BROADBAND SENSOR LOCATION SELECTION USING CONVEX OPTIMIZATION IN VERY LARGE SCALE ARRAYS

Yenming M. Lai, Radu Balan*

University of Maryland
AMSC Program and Department of Mathematics
College Park, MD 20742

Heiko Claussen, Justinian Rosca

Siemens Corporation
Corporate Technology
Princeton, NJ 08540

ABSTRACT

Consider a sensing system using a large number of N microphones placed in multiple dimensions to monitor a broadband acoustic field. Using all the microphones at once is impractical because of the amount of data generated. Instead, we choose a subset of D microphones to be active. Specifically, we wish to find the set of D microphones that minimizes the largest interference gain at multiple frequencies while monitoring a target of interest. A direct, combinatorial approach – testing all N choose D subsets of microphones – is impractical because of problem size. Instead, we use a convex optimization technique that induces sparsity through a $l1$-penalty to determine which subset of microphones to use. We test the robustness of the our solution through simulated annealing and compare its performance against a classical beamformer which maximizes SNR. Since switching from a subset of D microphones to another subset of D microphones at every sample is possible, we construct a space-time-frequency sampling scheme that achieves near optimal performance.

Index Terms— array processing, multi-frequency beam pattern design, sensor location selection, very large scale arrays, convex optimization, simulated annealing

1. INTRODUCTION

Consider a large scale sensor array having N sensors that monitors a surveillance area. Using all sensors simultaneously may be unreasonable in terms of power consumption and data processing. For example, for $N = 10000$ sensors and a data sampling rate of 100000 samples per second, the bandwidth requirement is 1Gsamples/sec. We could poll a subset of D sensors at any one given time instead. The N choose D number of choices of sensors allows for a myriad of sensor configurations, and the task is then to choose a subset that achieves our objective.

Alternatively, our problem is to place D sensors in a constrained region of space. We convert this non-convex optimization problem into a combinatorial problem by discretizing the possible set of sensor locations. In this context our approach can be seen as an optimal test design.

Assume the surveillance area consists of a set of point-like sources. We seek designs that *minimize the largest interference gain* from a potentially very large number of locations while still maintaining target unit gain. Throughout this paper, we assume the following four hypotheses:

H1. Sensor locations $(x, y, z$-coordinates$)$ are known. Their locations however can be arbitrary.

H2. The number of real interferences, their locations and spectral powers are unknown.

H3. The maximum number of sensors active at any time, D, is fixed in advance.

H4. The signals that we wish to monitor are wide band.

When sensors have local computational processing power, we make an additional hypothesis:

H5. Sensors can band-pass signals and they can switch data transmission on a sample-by-sample basis.

Under H5, the system can sample selectively the space-time-frequency domain.

Our system divides the surveillance area into a large number of small, non-overlapping target areas. For each fixed target area, we find a subset of sensors that minimizes the gain of a large number of virtual interferences. By doing so, the system is robust to a wide variety of unknown interference configurations. It then begins to iteratively scan each target area. By measuring the signal strength in each fixed target area, we are able to locate and then monitor the actual targets.
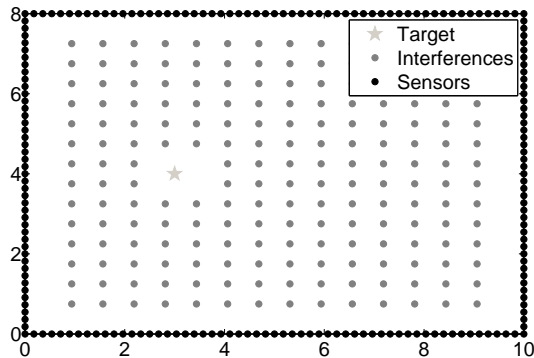

Fig. 1. Sample Scenario: $L = 193$ sources (1 target and 192 interferences), $N = 200$ sensors.

