ROBUST SUPERDIRECTIVE BEAMFORMER WITH OPTIMAL REGULARIZATION

Aviva Atkins, Yuval Ben-Hur, Israel Cohen

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel

Jacob Benesty

INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada

ABSTRACT

In this paper, we introduce an optimal beamformer design that facilitates a compromise between high directivity and low white noise amplification. The proposed beamformer involves a regularization factor, whose optimal value is determined using a simple and efficient one dimensional search algorithm. Simulation results demonstrate controlled tuning of various gain properties of the desired beamformer, and improved performance compared to a competing method.

***Index Terms*—** Microphone arrays, beamforming, delay-and-sum beamformer, superdirective beamformer, robust superdirective beamformer, supergain, white noise gain, directivity factor.

1. INTRODUCTION

Most fixed conventional beamformers are optimally designed for a given noise field. The most well-known beamformers are the delay-and-sum (DS), which maximizes the signal-to-noise ratio (SNR) gain under white noise conditions, and the superdirective beamformer [1, 2], which does the same only for diffuse noise. Realistic environments, however, are likely to impose several types of noises all at once. Unfortunately, it turns out that beamformers designed to operate solely under white noise perform poorly under diffuse noise, and vice-versa. Hence, extensive work has been done to find a superdirective beamformer with increased robustness to the white noise.

Cox et al. [1, 3] introduced an optimal beamformer which is derived when the white noise gain is constrained. Other methods suggested different variations on optimization problems, e.g., diagonal loading [4], or addressing microphone characteristics mismatch [5, 6], to solve this trade-off. Recently, Berkun et al. [7, 8] proposed robust approaches, which use a closed-form expression that enables tuning the beamformer's performance under various noise types. However, in almost all of these designs, the regularization factor, which is necessary for obtaining optimal results, is not easy to find. Often, the regularization factor is set by some heuristic considerations or some prior knowledge regarding the signal and the interference.

In this paper, we address the trade-off between the beamformer performances under white noise and diffuse noise by taking a slightly different approach. In Section 2, we present the signal model and the array setup as well as some basic performance measures. Section 3 summarizes the properties of conventional beamformers: DS, superdirective, and regularized superdirective. Next, in Section 4, we propose the usage of a combined noise field, composed of both white and diffuse noise. Considering this new noise model, we define the relevant SNR gain criterion and find the respective optimal beamformer. We then present a simple and computationally efficient search algorithm for calculating the optimal regularization factor. Section 5 shows simulation results, which demonstrate our design method and its improved performance compared to the combined beamformers method described in [7]. Section 6 concludes the paper and offers future research possibilities.

2. SIGNAL MODEL AND ARRAY SETUP

We consider a plane wave, in the farfield, that propagates in an anechoic acoustic environment at the speed of sound in air and impinges on a uniform linear sensor array consisting of M omnidirectional microphones. The distance between two successive sensors is equal to δ and the direction of the source signal to the array is parameterized by the azimuth angle θ. The steering vector (of length M) is therefore given by

$$\mathbf{d}(\omega,\theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0\cos\theta} & \cdots e^{-j(M-1)\omega\tau_0\cos\theta} \end{bmatrix}^T, \quad (1)$$

where the superscript ^{T} is the transpose operator, $j = \sqrt{-1}$ is the imaginary unit, $\omega = 2\pi f$ is the angular frequency, $f > 0$ is the temporal frequency, and $\tau_0 = \delta/c$ is the delay between two successive sensors at the angle $\theta = 0$. We are interested in superdirective [1, 3] or differential beamforming [9, 10], where the inter-element spacing, δ, is small, the main lobe is at the angle $\theta = 0$ (endfire direction), and the desired signal propagates from the same angle. With the conventional signal model [10], the observation signal vector (of length M) is

$$\begin{aligned}\mathbf{y}(\omega) &= \begin{bmatrix} Y_1(\omega) & Y_2(\omega) & \cdots & Y_M(\omega) \end{bmatrix}^T \\ &= \mathbf{x}(\omega) + \mathbf{v}(\omega) = \mathbf{d}(\omega)X(\omega) + \mathbf{v}(\omega),\end{aligned} \quad (2)$$

978-1-5090-2007-2/16/$31.00

where $Y_m(\omega)$ is the mth microphone signal, $\mathbf{x}(\omega) = \mathbf{d}(\omega) X(\omega)$, $X(\omega)$ is the desired signal, $\mathbf{d}(\omega) = \mathbf{d}(\omega, 0)$ is the steering vector at $\theta = 0$ (direction of the source), and $\mathbf{v}(\omega)$ is the additive noise signal vector. By applying a complex-valued linear filter, $\mathbf{h}(\omega)$, to the observation signal vector, we obtain the beamformer output [11]:

$$\begin{aligned} Z(\omega) &= \mathbf{h}^H(\omega)\mathbf{y}(\omega) \\ &= \mathbf{h}^H(\omega)\mathbf{d}(\omega)X(\omega) + \mathbf{h}^H(\omega)\mathbf{v}(\omega), \end{aligned} \tag{3}$$

where $Z(\omega)$ is an estimate of the desired signal, $X(\omega)$, and the superscript $(\cdot)^H$ is the conjugate-transpose operator. In our context, the distortionless constraint is desired, i.e., $\mathbf{h}^H(\omega)\mathbf{d}(\omega) = 1$.

