Manifold Sparse Beamforming

Baran Gözcü*, Afsaneh Asaei†, and Volkan Cevher*
*École Polytechnique Fédérale de Lausanne, Switzerland †Idiap Research Institute, Martigny, Switzerland
Emails: baran.goezcue@epfl.ch, afsaneh.asaei@idiap.ch, volkan.cevher@epfl.ch

***Abstract*—We consider the minimum variance distortionless response (MVDR) beamforming problems where the array covariance matrix is rank deficient. The conventional approach handles such rank-deficiencies via diagonal loading on the covariance matrix. In this setting, we show that the array weights for optimal signal estimation can admit a sparse representation on the array manifold. To exploit this structure, we propose a convex regularizer in a *grid-free* fashion, which requires semi-definite programming. We then provide numerical evidence showing that the new formulation can significantly outperform diagonal loading when the regularization parameters are correctly tuned.**

I. Introduction

Consider the standard array observation model:

$$x = a(\theta_s)s + \sum_{k=1}^{K} a(\theta_{I_k})I_k + n, \tag{1}$$

where $s \in \mathbb{C}^{1\times T}$ denotes T snapshots of an unknown narrowband signal impinging on an array of sensors from a known direction θ_s in the presence of K-interferers I_k each at unknown direction of arrivals θ_{I_k}; $x \in \mathbb{C}^{M\times T}$ denotes the array observations vector, $n \in \mathbb{C}^{M\times T}$ is an i.i.d. Gaussian noise, and $a(\theta)$ is the array manifold vector. We assume uniform linear arrays with half source wavelength element separation, and hence, $a(\theta)$ can be written as

$$a(\theta) = \begin{bmatrix} 1 \\ e^{-j\pi\cos(\theta)} \\ e^{-j2\pi\cos(\theta)} \\ \vdots \\ e^{-j(M-1)\pi\cos(\theta)} \end{bmatrix}, \text{ where } j = \sqrt{-1}. \tag{2}$$

Prediction of s given the array observations x is known as the beamforming problem in array signal processing [1], and also has broad applications in statistics and machine learning. In this paper, we study the following linear estimator $\widehat{s} = {w^*}^H x$ of the source s, where $.^H$ denotes the Hermitian operator; the so-called steering weights w^* is obtained via minimization of the expected prediction risk in terms of mean square error (MSE)

$$w^* \in \underset{w\in\mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x - s\|_2^2\}. \tag{3}$$

Unfortunately, we cannot hope to solve the optimization problem above since we only know the source bearing θ_s and not the source s.

Among beamforming techniques, the most popular one perhaps is the minimum variance distortionless response (MVDR), whose weights are obtained via

$$w^*_{\text{MVDR}} \in \underset{w\in\mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x\|_2^2\} \quad \text{s.t.} \quad \text{Re}\{w^H a(\theta_s)\} = 1. \tag{4}$$

In (4), the constraint $\text{Re}\{w^H a(\theta_s)\} = 1$ ensures that the ideal signal passes through with unity gain. Minimizing $\mathbb{E}\{\|w^H x\|_2^2\}$ subject to this constraint then suppresses all the other signals that do not resemble the ideal signal.

Note that the MVDR objective function can be written as $w^H\Sigma w$, where $\Sigma = \mathbb{E}\{xx^H\}$ is the true array covariance matrix. Since we do not in general know the distribution of the noise sources in the observation model, we simply replace the true covariance matrix with its empirical covariance estimate $\widehat{\Sigma} = xx^H$. Then the optimal MVDR solution is given by

$$\widehat{w}_{\text{MVDR}} = \frac{\widehat{\Sigma}^{-1} a(\theta_s)}{a(\theta_s)^H \widehat{\Sigma}^{-1} a(\theta_s)}. \tag{5}$$

Even in the idealistic scenarios (i.e., no calibration error, no bearing mismatch, and no additive noise), MVDR estimates can suffer significant performance degradation in practice. As a stylistic example, consider the case where empirical covariance estimate $\widehat{\Sigma}$ is rank-deficient, which often happens in MIMO communication applications or when we have a high signal-to-noise ratio (i.e., $n \approx 0$ in (1)). In this case, we cannot directly take the inverse of the empirical covariance matrix since the optimal steering weights w^* is non-unique.