2. PROBLEM FORMULATION

Consider the setup described earlier and sketched in Figure 1. We assume we know the sensor locations and then fix a target area. Since the number, locations, and power distributions of the interferences are unknown, we assume a large number of virtual interferences, say L. Our objective is to minimize the largest gain from this set of virtual interferences which guarantees a small gain for any one actual intereference located at any of the virtual interference locations.

Assume the zero indexed source is the target and the remaining L sources are interferences. For these "virtual" $L + 1$ sources

we use the direct-path model, where the transfer function between source l and sensor n at frequency f_i is given by $H_{l,n}(f_i) = \exp(-2\pi i f_i c \|r_l - r_n\|)/\|r_l - r_n\|$, $0 \le l \le L, 1 \le n \le N$, where r_n, r_l denote the position vectors of sensor n and source l respectively, and c is the wave propagation speed (e.g. speed of sound for acoustic signals). We use this simple direct-path model because we do not want to assume prior knowledge of wall locations and their reflection coefficients. For evaluation, we use a direct-path plus one reflection off each wall model.

Let $w_n(f_i)$ denote the processing weight for sensor n at frequency f_i. If $w_n(f_i) = 0$ then sensor n is inactive at frequency f_i. Assume that we have F distinct frequencies of interest, $f_1, f_2, \ldots, f_F$. Given F sets of N sensor weights, each interference l has F different gains, one for each frequency. Note that for each of the F frequencies, we use a different set of N sensor weights to calculate the interference gains at that frequency. Our objective is to minimize the maximum *sum of gains across F frequencies* while still achieving target unit gain at each frequency. Assumption H3, however, limits us to D simultaneously active sensors.

In this multi-frequency setup, when H5 is not satisfied, sensor n becomes active if $w_n(f_i) \ne 0$ for any i. Thus the number of non-zeros in the N-dimensional vector $\max_{1 \le i \le F} |w_n(f_i)|$ is then an appropriate measure of the number of active sensors. Let $\|w\|_0$ denote the pseudo-norm which counts the number of non-zeros in the vector w. We can now state our optimization problem:

$$
\begin{aligned}
&\min_{(w_n(f_i))_{i=1,n=1}^{F,N}} \quad \max_{1 \le l \le L} \sum_{i=1}^{F} \left| \sum_{n=1}^{N} H_{ln}(f_i) w_n(f_i) \right| \\
&\text{subject to} \quad \sum_{n=1}^{N} H_{0n}(f_i) w_n(f_i) = 1 \text{ for } i = 1, 2, \ldots, F \\
&\qquad\qquad \left\| \max_{1 \le i \le F} |w_n(f_i)| \right\|_0 \le D
\end{aligned}
$$
(1)

Note that this is a non convex optimization problem due to the l^0 pseudo-norm constraint $\|\max_{1 \le i \le F} |w_n(f_i)|\|_0 \le D$. When hypothesis H5 is satisfied, we can solve the optimization problem (1) independently for each frequency, and then implement an adaptive space-time-frequency sampling. We shall compare both sampling scenarios.

3. CONVEX OPTIMIZATION STRATEGY

Our method is inspired by LASSO regularization [1], a regression technique that minimizes the sum of squares of residual errors subject to the l^1 norm of the coefficients being less than a constant. Similar to the sparse signal and model estimation approach in [2], the l^0 pseudo-norm is replaced by the l^1 norm $\|w\|_1 = \sum_{n=1}^{N} |w_n(f)|$ which is then incorporated into the optimization criterion using a Lagrange multiplier λ. The optimization problem (1) is then replaced by the following convex optimization problem which we call the λ-method:

$$
\min_{\substack{(w_n(f_i)) \\ i=1 \\ n=1}}^{F} \max_{1 \le l \le L} \sum_{i=1}^{F} \left| \sum_{n=1}^{N} H_{ln}(f_i) w_n(f_i) \right| + \lambda \sum_{n=1}^{N} \max_{1 \le i \le F} |w_n(f_i)|
$$

$$
\text{subject to } \sum_{n=1}^{N} H_{0n}(f_i) w_n(f_i) = 1 \text{ for } i = 1, 2, \ldots, F
$$
(2)

