A REVIEW OF THE COMPRESSIVE SAMPLING FRAMEWORK IN THE LIGHTS OF SPHERICAL HARMONICS: APPLICATIONS TO DISTRIBUTED SPHERICAL ARRAYS

*Bruno Masiero** and Martin Pollow*

Institute of Technical Acoustics
RWTH Aachen University
Neustraße 50, 52066 Aachen, Germany
bma@akustik.rwth-aachen.de

ABSTRACT

"Compressive Sampling" proposes a new framework on how to effectively sample information (signals or any other physical phenomena) with a reduced number of sensors. The main idea behind this concept is that if the information to be sampled can be sparsely described in a space that is incoherent to the measurement space, then this information can be restored by ℓ_1 minimization. In this paper we describe the Compressive Sampling framework and present one example of application, namely, to sample an outgoing acoustic field with a distributed spherical array composed of a reduced number of sensing microphones without suffering from aliasing errors.

1. INTRODUCTION

"Compressive Sampling", also called "Compressed Sensing" or simply CS, postulates that the widely accepted Shannon's sampling theorem is a sufficient but not a necessary condition for the sampling theory, i.e., CS states that if a signal of interest can be described by a sparse set of basis, then it can be sampled at a rate lower than twice its maximum frequency (sub-Nyquist-rate). The ability of CS to recover data from fewer measurement points than the traditional sampling techniques relies on two assumptions: the sparseness of the signal of interest when described by a basis Ψ and the incoherence between the measurement basis Φ and the representation basis Ψ. Both assumptions will be discussed later in this paper.

The CS framework relies on linear programming algorithms to recover a sparse representation of the signal. This is achieved by minimizing the ℓ_1 norm of the observation vector instead of the in acoustics more commonly used ℓ_2 norm. The ℓ_1 norm has been sporadically used in the field of acoustics, especially when dealing with source separation and reflection/echo estimation and cancellation. In a recent work from Hörchens and de Vries a spherical microphone array was used to identify the direction and time of arrival of wall reflection in a concert hall [1]. Even though they did not explicitly relate their work to the CS theory, it is possible to directly relate one of the strategies presented in that paper to CS, since they seek with ℓ_1 minimization a set of sparse coefficients in a basis (plane wave basis) incoherent to the original measurement basis (spherical array basis).

As far as the authors are concerned, the first two works to explicitly apply the CS framework to spatial audio were published independently in October 2009. In their work Epain, Jin and van Schaik shortly presents the concepts of CS and describes the application of CS to spatial sound field analysis (with a compact microphone array) and synthesis (with a loudspeaker ring) [2]. Lilis, Angelosante and Giannakis published their results later that same month [3]. They applied the concept of CS to Wave Field Synthesis (WFS) reproduction, achieving an improved performance when not every loudspeaker of the array is simultaneously active. It is interesting to emphasize their statement that a recursive version of the ℓ_1 minimization can be implemented in real time.

In this paper we first provide a review of the compressive sampling framework. We then present an example of CS applied to spatial audio, dual to the results presented in [2]. We consider a distributed microphone array used to measure the radiation pattern of a source located at the center of this array.

2. COMPRESSIVE SAMPLING

The first works describing Compressive Sampling were devoted to image reconstruction and later came applications to reconstruction of signals from noisy measurements. Candès and Wakin give a very didactic example of the idea behind CS: in digital image acquisition, a huge amount of data is collected to later be thrown away by compression algorithms, so to allow easy storage and facilitate file exchange. It seems contradictory that consumers lust for cameras with high pixel resolution only to later compress the image to a hundredth of its original size in order to efficiently store these photos. What happens in this example is that a huge amount of data (pixels) acquired by the camera is transformed into another domain, e.g. the DCT, in which a small percentage of the transform coefficients containing the greater part of the total energy are kept and all other coefficients are discarded. Compressive Sampling tries to skip the steps of acquiring a great amount of pixels and then compressing them by trying to directly sample only the relevant (highly energetic) information about the image (or signal) of interest. In a nutshell, CS directly converts analog data into an already compressed digital form, so to obtain very compact signals from a reduced number of sensors. After the acquisition process, one has only to accordingly "decompress" the compact data to restore the originally measured data [4].

Another possible application of CS is to inverse problems, like the problems commonly dealt with in spherical acoustics. Consider, e.g., one wishes to estimate the radiation pattern of a sound source from M sensing points. This is usually done by calculating the spherical harmonics (SH) coefficients that deliver

*Scholarship holder from CNPq - Brazil.

the best approximation at the sensing points and interpolate the remaining surface from these coefficients.

