PARSIMONIOUS SOUND FIELD SYNTHESIS USING COMPRESSIVE SAMPLING

Georgios N. Lilis
Dept. of ECE
Cornell University
Ithaca, NY 14853

Daniele Angelosante, Georgios B. Giannakis
Dept. of ECE
University of Minnesota
Minneapolis, MN 55455

ABSTRACT
Reproducing a sampled sound field over a two-dimensional area using an array of loudspeakers is a problem with well-appreciated applications to acoustics and ultrasound treatment. Loudspeaker signal design has traditionally relied on a (possibly regularized) least-squares criterion. The fresh look advocated here, permeates benefits from advances in variable selection and compressive sampling by casting the sound field synthesis as a sparse linear regression problem that is solved by the least absolute shrinkage and selection operator (Lasso). Analysis and simulations demonstrate that the novel approach exhibits superb performance even for under-sampled sound fields, where least-squares methods yield inconsistent field reproduction. In addition, Lasso-based synthesis enables judicious placement of parsimonious loudspeaker arrays.

Index Terms— Wave field synthesis, least-squares, compressive sampling, sparse regression, Lasso.

1. INTRODUCTION
For more than fifty years there has been growing interest in reproducing accurately wave fields of primary natural sound sources, using multiple loudspeakers distributed in space. Applications to acoustics include reproduction of the primary source signals and also impressions regarding source locations and radiation characteristics.

Field reproduction using multiple sound sources dates back to [3]. Early attempts pursued the Kirchhoff-Helmholtz (KH) based approach that is limited by truncation and aliasing artifacts emerging in the reproduced sound fields due to the finite number of loudspeakers deployed [2]. Other popular alternatives aim to design signals of loudspeakers placed at prescribed locations so that after accounting for the known Green’s functions of the propagation medium, the synthesized wave field matches samples of the desired one in the least-squares (LS) sense. Such multi-point LS matching approaches are available in the time- or frequency-domains, or, through spherical harmonics; see e.g., [6], [7], [10] and references therein. In the simplest setting, the desired sound field is sensed (i.e., sampled) and matched with its LS approximant over a fixed grid. Accuracy of the LS fit depends on the number of sample points, and a large oversampling factor may be needed at high frequencies. In most cases of interest however, the overall reconstructed field is synthesized by a few loudspeakers close to the actual source(s) that produce the desired field. This prompts the idea that loudspeakers must be sparsely distributed in space, meaning that many loudspeakers placed on candidate positions on a selected grid should remain inactive – a feature that LS-based schemes cannot ensure.

Motivated by these facts, the present paper draws from signal processing advances in variable selection [9] and compressive sampling [4] to introduce a novel approach to wave field synthesis (WFS). Specifically, the LS cost is penalized with a scalar multiple of the \( \ell_1 \)-norm, and the minimizer of the aggregate objective function yields the so-called least-absolute shrinkage and selection operator (Lasso) [9] applied to the loudspeaker signals. The major contribution of Lasso-based WFS is twofold: (i) accurate reproduction is enabled with a few samples by exploiting the underlying sparsity; and (ii) judicious placement of a few loudspeakers over a dense grid achieves parsimonious synthesis of the desired field.

After modeling preliminaries, the problem is stated in Section 2. The novel algorithm for sound field reproduction is developed and discussed in Section 3. Simulated tests corroborating the merits of Lasso-based WFS and comparisons with its LS-based counterpart are provided in Section 4, while concluding remarks are given in Section 5.

Notation. Vectors (matrices) are denoted with lower (upper)-case boldface letters; \( (\cdot)^T \), \( (\cdot)^H \) and \( (\cdot)^\dagger \) stand for transpose, conjugate transpose, and the Moore-Penrose pseudoinverse, respectively. Symbol \( \Re (\mathbb{C}) \) denotes the real (complex) field. The \( p \)th entry of \( \mathbf{x} \in \mathbb{C}^P \) is denoted as \( x(p) \), and the \( (p, q) \)th entry of matrix \( \mathbf{R} \in \mathbb{C}^{P \times Q} \) as \( R(p, q) \); \( \| \mathbf{x} \|_1 \) and \( \| \mathbf{x} \|_2 \) are the \( \ell_1 \) and \( \ell_2 \) norms of \( \mathbf{x} \in \mathbb{C}^{n \times 1} \), respectively.

2. PRELIMINARIES AND PROBLEM STATEMENT

2.1. Modeling
Consider the sound field synthesis setup under the standard operating conditions:
(c1) Sound waves propagate in free space through a homogeneous and linear medium with speed $c$.

(c2) The spectrum $p_d$ of the desired two-dimensional field is sampled at $M$ locations $\{c_m\}_{m=1}^M \in S \subset \mathbb{R}^2$ at a given angular frequency $\omega_0 := 2\pi f_0$. Let $p_d \in \mathbb{C}^M$ denote a vector collecting the samples $p_d(m) \in \mathbb{C}$.