For very large λ, the penalty term forces many of the sensors to become inactive. Specifically, let W be the N by F matrix of sensor weights produced by the optimization. With a large λ penalty, many of the rows of W contain only values very close to zero. If the nth row only contains such small values, sensor n is then inactive. As λ decreases the penalty term becomes less expensive and more sensors become active. At the limit, $\lambda = 0$, all sensors are allowed to be active. We fine tune λ until we get D non-zero components. Specifically, this is accomplished when $g_{n_{D+1}}$ is significantly smaller than g_{n_1} (e.g. by 3 orders of magnitude), where $g_n = \max_{1 \le i \le F} |w_n(f_i)|$. We then solve (2) using this λ. The final weights are obtained by solving a second time (2) restricted to the sub matrix indexed by $\mathcal{A}$ and $\lambda = 0$, commonly known as the debiasing step [1].

When we optimize over multiple frequencies (2), the support of the chosen microphones remains fixed over the frequencies. However, we can also run (2) for each individual frequency of interest. Then the support of the microphones can vary over frequencies. If the frequencies are chosen with uniform spacing, we can use non-uniform sampling to reconstuct the signals of interest at each frequency. Specifically assume we have a total of R frequency bands (e.g. $R = 4$ as in the results below). Then each sensor has an additional R-channel filter bank each processing disjoint frequency bands. The filter bank outputs are downsampled by R and the corresponding sample streams are sent according the transmission policy obtained in the optimization step. The central controller reconstructs the broadband signal by putting together the disjoint frequency bands.

4. SIMULATED ANNEALING

We test the robustness of the solution found by convex optimization through simulated annealing (SA). Simulated annealing is a simple randomized technique for iterative improvement introduced in [3]. SA will probabilistically accept worse transitions in order to avoid local minima. In our case, SA minimizes the objective function given by the largest gain for all interference positions for a fixed-size subset of microphones over locations of the microphones in the subset. More precisely, given a fixed number of D microphones, we run the same convex optimization problem as the λ method (e.g. find the filter weights that minimize the maximum valid interference over a coarse grid while maintaining a target gain of unity) but with λ *set to 0*. The objective function value f is then the largest gain of the valid interferences. If the optimization program fails for any reason, the objective function value is set to infinity. The *initial temperature* is set to 100, and the *annealing schedule* decreases the temperature by 5% each iteration. We limit the length of the Markov walk to a *total length* of 1200 iterations. The initial search point is the point produced by the λ found through an iterative binary search that produces $D = 32$ microphones. We assume that the target location, room size, and the frequency of interest are fixed.

5. RELATION TO PRIOR WORK

In this section we compare and relate our approach to four problems in the literature: the beam pattern design, grid-based beamforming, compressive sampling, and greedy selection.

Beampattern Design

Lebret and Boyd [4] showed that given arbitrary microphone locations, finding the set of sensor weights that minimizes the maximum interference for multiple frequencies could be formulated as a convex optimization problem. They modeled sources as point sources, complex exponentials decaying as a function of distance. Specifically, the problem could be showed equivalent with a Second Order Cone Programming (SOCP) problem and thus efficiently

solved by interior point methods [5]. Our work differs in that we simultaneously search for filter weights for each frequency of interest that minimize the maximum interference at the corresponding frequency. In contrast, Lebret and Boyd search for only a single set of filter weights that minimizes the maximum inteference for multiple frequencies. Our approach is similar in spirit to that of Ling et al. [6]. The authors followed the approach of [4] but added an additional l^1 norm penalty to the weights. The l^1 penalty sparsifies the sensor weights and therefore sensor locations. Specifically, the l^1 penalty on weights causes many of the filter weights to be close to zero in magnitude. Our work differs in the following three aspects. First, our l^1 penalty encourages the use of sensor locations that perform well across frequencies when the sensor is allowed to use a different weight for each frequency. Second, we allow both sensors and intereferences to lie in multiple dimensions. Third, we do not add an l^2 penalty on the weights to protect against large gains. Instead, after optimizing over a large set of sensors with a l^1 penalty, we select the sensor locations with the largest weight values across frequencies and then reoptimize over this subset of locations. During the reoptimization, we exclude the l^1 penalty.

