$\mathbf{m} = \Phi\mathbf{b}, \tag{3}$$

using the following conventions. Discrete time column vectors of length N_T associated to continuous time signals $x(t)$ are represented by small bold letters $\boldsymbol{x}$. For a given delay τ we denote $\boldsymbol{x}^\tau[n] \triangleq x(nT_s - \tau)$. The dot product is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle \triangleq \mathbf{x}^H.\mathbf{y}$ with $(\cdot)^H$ the Hermitian adjoint. The following compact notations are

used, with $(\cdot)^T$ the transpose:

$$\mathbf{m}_j^T \triangleq \left[\mathbf{m}_{j,1}^T, \ldots, \mathbf{m}_{j,N_P}^T\right], \quad \mathbf{m}^T \triangleq \left[\mathbf{m}_1^T, \ldots, \mathbf{m}_{N_R}^T\right], \tag{4}$$

$$\mathbf{b}^T \triangleq \left[\mathbf{b}_1^T, \ldots, \mathbf{b}_M^T\right]. \tag{5}$$

$$\mathbf{a}_{j,p,k} \triangleq \sum_{i=1}^{N_E} \mathbf{e}_{i,p}^{\tau_{ikj}}, \quad \tau_{ikj} \triangleq \tau_{ik} + \tau_{kj} \tag{6}$$

$$\mathbf{a}_{j,k}^T \triangleq \left[\mathbf{a}_{j,1,k}^T, \ldots, \mathbf{a}_{j,N_P,k}^T\right], \tag{7}$$

$$\mathbf{a}_k^T \triangleq \left[\mathbf{a}_{1,k}^T, \ldots, \mathbf{a}_{N_R,k}^T\right], \tag{8}$$

$$\Phi \triangleq \left[\mathbf{a}_1, \ldots, \mathbf{a}_M\right]. \tag{9}$$

In other words, $\mathbf{m} \in \mathbb{C}^{N_P N_R N_T}$ is the full measurement, $\mathbf{b} \in \mathbb{C}^M$ is the vector with the unknown backscattering coefficients and $\Phi \in \mathbb{C}^{N_P N_R N_T \times M}$ is made with delayed versions of the emitted signals.

3. STANDARD BEAMFORMING APPROACH

Beamforming (BF) is a standard imaging approach to estimate the backscattering coefficients $\mathbf{b}$ as a linear combination of the measurements, for specific configurations of the emission [1].

3.1. Emission

In the classical approach, at each ping p, all emitters are simultaneously active. Introducing small delays between them allow to concentrate the emitted energy in a specific 2D-slice by forming a *beam*. Illuminating only a region of interest shall improve the signal-to-noise ratio (SNR). Moreover, it reduces the computational cost, by dividing the 3D imaging problem into several independent, smaller 2D subproblems. However, by nature, this strategy does not take into account across-slice structure.

We target a more flexible system, able to deal with 3D and to adopt various configurations. At each ping, we choose to have only one active emittor (*mono-emission*), which insonifies then the whole half-space below the ship. Thanks to linearity, such a configuration can be used to simulate the previous one, but also a wider variety of schemes. As the insonification is global, information about all the scene to image is carried at each ping, opening to a better scalability and versatility in the number of pings and acquisition time, even though at the price of a possible lower SNR.

3.2. Imaging

In the classical beamforming approach, images are formed by scanning the space. Reception beams, perpendicular to the emission beams, are synthetically formed for each ping. The imaging process basically amounts to a linear estimate, similar to matched filtering $\hat{\mathbf{b}} \triangleq \Phi^H \mathbf{m}$. One of the main drawbacks of this method is the appearance, in the reconstructed image, of energetic artifacts around bright points (*e.g.* points of the wheel where a specular reflexion occur). These ghost reflectors are due to the *sidelobes* in the antenna directivity, or, in our formalism, as due to the non-zero off-diagonal terms in the matrix $\Phi^H \Phi$. Any linear estimator of $\mathbf{b}$ is likely to suffer from these sidelobes.

Classical beamforming neglects an important feature of the scene to image: its sparsity in the 3D space. We will show in section 4 how this sparsity can be taken into account, and used to build an image from the mono-emission sonar measurements. Being able to deal with 3D-insonification and a sparse model will then open many doors to new acquisition strategies.