As a result, we typically use diagonal loading of the empirical covariance matrix by substituting $\widehat{\Sigma} = xx^H + \lambda_{dl}\mathbb{I}$ in (5) where $\mathbb{I}$ denotes the identity matrix and $\lambda_{dl} > 0$ is the regularization parameter [2, 3]. This operation corresponds to the following Tikhonov regularized problem

$$\widehat{w}_{dl} = \underset{\mathrm{Re}\{w^H a(\theta_s)\}=1}{\arg\min} \; w^H \widehat{\Sigma} w + \lambda_{dl} \|w\|_2^2. \tag{6}$$

This paper argues that one can obtain a better prediction performance (3) by exploiting a heretofore unexplored structure in beamforming: sparsity of the weight vector over the array manifold. Sparsity enforcing has been shown to be beneficial in terms of sidelobe suppression of the beampattern [4]; this paper however considers a fundamentally different formulation [5]. We illustrate that when there is no additive noise in the array observations, the ideal steering weights can be represented sparsely in terms of the manifold vectors. While it is well-known that sparsity regularization helps in empirical risk minimization, to the best of our knowledge, this has not been demonstrated in beamforming problems.

We use this motivation to propose a new regularizer based on the atomic norm formulation of [6]. Hence, we dub our formulation sparse manifold beamforming as the weights are regularized using the sparsity assumption on the manifold vectors. We numerically illustrate that atomic beamforming can significantly improve the prediction performance by comparing diagonal loading against the atomic norm regularization using oracle parameters λ (i.e., optimized using the true prediction risk). For the numerical solution of the proposed formulation, we use a semi-definite relaxation and provide a grid free solution on the array manifold.

The article is organized as follows. In Section II we establish that MVDR is empirical risk minimization of the linear prediction when interferers are uncorrelated with the source signal. The fundamentals of an atomic beamformer and the motivation behind its use are given in Section III and Section IV respectively. We explain the simulation results in Section V and discuss the benefits of our technique. The conclusions are drawn in Section VI.

II. A *Risky* TALE OF MVDR

Here, we provide some relevant beamforming preliminaries for signal prediction for the sake of completeness. We first show that when the source signal is uncorrelated with the interferers, the MVDR beamformer approximately optimizes the empirical signal prediction MSE risk. To see this, we reconsider (3) assuming that $\mathbb{E}\{s^H I_k\} = 0$ $(k = 1, \ldots, K)$ and $\mathbb{E}\{s^H n\} = 0$:

$$\begin{aligned} w^* &\in \underset{w \in \mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x - s\|_2^2\} \\ &\in \underset{w \in \mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x\|_2^2 + \|s\|_2^2 - 2\mathrm{Re}\{s^H w^H x\}\} \\ &\in \underset{w \in \mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x\|_2^2\} + \mathbb{E}\{\|s\|_2^2\} \\ &\quad - 2\mathrm{Re}\, \mathbb{E}\left\{ s^H w^H \left(a(\theta_s) s + \sum_{k=1}^{K} a(\theta_{I_k}) I_k + n \right) \right\} \\ &\in \underset{w \in \mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x\|_2^2\} - 2\mathrm{Re}\{w^H a(\theta_s)\} \times \mathbb{E}\{\|s\|_2^2\}, \end{aligned} \tag{7}$$

where the last line can be written as follows for some constant c due to the convexity of the terms:

$$w^* \in \underset{w \in \mathbb{C}^M}{\arg\min}\, \mathbb{E}\{\|w^H x\|_2^2\} \quad \text{s.t.} \quad \mathrm{Re}\{w^H a(\theta_s)\} = c. \tag{8}$$

In the absence of prior information on the source signal power, MVDR stipulates that we pass the source signal untouched, and hence, enforces $c = 1$. We emphasize that this may not be the optimal solution to the risk minimization problem according to our derivation above.

III. ATOMIC BEAMFORMING

When $\widehat{\Sigma}$ is rank deficient, we typically use Tikhonov regularization as described in (6) to obtain a unique solution (and hence replace $\in$ with equality). This is also known to improve performance when the condition number of the empirical covariance matrix is high. This solution is often referred to as *diagonal loading MVDR*. In this work, we instead propose a new regularization for MVDR:

$$w_A^* = \underset{\mathrm{Re}\{w^H a(\theta_s)\}=1}{\arg\min} \; w^H \widehat{\Sigma} w + \lambda_A \|w\|_{\mathcal{A}}, \tag{9}$$

where $\|w\|_{\mathcal{A}}$ is the so-called atomic norm defined as

$$\begin{aligned} \|w\|_{\mathcal{A}} &= \inf\{t > 0 \,|\, w \in t\mathrm{conv}(\mathcal{A})\} \\ &= \inf\left\{ \sum c_i | w = \sum_{a_i \in \mathcal{A}} c_i a_i, c_i \geqslant 0 \right\}. \end{aligned} \tag{10}$$

The atomic norm is a convex norm that allows one to decompose a vector w into its atoms a in a set $\mathcal{A}$—possibly—with infinite cardinality. When the atoms are not symmetric with respect to the origin, it is defined using a convex gauge [6].