However, when the number of sensing positions is smaller than the number of SH basis necessary to describe this radiation pattern, we have infinite many surfaces that match the measured values and we cannot be sure which of these surfaces we just sampled. Nevertheless, if the basis for the sensing domain Φ (e.g. sampled sphere) and the projection domain Ψ (e.g. spherical harmonics) are incoherent, then an efficient sensing is possible; meaning that the exact sound field could be reconstructed from a reduced number of microphones, as long as this field can be sparsely described in the given representation basis. The CS framework senses the signal at a low sampling rate, without trying to completely comprehend the signal of interest, and later relies on computational power to recover the signal from an apparently insufficient set of measurements [4].

To allow nearly perfect signal reconstruction from a reduced number of sensors CS heavily relies in two principles, highlighted in the previous paragraph: *sparseness* (a characteristic of the signal of interest) and *incoherence* (a characteristic of the sampling strategy).

2.1. Sparseness

A vector of length N is said to be *sparse* if just $S < N$ of its elements are non-zero. If the remaining elements are not zero but their energy is considerably lower than that of the first S more energetic coefficients, then the vector is called *compressible*. The concept of sparseness is important in the CS framework since CS is only applicable to signals that can be concisely (sparsely) described in the representation basis Ψ.

We give two examples of sparse/compressible signals in the spatial acoustic context. In [2] the authors state that often a sound field is generated by one or a few sound sources. If this sound fields is represented in a "source" or "plane wave" domain, then this representation is sparse. Another example would be the directivity of an omnidirectional dodecahedron loudspeaker when represented in the spherical harmonics domain. In low frequencies it radiates omnidirectionally, being described by only one SH coefficient. At higher frequencies it no longer radiates omnidirectionally, but its radiation pattern can still be described by a few SH coefficients. For instance if we measure its radiation at 20 kHz, a frequency region where its radiation is nowhere near omnidirectional, and calculate the SH coefficients up to order 40 we verify that 95% of the radiated energy is concentrated on 106 out of 1681 coefficients, thus a compressible representation.

2.2. Incoherence

We are accustomed to the concept of time and frequency domains and it is common knowledge that an impulse in time domain has a wide-band representation in frequency domain. If the description of one event is local (i.e. sparse) in one domain while spread (i.e. dense) in the other domain, then these two domains can be called incoherent. Coherence between two bases is defined as

$$\mu(\Psi, \Phi) = \sqrt{n} \cdot \max_{k,j} |\langle \phi_k, \psi_j \rangle| . \quad (1)$$

Thus, one can say that time and frequency domains are incoherent as their basis pair present $\mu(\Psi, \Phi) = 1$, yielding minimum coherence (or maximum incoherence).

For applications such as image processing, wavelets and noiselets bases are commonly used to provide incoherent sampling. In this paper we will restrain ourselves to the traditional basis of spherical harmonics and Dirac impulse distribution on the sphere, even though we are aware research has already been done with wavelet like spherical basis [5].

3. MINIMIZATION

We strive to estimate a vector of length N' out of $M < N'$ measurements, leading to an under-determined system of equations. As well known, such systems have an infinite number of solutions. Taking the solution with the minimal ℓ_2 norm provides the minimum energy solution and can be easily calculated. On the other hand, when using incoherent basis a sparse solution is expected and a norm that promotes sparsity should be preferred.

Minimization of the ℓ_0 norm would be the obvious choice when a sparse response is expected. But to solve a ℓ_0 minimization is at the moment still a very time consuming operation, based on iterative algorithms or greedy heuristic. Meanwhile, ℓ_1 minimization can be efficiently calculated using linear programming methods and also delivers sparse results (ℓ_1 minimization has been in use for almost 40 years as a sparsity promoting norm). Even though ℓ_1 is a softer requirement for obtaining sparse solutions, Candès and Wakin argues that if the vector is sufficiently sparse, recovery via ℓ_1 minimization is provably exact [4].

4. RANDOM SENSING

In [4] the authors discuss the "restricted isometry property" (RIP). Briefly speaking, matrices that obey the RIP preserve, at a certain degree, the length of a multiplied sparse vector; implying that such vectors cannot be in its null space.