4. OMP-LIKE RECONSTRUCTION ALGORITHM

We introduce our approach by noticing that the coefficients $\mathbf{b}_k$ are expected to be zero for most of the voxels (voxel containing only water). Exploiting this sparse nature, we can formulate a non-linear estimator from known greedy algorithms that have been developed for generic sparse recovery problems, such as Orthogonal Matching Pursuit [7] (OMP, see Algorithm 1). We propose algorithmic enhancements for the configuration where only one emittor is active at a time ($\mathbf{e}_{ip} \triangleq \delta_{ip}\mathbf{e}$).

4.1. Discretization and dimensionality issues

Discretizing the full tank with a step of λ would results in a $M = 48.10^6$-point grid, and, in time, the measurement length is about $N_P N_R N_T = 13.10^6$ samples. This makes the problem obviously intractable. Before detailing an algorithmic proposal, we alleviate this dimensionality burden in two ways.

We focus on the region of interest defined by $0.6\,m \leq r \leq 1.2\,m$, $-1.4° \leq \theta \leq 6.4°$, $-4.7 \leq \phi \leq 34.4°$. This area is depicted in the right part of Fig. 1 using dashed lines. A tradeoff between resolution and complexity must be found. We have chosen a discretization using 61 uniform radial steps ($\Delta r \approx 9.9\,mm$), 16 uniform angular steps along θ ($\Delta\theta \approx 0.52°$), and 72 uniform angular steps along ϕ ($\Delta\phi \approx 0.54°$). In the region of interest, each voxel is approximately a cube with edge length of $1\,cm \approx 3\lambda$. As shown below, scanning this region remains a large-size problem.

Signals can be downsampled at $F_s' = F_s/5$ to reduce their size according to their bandwidth. Then, they are cropped according to the radial limitation of the region of interest. As a result, the downsampled cropped measurements are $N_T = 324$-sample series for each of the $N_R = 64$ receivers and of the $N_P = 64$ pings.

The problem formulation (3) is now realistic, with $N_P N_R N_T = 1,327,104$ measures, and $M = 61 \times 16 \times 72 = 70,272$ voxels.

4.2. Reconstruction algorithm

Assuming that $\mathbf{b}$ is sparse, it can be estimated by many sparsity-based methods. Here, we propose to use an approach based on the OMP algorithm, which has a relatively affordable computational cost to face a high-size problem. Algorithm 1 details a naive implementation of OMP. Several days of computations and a hundred gigabytes of memory would be necessary to run such a program. The memory issue comes from the size of the dictionary. As a result, each atom must be built when needed and then forgotten. At each iteration, two stages are computationally demanding: the atom selection and the sparse representation update. The complexity of the latter is cubic and can be reduced – as proposed in [8] – to a quadratic complexity by using a block matrix inversion of $\Phi_\Omega^* \Phi_\Omega$, where the results from the previous iteration is used. The complexity is approximately reduced by a factor $K_{\max}$ (about one thousand in our experiments). Reducing the complexity of the atom selection stage remains the bottleneck addressed in the next section.

4.3. Fast atom selection

The cost for reconstructing all atoms and the inner products with the residual at each iteration is about $O(MN_T N_R N_P)$, where M is the total number of atoms. We propose two strategies to reduce this cost. We will denote by $\bar\tau_{ikj} \triangleq [\tau_{ikj}/F_s'] F_s'$ the approximation of $\bar\tau_{ikj}$ using the sampling frequency F_s'. We denote the part of $\mathbf{r}$ related to ping p and receiver j by $\mathbf{r}_{p,j}$.

Algorithm 1 OMP

> *Initialisation:* $\mathbf{r} \leftarrow \mathbf{m}$; $\Omega \leftarrow \emptyset$.
> **for** $K = 1$ to $K_{\max}$ **do**
>   *Atom selection:* $\widehat{k} \leftarrow \arg\max_k |\langle \mathbf{a}_k, \mathbf{r} \rangle|$
> $\Omega \leftarrow \Omega \cup \left\{\widehat{k}\right\}$
>   *Sparse representation update:* $\widehat{\mathbf{b}}_\Omega \leftarrow \Phi_\Omega^+ \mathbf{m}$
>   *Residue update:* $\mathbf{r} \leftarrow \mathbf{m} - \Phi_\Omega \widehat{\mathbf{b}}_\Omega$
> **end for**
> *Output:* $\widehat{\mathbf{b}}^{\mathrm{OMP}} \leftarrow \widehat{\mathbf{b}}_\Omega$.