In our case, we take the complex manifold vectors (2) as our atoms, and hence

$$\mathcal{A} = \{a(\theta), \theta \in [0, \pi]\}.$$

We can exploit the fact that our manifold vectors $a(\theta)$ have the same structure of spectral estimation problem on continuous normal frequency grid [6]. Therefore, following the formulation of the line spectral estimation problem presented in [6], we can translate our minimization problem (9) into

$$\begin{aligned} \underset{t,u,w}{\text{minimize}} \quad & w^H \widehat{\Sigma} w + \frac{\lambda_A}{2}(t + u_1) \\ \text{subject to} \quad & \begin{bmatrix} T(u) & w \\ w* & t \end{bmatrix} \succeq 0 \\ & \text{Re}\{w^H a(\theta_s)\} = 1, \end{aligned} \tag{11}$$

where T is a Toeplitz operator applied on the vector $u = [u_1 u_2 \dots u_M]^T$, u_1 being the first element of it. This is a semidefinite programming (SDP) that can be solved with CVX package in MATLAB [7]. The atomic MVDR yields significantly lower prediction risk compared to the diagonal loading approach as it will be seen in Section V. The following section provides some insights on the sparsity structure of the optimal weights w^* on the array manifold.

IV. Sparsity of w^* on the Array Manifold

In this section we give a heuristic justification for the sparsity of optimum weight vector on our atomic set $\mathcal{A}$. For concreteness, we consider the case where the additive noise is insignificant compared to the source and interferers so that the array covariance matrix is rank deficient. We now revisit (7) to understand the subspace restrictions on the solution $w^* \in$

$$\underset{\text{Re}\{w^H a(\theta_s)\}=1}{\arg\min} \quad \mathbb{E}\left\{ \left\| w^H \left(a(\theta_s)s + \sum_{k=1}^{K} a(\theta_{I_k}) I_k \right) \right\|_2^2 \right\}.$$

Assuming that the source and the interferers are uncorrelated, we obtain the following optimization problem for w^*

$$\underset{\text{Re}\{w^H a(\theta_s)\}=1}{\arg\min} \quad \mathbb{E}\left\| w^H a(\theta_s) s \right\|_2^2 + \mathbb{E}\left\| \sum_{k=1}^{K} w^H a(\theta_{I_k}) I_k \right\|_2^2.$$

Now, we are going to replace the constraint $\text{Re}\{w^H a(\theta_s)\} = 1$ with the constraint $w^H a(\theta_s) = 1$. Note that this constraint is more stringent since it forces the imaginary part of $w^H a$ to be exactly zero as opposed to be free. If we can find sparse solutions here, we can always find sparse solutions to the original problem. Under this modification, we have the following problem

$$w^* \in \underset{w^H a(\theta_s)=1}{\arg\min} \quad \mathbb{E}\left\| \sum_{k=1}^{K} w^H a(\theta_{I_k}) I_k \right\|_2^2.$$

This suggests that if the modified weight vector can satisfy the nulling equations $w^H a(\theta_{I_k}) = 0$ for $k = 1, ..., K$, then the error is minimized. Assuming that w^* could be written as a S-sparse linear combination of atoms $w^* = \sum_{k=1}^{S} c_k a(\theta_k)$, we would like to see if the following linear system is feasible

$$\begin{bmatrix} a(\theta_s)^H \\ a(\theta_{I_1})^H \\ a(\theta_{I_2})^H \\ \vdots \\ a(\theta_{I_K})^H \end{bmatrix} [a(\theta_1) a(\theta_2) \dots a(\theta_S)] \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_S \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

We claim that when the interferer angles θ_{I_k}s are random, w^* lies in a $M - K + 1$ dimensional subspace which could be spanned by $S = M - K + 1$ atoms. Therefore this set of equations can have a solution when $K + 1 \leqslant M - K + 1$, i.e., $K \leqslant \frac{M}{2}$. Hence, the original problem can have sparse solutions when the number of interferers is small relative to the number of array elements.

The sparsity regularization can be beneficial as it restricts the search space to the null space of the interferences. Section V demonstrates that the performance improvement obtained by sparsity regularization even persists when $K > \frac{M}{2}$. At this point, the theoretical justification for that is not clear. We believe that the sparsity of the weight vectors on the manifold improves the robustness of beamforming when the empirical covariance estimates are used. We leave the rigorous discussion on this topic for future work.

V. Numerical experiments

In this section we compare the performances of diagonal loading (6) and atomic beamforming (11) based on MATLAB CVX simulations. To see the impact on the actual signal prediction error, we tune the regularizer parameters λ via an oracle model which is basically a crude search for the best hyperparameter having access to the desired source signals. To implement the oracle, we first perform an exhaustive grid search over a limited number of regularization parameters λ within an specified range by calculating the actual prediction risk to narrow down the regularization parameter range. We then use continuous optimization to obtain the regularizer values using the `fmincon` function of MATLAB for which a good initial point is provided by the grid search.