CS theory argues that an ideal sensing matrix should obey the RIP. In [4] some matrices that obey RIP are discussed and a special focus is placed at randomness. The main property of matrices that obey the RIP condition is that, regardless of having linearly dependent columns, subsets from these columns are nearly orthogonal. Another way to generate matrices that obey the RIP is to take an orthogonal matrix and randomly select M of its rows. In the spherical harmonic case this is equivalent to randomly selecting M sampling points on the sphere. A matrix with this format will obey the RIP with elevated probability as long as

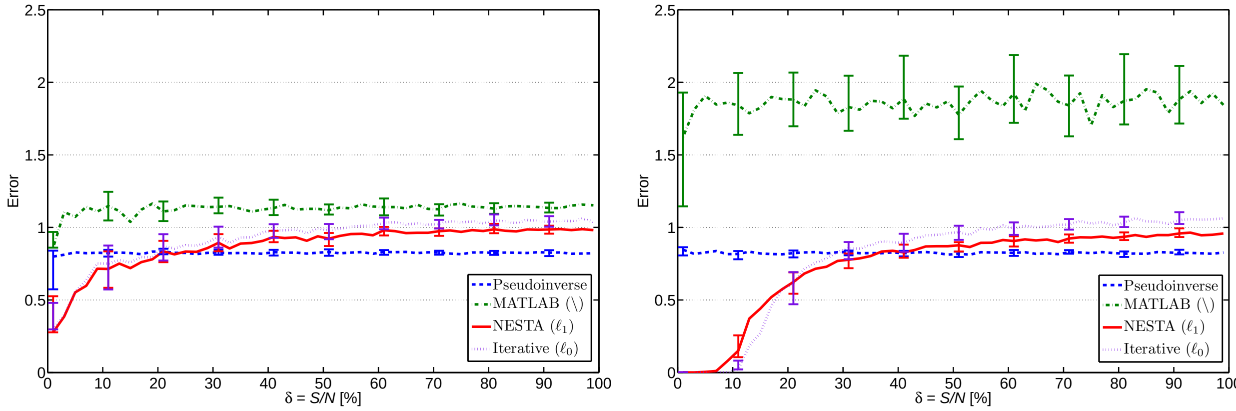
$$M \geq C \cdot S \log\left(N'/S\right), \quad (2)$$

where C is a constant dependent on the matrix (but usually small).

Candès and Wakin conclude that "randomized matrices together with ℓ_1 minimization is a near-optimal sensing strategy" [4]. This is the framework we will apply in the context of spherical harmonics in the following sections.

5. COMPRESSIVE SAMPLING IN THE SPHERICAL DOMAIN

In the next section we consider a sound source placed in the center of a sphere. The sound field radiated by the source in the farfield can be described by the spherical harmonics coefficients $\hat{\boldsymbol{p}}$ of length N. The sound pressure generated at the sampling



(a) (b)

Figure 1: Variation of the estimation error with the sparseness level δ averaged over 50 trials. $\hat{\boldsymbol{p}}$ had its length fixed to $N = 100$. There were $S = 32$ sampling points and no additive noise. (a) Regularly distributed array. (b) Randomly distributed array.

points is

$$\boldsymbol{p} = Y\hat{\boldsymbol{p}}, \tag{3}$$

where $\boldsymbol{Y}$ is the orthonormal spherical harmonic basis. Our interest is to estimate the spherical harmonics coefficients $\hat{\boldsymbol{p}}$ from the measurement vector $\boldsymbol{p}$. This estimation is usually performed as

$$\hat{\boldsymbol{p}}' = \boldsymbol{Y}^{\dagger}\boldsymbol{p}, \tag{4}$$

where $\boldsymbol{Y}^{\dagger}$ is the pseudoinverse of $\boldsymbol{Y}$ and $\hat{\boldsymbol{p}}'$ is the approximation of $\hat{\boldsymbol{p}}$ in the ℓ_2 sense. We can, instead, use a convex optimization program to minimize the ℓ_1-norm[1] of $\hat{\boldsymbol{p}}'$:

$$\min_{\hat{\boldsymbol{p}}' \in \mathbb{R}^N} \left\|\hat{\boldsymbol{p}}'\right\|_1 \text{ subject to } \left\|\boldsymbol{Y}\hat{\boldsymbol{p}}' - \boldsymbol{p}\right\|_2 \leq \epsilon. \tag{5}$$

ϵ is a small constant proportional to the measurement noise.

The normalized error between original and the estimated values is defined as

$$e \triangleq \frac{\|\hat{\boldsymbol{p}}' - \hat{\boldsymbol{p}}\|_2}{\|\hat{\boldsymbol{p}}\|_2}. \tag{6}$$

6. RESULTS

Epain *et al.* presented an example of a compact microphone array for incoming waves using the ideas from CS [2]. Here we would like to tackle its dual problem, i.e., a distributed microphone array to sense outgoing waves. In the following example we consider a spherical microphone array with the sound source positioned exactly at its center[2]. We wish to estimate the source's radiation pattern up to higher orders from a reduced number of sample points, using the CS framework to avoid aliasing errors.