Atom selection based on FFT-filtering: the first strategy relies on the structure of the atoms, which are only formed by delayed versions of the elementary emission pattern $\mathbf{e}$ (see eq. (6)-(8)). The strategy is close to [9]:

1. first, compute the auxiliary result $\mathbf{e} * \mathbf{r}$ using an FFT filtering technique, the resulting complexity being $O(N_T \log N_e)$, where $N_e \ll N_T$ is the length of $\mathbf{e}$ when sampled at the same sampling frequency as $\mathbf{r}$;

2. then, approximate τ_{ikj} by $\bar\tau_{ikj}$ and according to eq. (6), use $\langle \mathbf{a}_k, \mathbf{r} \rangle \approx \sum_{j=1}^{N_R} \sum_{i=1}^{N_E} [\mathbf{e} * \mathbf{r}_{p=i,j}](\bar\tau_{ikj}) e^{2j\pi F_e(\tau_{ikj} - \bar\tau_{ikj})}$, with a complexity of $O(MN_R N_E)$.

Approximate atom selection: in the second strategy, we approximate the contribution of $\mathbf{r}_{p,j}$ to the inner product by the single, most significant value $\mathbf{r}_{p=i,j}(\bar\tau_{ikj})$ so that $\langle \mathbf{a}_k, \mathbf{r} \rangle$ is approximated by

$$\sum_{j=1}^{N_R} \sum_{i=1}^{N_E} \mathbf{r}_{p=i,j}(\bar\tau_{ikj}) e^{2i\pi F_e(\tau_{ikj} - \bar\tau_{ikj})}, \qquad (10)$$

the complexity being $O(MN_E N_R)$.

5. IMPLEMENTATION AND RESULTS

A significant variation in the level of the recorded signals was observed, which revealed a calibration problem at the arrays. In particular, the acquisition gains of some transducers were below -10 dB compared to an estimation of the average acquisition gain. In order to circumvent this issue, 12 receivers were completely discarded when applying the sparse reconstruction approach. Nevertheless, these transducers were not removed during the classical imaging approach; amplitude mismatch doesn't adversely affect the classical approach, as long as there is no phase mismatch.

Both OMP algorithms were performed for $K_{max} = 1200$ iterations. The computation time was about 8 hours for both selection tactics. Each algorithm provided a residual $\mathbf{r}$ and a sparse set of coefficients $\mathbf{b}_k, k \in \Omega$. An image was built in a hybrid way representing both the sparse coefficient values and the magnitude of the residual projected on the non-selected atoms:

$$I_k \triangleq \begin{cases} |\mathbf{b}|_k, & k \in \Omega \\ |\langle \mathbf{a}_k, \mathbf{r} \rangle|, & k \in \Omega^c. \end{cases} \qquad (11)$$