A. Experiment Setup

The number of sensors is taken as $M = 8$ whereas the total number of snapshots is $T = 80$. For one source

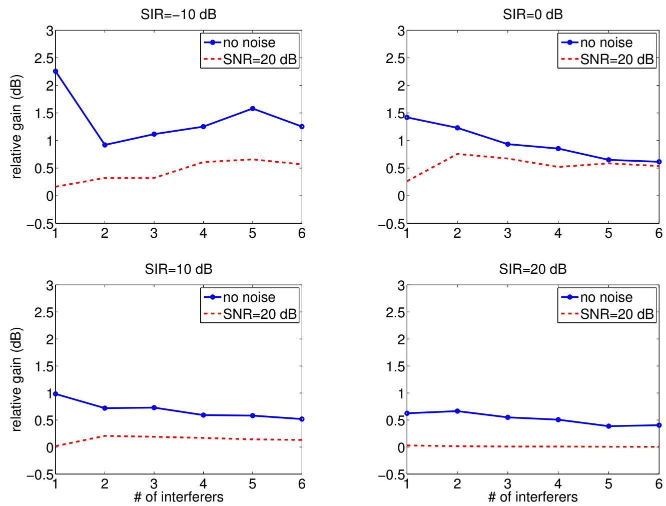

Fig. 1: Relative mean square error gain of atomic beamforming over diagonal loading vs. number of interferers

signal and $K = 1, 2, \ldots, M - 2$ interferers, the optimal weights are calculated for diagonal loading and atomic beamforming based on (6) and (11), respectively.

The experiments are repeated for 2500 runs each of which included, beside different number of interferers, different signal qualities: no noise and SNR= 20 dB with $-10, 0, 10, 20$ dB signal-to-interferer ratio (SIR) values. Note that the SIR value is defined as the total signal to interference ratio, therefore individual power of interferers decreases as the number of interferers increases.

B. Evaluation Results

The results of a noiseless scenario are depicted in Fig. 1. The relative error gain is defined as

$$10 \log_{10} \|w_A^* x - s\|_2^2 - 10 \log_{10} \|w_{dl}^* x - s\|_2^2. \qquad (12)$$

We can see that at low number of interferers and at a low SIR value the sparsity constraint on the weight vectors helps to gain several dB of improvement in terms of MSE prediction risk. However at higher SIR, contribution of atomic beamforming is less since the nulling equations forcing the sparsity in atomic manifold set becomes less and less significant. On the other hand, Fig. 1 also demonstrates that when there is noise in the measured signal, this effect is less visible since the empirical matrices are not rank deficient anymore and the atomic regularization is less effective.

Furthermore, we compare the CPU time for diagonal loading as well as atomic MVDR which confirms that both methods could perform real-time, although atomic beamformer requires slightly more computations. It should be noted that the use of atomic norm regularization for a grid-free sparse recovery is computationally more favorable than the discretized version solved using Lasso [8] as the complexity of the former is $O(M^{3.5})$ whereas the discretized version on a grid of G cells has a complexity of $O(G)$ which can be greater for a typically large grid.

VI. Conclusion

In this paper, we propose a new regularization approach for beamforming to handle the rank deficient covariance matrices. Our approach is based on enforcing a sparse and grid-free representation of optimum weights vector on the array manifold. To exploit this structure, we define a convex regularizer, and formulate our beamforming approach as a semidefinite programming problem. Numerical results illustrate that in our setting, one can obtain substantial performance improvement over the traditional diagonal loading approach. Although the results are quite encouraging, many aspects of this research are yet to be explored and investigated rigorously and extensively.

VII. Acknowledgments

Baran Gözcü and Volkan Cevher are supported in part by the European Commission under Grant MIRG-268398, ERC Future Proof and SNF 200021-132548. Afsaneh Asaei is supported by the Swiss National Science Foundation under the National Center of Competence in Research (NCCR) on Interactive Multi-modal Information Management (IM2).

References

[1] D. H. Johnson and D. E. Dudgeon, *Array signal processing: concepts and techniques*. Simon & Schuster, 1992.
[2] H. Cox, R. M. Zeskind, and M. H. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 35, 1987.
[3] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Processing, Special Issue: Advances in Sensor Array Processing*, 2013.
[4] Y. Zhang, B. P. Ng, and Q. Wan, "Sidelobe suppression for adaptive beamforming with sparse constraint on beam pattern," *Electronics Letters*, vol. 44, 2008.
[5] A. Asaei, "Model-based sparse component analysis for multiparty distant speech recognition," Ph.D. dissertation, Ecole Polytechnique Federal de Lausanne (EPFL), 2013.
[6] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation, http://arxiv.org/abs/1204.0562," 2013.
[7] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," Sep. 2012.
[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, 1994.