(c3) Vector $p_d$ is to be reproduced using $L$ monopole sound sources (loudspeakers) located at positions $\{c_l\}_{l=1}^L \in S$. The loudspeakers emit pure tones at $\omega_0$, amplified, and shifted in time. Time shifts correspond to complex scaling factors $s(\ell) \in \mathbb{C}$ in the frequency domain. Let $s \in \mathbb{C}^L$ denote the vector formed by concatenating these factors.

Although three-dimensional sound fields and reflective propagation environments can be also accommodated, (c1) is adopted here for simplicity in exposition. Since the medium is linear and homogeneous, the Green’s function capturing the frequency response of the propagation from the $\ell$-th loudspeaker to the $m$-th sampling location is

$$G(m, \ell; \omega_0) := \frac{e^{-j\frac{2\pi}{\lambda}||c_m - c_\ell||_2}}{4\pi ||c_m - c_\ell||_2}.$$  \hspace{1cm} (1)

Let $G \in \mathbb{C}^{M \times L}$ denote a matrix with $(m, \ell)$th entry given by $G(m, \ell; \omega_0)$; and $p_r := [p_r(1), \ldots, p_r(M)]^T \in \mathbb{C}^M$ a vector formed by the reproduced spectrum of the field at locations $\{c_m\}_{m=1}^M$. Under (c1)-(c3), the latter can be expressed as

$$p_r := Gs.$$ \hspace{1cm} (2)

Given $G$ and $p_d$, the goal of WFS is to select $s$ so that the reproduced $p_r$ matches “best” the desired $p_d$ over an area $\mathcal{R} \subset S$, “best” in a suitably defined fidelity criterion. The extra objectives in the present paper are to judiciously position a minimal number of loudspeakers, and also attain the WFS goal when the matrix $G$ is fat (under-sampled case).

2.2. WFS Using LS and Regularized LS

If LS is used as the fidelity criterion, WFS seeks the solution of the following minimization problem (see e.g., [7])

$$\hat{s}^{LS} := \arg \min_s \|p_d - Gs\|_2.$$ \hspace{1cm} (3)

If $G$ is tall (i.e., $M \geq L$) and has full column rank $L$, the unique solution of (3) is

$$\hat{s}^{LS} = G^\dagger p_d$$ \hspace{1cm} (4)

where $G^\dagger = (G^H G)^{-1}G^H$. If the tall $G$ is rank deficient, then (4) corresponds to the solution of (3) with minimum $\ell_2$ norm. Performance of the LS design in (4) depends on the condition number of $G^H G$, that is the ratio between the largest and smallest eigenvalues. In fact, if the condition number of $G$ is very large, the sound field will not be reproduced accurately over $\mathcal{R}$. Furthermore, the available total power of reproducing sources may be limited; e.g., to prolong the life of battery-operated wireless loudspeakers.

For these reasons, the standard LS criterion is often regularized to yield [6]

$$\hat{s}^{LS^2} := \arg \min_s \left[\|p_d - Gs\|_2^2 + \gamma \|s\|_1^2\right].$$ \hspace{1cm} (5)

The solution of (5) is also available in closed form as

$$\hat{s}^{LS^2} = (G^H G + \gamma I_L)^{-1}G^H p_d$$ \hspace{1cm} (6)

where the regularization parameter $\gamma$ is chosen to meet the power constraints, and render the problem well-conditioned.

For any estimated vector $\hat{s}$, the reproduction error and the transmitted power are given, correspondingly, by

$$E := \int_{\mathcal{R}} |p_d(\mathbf{r}) - p_r(\mathbf{r})|^2 d\mathbf{r}$$ \hspace{1cm} (7)

$$P := \|\hat{s}\|_2^2$$ \hspace{1cm} (8)

where $p_d(\mathbf{r})$ and $p_r(\mathbf{r})$ denote the desired and reproduced spectra of the sound fields at location $\mathbf{r} \in \mathcal{R}$, respectively. Note that all points on the error-power curve can be attained by suitably varying $\gamma$ over $[0, +\infty)$.

3. WFS Using Compressive Sampling

Suppose that the desired sound field is to be reproduced using a minimal number of loudspeakers placed close to the actual source(s). In this setup, LS allocates power to all loudspeakers ignoring the fact that the actual source(s) maybe located close to the loudspeakers.

Recent advances in statistics and sub-Nyquist sampling have established that when the signal of interest is sparse over a given basis, replacing the $\ell_2$-norm with the $\ell_1$-norm penalty enables parsimonious variable selection, meaning that the final estimate contains many zero entries [4, 9]. To effect such a sparse solution, (3) or (5) is replaced by the Lasso

$$\hat{s}^{Lasso} := \arg \min_s \left[\|p_d - Gs\|_2^2 + \lambda \|s\|_1\right]$$ \hspace{1cm} (9)

where the scalar $\lambda$ controls the degree of sparsity (the larger $\lambda$ is, the more zeros $\hat{s}^{Lasso}$ contains).