The resulting figures are shown in grey scale in Fig. 2. Focusing on column (a) of Fig. 2, it can be observed that there are regions with very large amplitudes at $x = 0.6$ and $x = 1.1\,m$. This is due to specular echos, which result to the appearance of side lobes, *i.e.* spatial spreading along ϕ and θ. This can lead to misintepretation of the object shape. While the main shape of the scatter object is more or less apparent with all approaches, it can be seen that the
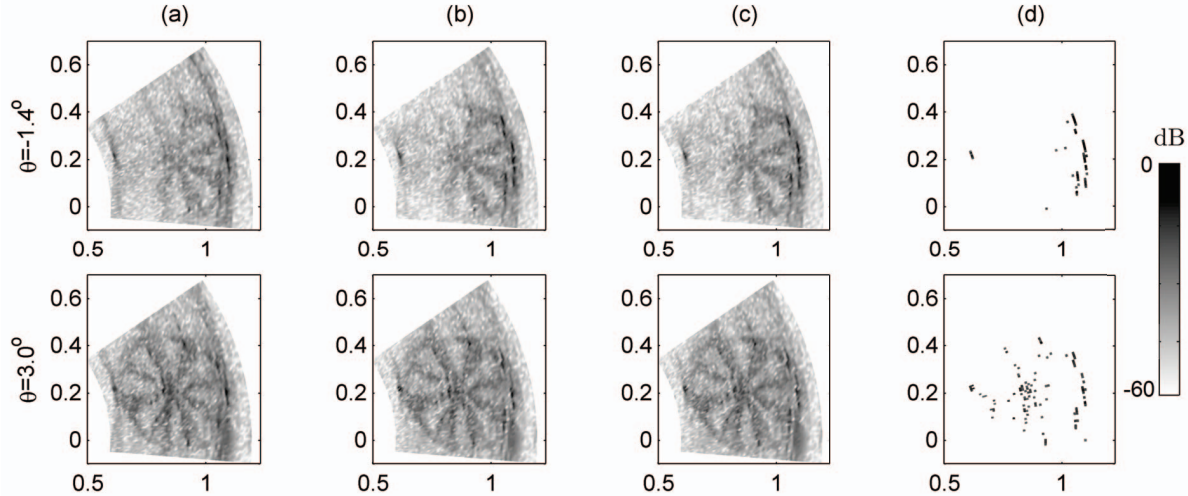
Fig. 2. Images obtained by classical imaging (a), by OMP with approximate selection (b), by OMP with FFT based selection step (c). The selected voxels, as obtained by the last method, are shown with respective amplitudes (d). Upper row corresponds to the slice $\theta = -1.4^o$ and lower row to slice at $\theta = 3^o$. All figures are displayed at the same dynamic range and the colour scale is normalized with respect to the maximum voxel amplitude. The *x* axis corresponds to range and the *y* axis to the y coordinate

side lobes in the sparse formulation are significantly reduced, and the image quality is thus improved.

To be noticed is the fact that after 1200 iterations, the ratio between the initial energy of the measurement to the energy of the residual, calculated as $20 \log_{10} \left\| \frac{\mathbf{m}}{\mathbf{r}} \right\|_2$, was only +0.6 dB for OMP algorithms. This rather small convergence is believed to be a consequence of the calibration issues mentioned earlier and possibly, of modeling errors concerning the relevant position of the two antennas or the calculation of the sound speed. Given a better modeling or estimation of these parameters, it is believed that the sparse formulation will further enhance the imaging quality compared to the standard approach.

6. CONCLUSION AND PERSPECTIVES

In this paper, we have applied sparse reconstruction to an underwater acoustic imaging problem. We highlighted the very large amount of data it implies to deal with, the issues it raises, and showed that this burden can be overcome in our case by applying an optimized version of OMP algorithm. Although under conditions of poor signal-to-noise ratio and with serious calibration issues concerning the receiving antenna, the sparse reconstruction approach proved to be a viable option to produce images of the target, and even to bring some improvement in certain aspects of the image quality, compared to standard imaging. Though perfectible, the sparse approach shows here its potential on real world data.

Besides possible refinements of the sparse model itself and improvement of the computation speed, some other deeper changes must be considered. One of the major interest of modelling the imaging problem as a sparse inverse problem, is the possibility to exploit the compressive sampling framework. Some major changes in the acquisition schemes can be performed: for instance, introduction of randomness in the emitted sequences, and different distributions of the transducers utilities (for instance, placing receivers and emitters on both antennas, with irregular distribution, which is allowed by our flexible sonar). This framework lets foresee dramatic improvements in terms of acquisition time and data flow, which is a major issue in underwater acoustics.

7. REFERENCES

[1] X. Lurton, *An introduction to underwater acoustics : principles and applications*, Springer, 2002.

[2] E. J. Candès, J. Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, vol. 59, pp. 1207–1223, 2006.

[3] David L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] Richard Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, July 2007.

[5] M.A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[6] P. Boufounos, "Compressive sensing for over-the-air ultrasound," in *Proc. of ICASSP*, Prague, Czech Republic, May 2011, pp. 5972–5975.

[7] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27 th Ann. Asilomar Conf. Signals, Systems, Computers*, 1993, pp. 40–44.

[8] T. Blumensath and M.E. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370 – 2382, june 2008.

[9] S. Krstulovic and R. Gribonval, "MPTK: Matching pursuit made tractable," in *Proc. of ICASSP*, May 2006.