When the pseudoinverse is used to evaluate measurements from an array with a reduced number of microphones, e.g. 32 units, we are restricted to work with a relatively small SH order. Usually one would calculate only up to order $4 = \left(\lfloor\sqrt{32}\rfloor - 1\right)$. If we attempt to calculate $\hat{\boldsymbol{p}}'$ at higher SH orders we will suffer from aliasing and energy will migrate to the higher order coefficients [7], resulting in an urchin-like plot, since this is the least energetic surface that contains the measured points. When, instead of minimizing the ℓ_2-norm, we minimize Eq. 5, we can estimate up to higher orders without the effect of aliasing, again, only if the measured surface have a sparse representation in the used representation domain (in this case the spherical harmonics domain). If instead of a regular grid a randomly distributed array is used, an even better approximation is achieved.

To better illustrate the affirmations made in the last paragraph we present here a numeric example. We first analyze the influence of the vector's sparseness to the estimation error. We define $N' = N = 100$ (SH order 9) and we sample at $M = 32$ points. For each value of S the index and (complex) value of the S non-zero coefficients is randomly chosen. We calculate $\hat{\boldsymbol{p}}'$ using four different algorithms: (1) pseudoinverse, (2) MATLAB operator "\", which delivers a rank limited ℓ_2 minimization, (3) ℓ_1 minimization using the NESTA toolbox [8] and (4) reweighted ℓ_1 minimization, which approximates the ℓ_0 solution.

Fig. 1 displays to each value of S the mean value with standard variation from 50 trials. We verify that ℓ_1 and ℓ_0 minimization deliver considerably better results than the ℓ_2 minimization when $\hat{\boldsymbol{p}}$ is sufficiently sparse. CS allows the estimation of a vector of size $N = 100$ out of 32 sample points without suffering from aliasing if $S < 10$. Also important to note is that a regular distribution of the sample points, as presented in Fig. 1(a), has a performance significantly worse than when the points are randomly distribute, as seen in Fig. 1(b).

The results present in Fig. 1 were simulated in a noiseless fashion. The next question that arises is if CS is robust to noise. We now repeat the simulation fixing $S = 9$ and varying the signal-to-noise ratio (SNR). Sample points were randomly distributed. The results presented in Fig. 2 shows that the performance of all tested minimization algorithms is linearly proportional to the SNR. The performance of the ℓ_1 and ℓ_0 minimization for low SNR values saturates at the chosen value of ϵ.

So far the simulations were performed for a fixed value of N'. But if we do not know the size of the support of $\hat{\boldsymbol{p}}$, than we cannot a priori decide the value of N'. The same simulation is repeated once again, with the length of $\hat{\boldsymbol{p}}$ fixed to $N = 49$ and $S = 9$, $M = 32$ and SNR = 40 dB. Fig. 3 shows how the error varies with the length of $\hat{\boldsymbol{p}}'$ (given in SH order = $\sqrt{N'} - 1$). We see that for $N' < N$ all algorithms perform identically, as this is still an over-determined system and all algorithms minimize $\|\boldsymbol{Y}\hat{\boldsymbol{p}}' - \boldsymbol{p}\|_2$. At $N' = 6$ the ℓ_2 minimization algorithms reach a minimum and for values of $N' > N$ the estimation error increases, caused by the aliasing effect already previously

[1] $\|\mathrm{x}\|_1 \triangleq \sum_i |x_i|$

[2] Sources out of center will not be dealt with here. For more details about this topic, please refer to [6].

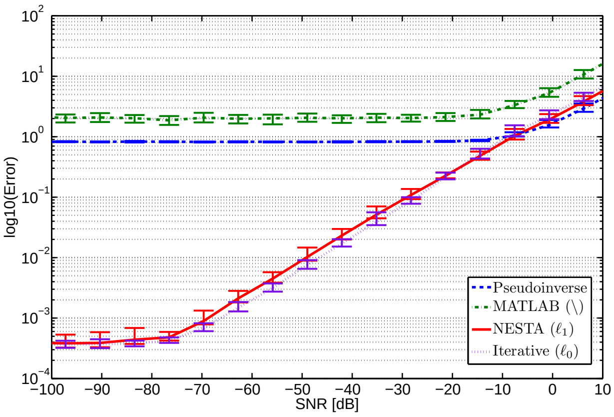


Figure 2: Variation of the estimation error with the signal-to-noise level averaged over 50 trials. $\hat{\boldsymbol{p}}$ had its length fixed to $N = 100$ and $S = 6$ non-zero elements. There were $S = 32$ randomly distributed sampling points.