Grid-based Beamforming Brandstein and Ward modeled an acoustic enclosure as a rectangular grid of point sources that included reverbeations by allowing for sources to lie outside the room [7], an idea drawn from the "image" model where reflections of sources off of walls are modeled as virtual sources lying outside the room [8]. Grids were labeled either as sources or interferences based on prior knowledge. Sensor weights were calculated to maximize the optimization criterion given by the ratio between source gains and interference gains.

Compressive Sampling The widespread compressive sampling problem (see [9, 10]) is to to minimize the l^0 pseudo-norm of a vector x subject to a linear constraint $Ax = b$ (in the absence of noise), or an inequality $\|Ax - b\|_p \leq \varepsilon$ (in the presence of noise). Using Lagrange multipliers this problem becomes

$$\min_x \|Ax - b\|_p + \lambda \|x\|_0 \tag{3}$$

Problem (1), in the single frequency case ($F = 1$), can be brought to this form if we make the *additional* assumption that a specific sensor is active. For simplicity of notation assume that we know $w_N(f) \neq 0$. Then we solve for w_N from the linear constraint $\langle w, \bar{H}_0 \rangle = 1$ and substitute back into Hw. Denoting $\tilde{w}$ the $N - 1$ vector of the first $N - 1$ components of w, and A the $L \times (N - 1)$ matrix of entries $A_{l,n} = H_{l,n} - \frac{H_{0,n} H_{l,N}}{H_{0,N}}$ and b the L vector $b_l = -\frac{H_{l,N}}{H_{0,N}}$, (1) becomes

$$\min_{\|\tilde{w}\|_0 \leq D-1} \|A\tilde{w} - b\|_\infty \tag{4}$$

which turns into a problem similar in form to (3):

$$\min_{\tilde{w}} \|A\tilde{w} - b\|_\infty + \lambda \|\tilde{w}\|_0. \tag{5}$$

In the multi-frequency case, A,b, and $\tilde{w}$ are all frequency dependent. In (3), the 0 quasi-norm becomes the mixed-norm $(\infty, 0)$, where (∞) is the maximum acorss frequenices and (0) is the number of non-zeros among these maximums. In (4), the ∞-norm becomes the mixed-norm $(1, \infty)$ norm, where (1) is the summation across frequencies and (∞) is the maximum of across this resulting sum.

Greedy Selection In [11], Bertrand and Moonen propose two algorithms, greedy addition and greedy deletion, to select a subset of sensors. In greedy addition, they choose D subset of sensors out of N possible sensors by running D trials. At the n-th trial ($n = 1, 2, \ldots, D$), $N - n + 1$ individual candidate sensors are tested by adding each candidate sensor to the previously selected $n - 1$ sensors and measuring the n sensors performance. The sensor which produced the most improvement in the objective function when added to the previous $n-1$ sensors is then added to the selected group of sensors. In a corresponding fashion, in greedy deletion, they choose D sensors by starting with N sensors and then deleting $N - D$ members one by one. $N - D$ trials are run and during a trial a sensor is deleted if its contribution to the group's performance is least among the other group members. A sensor's contribution to the group performance is measured by how the group performs without the sensor. Though their objective function of minimizing mean square error in signal estimation is different from ours, their greedy approaches provide alternative strategies to choosing a subset of sensors. However, their approach differs from ours in that it is combinatorial by nature.