The minimization in (9) can be carried out using fast convex solvers with computational complexity comparable to LS [5]. The LS cost penalized with the scaled $\ell_1$-norm, yields reliable estimates both when $G$ is square or tall and ill-conditioned, but also when $G$ is fat [4, 9]. Similar to LS, one can attain any point on the error-power curve of the Lasso by varying $\lambda \in [0, +\infty)$.

When tailored for WFS, we have shown that Lasso offers the following distinct advantages over LS.$^1$

$^1$Detailed derivations are omitted due to lack of space, but can be found in the journal version of the present paper [8].
Lasso estimates are efficiently computed using coordinate descent. Indeed, when all but one entry of $\hat{s}_{\text{Lasso}}$ is unknown, the scalar solution of (9) is available in closed form, even in the complex case needed here; see also [5, 1] for related claims when $G$ is real.

Similar to recursive LS, Lasso can be implemented on-line [1] – an attractive feature when it comes to reproducing fields generated by moving sources.

Lasso outperforms (regularized) LS when the desired field is sampled at less locations than the number of loudspeakers ($M < L$).

Consider a dense grid of $L_c \gg L$ candidate loudspeaker locations. Solving (9) with $G \in \mathbb{C}^{M \times L_c}$ for a sequence of increasing $\lambda$ values, yields Lasso estimates with an increasing number of zero entries, which correspond to inactive loudspeakers. For specific values of $\lambda$, the number of non-zero entries (corresponding to active loudspeakers) equals $L$. This facilitates selection of $L$ favorable out of the $L_c$ candidate loudspeaker positions.

4. SIMULATED TESTS

4.1. Sound field reproduction via compressive sampling

An octagonal loudspeaker array comprising $L = 408$ pure tone sources with frequency $f_o = 1\text{KHz}$ is used to reproduce the sound field $S = [0,10] \times [0,10]$ of a point source located at (8,9) over the listening region $R = [3,7] \times [3,7]$ inside the octagonal array; see Figure 3. The desired acoustic field is sampled uniformly over $R$ at $M$ locations. Figures 1 and 2 show the error versus power curves for the regularized LS and the Lasso applied with either $M = 400$ samples (undersampled case), or, $M = 160,000$ samples (oversampled case). In the undersampled case, proper selection of $\lambda$ allows Lasso to exhibit considerably lower error than LS for the same power (compare solid and dashed curves in Fig. 1). Proper selection of $\lambda$ can be performed through cross-validation [9]. As expected, when the field is oversampled, LS outperforms the Lasso; see Fig. 2.

Figure 3 depicts the desired sound field and the fields reconstructed by the Lasso and the LS using $M = 400$ samples for the values of $\lambda$ and $\gamma$ that achieve the minimum squared error. Notice that only a few loudspeakers close to the source are active in the Lasso-based reproduction. The LS-based method on the other hand, overfits the sound field at the sampling points by allocating power to all loudspeakers. This overfitting that is inherent to the LS criterion causes unavoidable artifacts to emerge in the reproduction.

4.2. Judicious loudspeaker placement

The same octagonal geometry, desired field characteristics and listening area, are considered again here as in Section 4.1. For both LS and Lasso based methods, knowledge of favorable loudspeaker positions is not assumed. Because Lasso is capable of selecting subsets of locations by tuning $\lambda$, it is applied to a dense grid of $L_c = 408$ candidate locations, out of which $L \ll L_c$ are chosen. For fairness, LS is applied to uniform loudspeaker arrays also with $L$ elements. All elements are positioned on the same octagon. Figures 4 and 5 depict the error versus power curve for LS obtained by varying $\gamma \in [0, +\infty)$, along with the (power, error) points obtained by Lasso when the value of $\lambda$ is set such that $L$ active sources are selected out of $L_c$. Six different arrays are considered having $L_1 = 208$, $L_2 = 108$, $L_3 = 88$, $L_4 = 48$, $L_5 = 32$, and $L_6 = 28$ loudspeakers, respectively. In all cases tested, Lasso achieves lower reproduction error than LS for the same power.
Figure 3: Desired field (left); Lasso reproduction (center); LS reproduction (right). Undersampled case, $M = 400$.

Figure 4: Selection of loudspeakers.

Figure 5: Selection of loudspeakers (zoom-in).

5. CONCLUSIONS

A novel method for sound field reproduction has been developed using compressive sampling, and the related Lasso approach. Under the common setup where desired fields are induced by a few (or single) sources, and consequently exhibit sparsity in space, the resultant Lasso-based algorithm is more robust to ill conditioning of the Green’s matrix and more accurate than LS-based alternatives especially in reproducing sparsely sampled sound fields. As Lasso effects sparsity, the novel method also selects properly the positions of a minimal number of loudspeakers from a grid of candidate locations. Simulated tests confirmed these merits relative to the LS algorithm.

6. REFERENCES