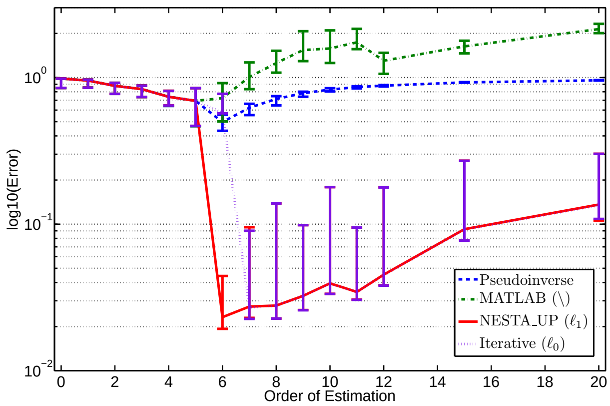


Figure 3: Variation of the estimation error with the order of the estimation vector averaged over 50 trials. $\hat{\boldsymbol{p}}$ had its length fixed to $N = 49$ and $S = 9$ non-zero elements. There were $S = 32$ randomly distributed sampling points and SNR = 40 dB.

discussed. Meanwhile, the sparsity promoting minimization algorithms achieve a much better performance, which furthermore does not degrade as quickly as the ℓ_2 minimization does. This example indicates that CS is relatively robust to mistakes in estimating the order of $\hat{\boldsymbol{p}}$.

We did not vary the value of M. It should be clear that increasing the value of M will reduce the approximation error; the interesting result is that for the same number of sampling points, a random distribution is likely to have a better performance than regularly distributed sensing points. It is also important to mention that the time taken to calculate the ℓ_1 and ℓ_2 minimization usually lies in the same order of magnitude.

7. DISCUSSION

In this paper we showed how Compressive Sampling can be used to measure the radiation pattern of sound sources with a reduced number of sensors. For the example discussed in this paper, spherical harmonics can be selected as an incoherent domain and ℓ_1 minimization can correctly recover the sparse spherical harmonic coefficients that describe the radiated sound field using three times fewer sensors than that required by traditional ℓ_2 minimization.

This result can be applied to reduce the time necessary to measure directivity balloons of loudspeakers or to reduce the number of points needed to measure individual "head-related transfer functions" (HRTFs). In this case, spherical harmonics (SH) might not be an adequate basis, since high frequency HRTFs cannot be considered sparse in the SH domain. A set of basis extracted from the "principal component analysis" of a group of individual HRTFs, might be a good candidate for an incoherent representation domain.

For the dual problem of sound field synthesis with a compact loudspeaker array, CS might also deliver an interesting counterpoint to the usual ℓ_2 minimization, as a solution with a reduced number of active loudspeakers would be striven. As shown by Epain *et al.* [2] and Lilis *et al.* [3] for the distributed source case, such a solution may deliver a sound field with a wider region of similarity to the original field.

8. ACKNOWLEDGMENTS

We would like to thanks Prof. Dr. Vítor Nascimento from EP-USP, Brazil, for the fruitful discussions that led to the results presented in this paper.

9. REFERENCES

[1] L. Hörchens and D. de Vries, "Plane wave decomposition: an inverse problem in array processing," in *Fortschritte der Akustik - DAGA*, 2009.

[2] N. Epain, C. Jin, and A. V. Schaik, "The application of compressive sampling to the analysis and synthesis of spatial sound fields," in *Proceedings of 127th AES Convention*, New York, USA, oct 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15052

[3] G. Lilis, D. Angelosante, and G. Giannakis, "Parsimonious sound field synthesis using compressive sampling," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09*, 2009, pp. 253–256.

[4] E. Candès and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21 –30, march 2008.

[5] M. Shahram, D. Donoho, and J. Starck, "Multiscale Representation for Data on the Sphere, and Applications to Geopotential Data," in *Proc. of SPIE*, vol. 6701. Citeseer, 2007, pp. 67 010A–1.

[6] M. Pollow, G. Behler, and B. Masiero, "Measuring directivities of natural sound sources with a spherical microphone array," in *Proceedings of Ambisonics Symposium*. Graz, AT: IEM Graz, Jun 2009.

[7] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, p. 1003–1010, mar 2007.

[8] S. Becker, J. Bobin, and E. Candès, "Nesta: A fast and accurate first-order method for sparse recovery," California Institute of Technology, Tech. Rep., apr 2009. [Online]. Available: http://www.acm.caltech.edu/~nesta