6. EXPERIMENTAL RESULTS

We run experiments by optimizing over a simpler model and then evaulating over a more sophisticated model. Our optimization model is as follows: The room size is 10m by 8m. The target of interest is located at (3m,4m). There are 1000 possible microphone locations located along the perimeter of the rectangular room. We optimize over four freqencies of 250,500,750,and 1000 Hz. There are 6200 virtual interferences, and a direct path model is used to calculate the transfer functions. We do not place these interferences within 0.5 m of the perimter of the room or the target. The evaluation model differs from the optimization model in two ways: There is a denser set of interferences, 620000, and we include reflections for each of these virtual interferences. We run five types of experiments to compare the performance of the λ-method. First, we run the optimization problem (2), simultaneously optimizing over the four frequencies. This setup fixes the support of the microphone setup across all 4 frequencies. Second, we run the optimization problem (2) again four times, once at each individual frequency. The support of the chosen microphones are then allowed to vary over frequencies. Third, we randomly perturb the set of sensors found by the multi-frequency optimization of experiment (1) using simulated annealing to see how much better a solution we can find. Fourth, we again use simulated annealing the perturb the set of microphones found by the single-frequency optimization of experiment (2). Fifth, we test the performance of the beamformer that maximizes the signal to noise ratio (SNR). This beamformer can be shown to be the set of sensors that lie closest to the target.

Table (1) shows the worst interference gain in dB for the five setups using the evaluation model. The results show our multi-freqency λ-processor (column 2) outperforming the beamformer that maximizes SNR (column 1) for every frequency. The single-frequency λ-processor performs better than multi-frequency λ-processor since microphone locations are allowed to vary across frequencies. Simulated annealing sometimes but not always finds better performing solutions when measured with the evaluation model. By algorithm construction, simulated annealing finds a solution at least as good as the initial point when measured on the optimization model. The below figures show the beam patterns for both the λ-method and the maximum-SNR beamformer at 1000 Hz along with the placement of the microphones. Results are in dB, with unit target gain (0 dB). We also compare the expected value and variance of the maximum gain among a varying number of randomly placed interferences in the surveillance area. Finally, we show the sharp drop in filter weights produced by the multi-frequency λ-processor.

7. CONCLUSIONS

We aim to utilize a very large number of available sensors by using customized subsets of sensors to monitor specific areas of interest. This selective sampling of sensors then produces reasonable amounts of data to be processed. An equivalent problem to our sen-

f[Hz]	BF	MF λ	SA MF λ	SF λ	SA SF λ
250	15.6	9.2	3.9	-1.19	-3.1
500	14.5	7.6	4.3	3.3	2.1
750	12.5	3.4	4.1	-0.9	-0.2
1000	10.4	2.4	5.8	1.5	0.1

Table 1. Worst interference gains[dB] for: Beamformer, MultiFreq λ-processor, SA MultiFreq λ, SingleFreq λ-processor, SA Single-Freq. λ

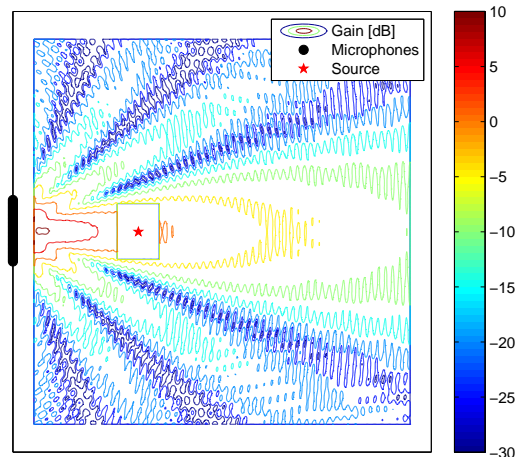sor subset selection is sensor spacing. Our optimization criterion finds sensors that suppresses a large number of interferences across multiple frequencies while monitoring a target location. We allow the subset of sensors we choose to have different weights for different frequencies of interest. We show that our multi-frequency LASSO-inspired convex optimization technique can find subsets of sensors that give reasonable performance on evaluation models that contain large number of virtual interferences and reflections even though the optimization criterion assumes many fewer virtual interferences and no reflections. If frequencies of interest are uniformly spaced, we can acheive even better performance by allowing the active sensor subset to change over frequencies and then reusing space-time-frequency sampling to recover our signal of interest.
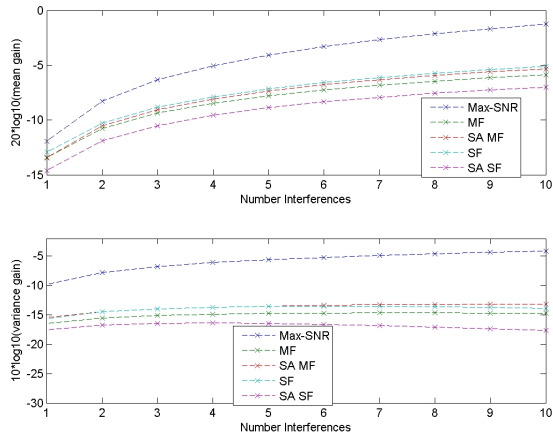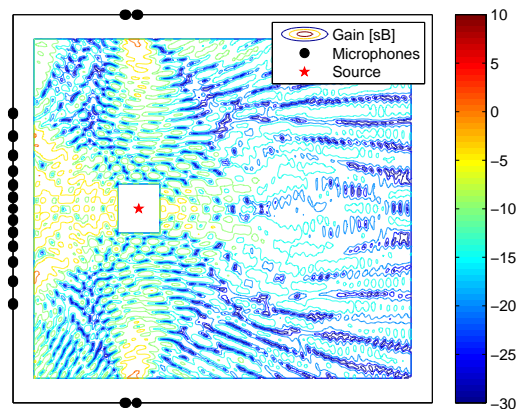

Fig. 2. Comparison of mean and variance of maximum gain for varying number of randomly placed interferences at 1000 Hz

Fig. 3. Gain map for classical beamformer that maximizes SNR at 1000 Hz

Fig. 4. Gain map for beamformer found by multi-frequency λ-method at 1000 Hz

8. REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society (Series B)*, vol. 58, pp. 267–288, 1996.

[2] E.J. Candes and Y. Plan, "Near-ideal model selection by l1 minimization," *Annals of Statistics*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[3] S. Kirkpatrick and MP Vecchi, "Optimization by simmulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[4] H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," *Signal Processing, IEEE Transactions on*, vol. 45, no. 3, pp. 526–532, 1997.

[5] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear algebra and its applications*, vol. 284, no. 1, pp. 193–228, 1998.

[6] C. Ling, S. Wee, C. Wei, and Y. Zhu Liang, "Linear sparse array synthesis via convex optimization," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 2010, pp. 4233–4236.

[7] M. S. Brandstein and D. B. Ward, "Cell-based beamforming (ce-babe) for speech acquisition with microphone arrays," *Speech and Audio Processing, IEEE Transactions on*, vol. 8, no. 6, pp. 738–743, 2000.

[8] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, 1979.

[9] R.G. Baraniuk, E. Candes, R. Nowak, and M. Vetterli, "Compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 12–13, 2008.

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

[11] Alexander Bertrand and Marc Moonen, "Efficient sensor subset selection and link failure response for linear mmse signal
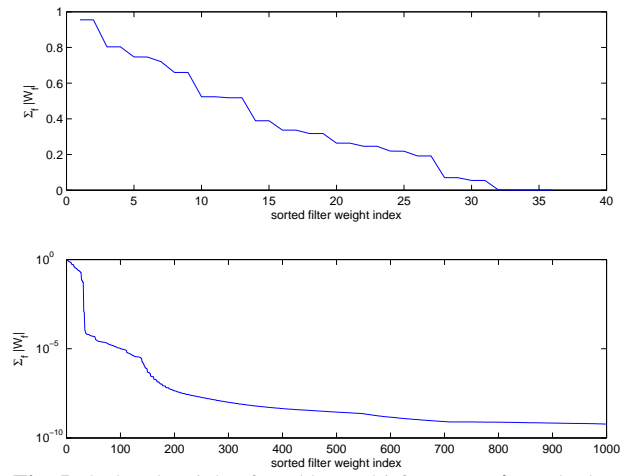
Fig. 5. Ordered weights found by multi-frequency λ-method

estimation in wireless sensor networks," in *Proc. of the European signal processing conference (EUSIPCO)*, 2010, pp. 1092–1096.