3. PERFORMANCE MEASURES AND CONVENTIONAL BEAMFORMERS

The first important measures are the input and output SNRs. Taking the first microphone as a reference, we can define the input SNR as

$$\text{iSNR}(\omega) = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)},$$

where $\phi_X(\omega) = E\left[|X(\omega)|^2\right]$ and $\phi_{V_1}(\omega) = E\left[|V_1(\omega)|^2\right]$ are the variances of $X(\omega)$ and $V_1(\omega)$, respectively, with $E[\cdot]$ denoting mathematical expectation. The output SNR is defined as

$$\text{oSNR}[\mathbf{h}(\omega)] = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)} \times \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega)\right|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\mathbf{v}}(\omega)\mathbf{h}(\omega)},$$

where $\mathbf{\Gamma}_{\mathbf{v}}(\omega) = \frac{E\left[\mathbf{v}(\omega)\mathbf{v}^H(\omega)\right]}{\phi_{V_1}(\omega)}$ is the pseudo-coherence matrix of $\mathbf{v}(\omega)$. From the two previous definitions, we deduce the gain in SNR:

$$\mathcal{G}[\mathbf{h}(\omega)] = \frac{\text{oSNR}[\mathbf{h}(\omega)]}{\text{iSNR}(\omega)} = \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega)\right|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\mathbf{v}}(\omega)\mathbf{h}(\omega)}. \tag{4}$$

The most convenient way to evaluate the sensitivity of the array to some of its imperfections such as sensor noise is via the so-called white noise gain (WNG), which is defined by plugging $\mathbf{\Gamma}_{\mathbf{v}}(\omega) = \mathbf{I}_M$ ($\mathbf{I}_M$ is the $M \times M$ identity matrix) into (4):

$$\mathcal{W}[\mathbf{h}(\omega)] = \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega)\right|^2}{\mathbf{h}^H(\omega)\mathbf{h}(\omega)} \leq M. \tag{5}$$

It is easy to see that $\mathcal{W}[\mathbf{h}(\omega)]$ is maximized with the well-known DS beamformer:

$$\mathbf{h}_{\text{DS}}(\omega) = \frac{\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)\mathbf{d}(\omega)} = \frac{\mathbf{d}(\omega)}{M}. \tag{6}$$

Another important measure, which quantifies how the microphone array performs in the presence of reverberation, is the directivity factor (DF). Considering the spherically isotropic (diffuse) noise field, the DF is defined as

$$\mathcal{D}[\mathbf{h}(\omega)] = \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega)\right|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{d}}(\omega)\mathbf{h}(\omega)} \leq M^2, \tag{7}$$

where $\mathbf{\Gamma}_{\text{d}}(\omega) = \frac{1}{2}\int_0^\pi \mathbf{d}(\omega,\theta)\mathbf{d}^H(\omega,\theta)\sin\theta d\theta$. It can be verified that the elements of the $M \times M$ matrix $\mathbf{\Gamma}_{\text{d}}(\omega)$ are

$$[\mathbf{\Gamma}_{\text{d}}(\omega)]_{ij} = \frac{\sin[\omega(j-i)\tau_0]}{\omega(j-i)\tau_0} = \text{sinc}[\omega(j-i)\tau_0].$$

It can be shown that $\mathcal{D}[\mathbf{h}(\omega)]$ is maximized with the conventional superdirective (SD) beamformer [3]:

$$\mathbf{h}_{\text{SD}}(\omega) = \frac{\mathbf{\Gamma}_{\text{d}}^{-1}(\omega)\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)\mathbf{\Gamma}_{\text{d}}^{-1}(\omega)\mathbf{d}(\omega)}. \tag{8}$$

This filter is a particular form of the celebrated minimum variance distortionless response (MVDR) beamformer [12, 13]. While the DS beamformer maximizes the WNG and never amplifies the diffuse noise since $\mathcal{D}[\mathbf{h}_{\text{DS}}(\omega)] \geq 1$, it performs poorly in reverberant and noisy environments, even with a large number of microphones, because its DF is relatively low. On the other hand, with the superdirective beamformer we can obtain a DF close to M^2, which is good for speech enhancement (i.e., dereverberation and noise reduction), but the WNG can be much smaller than 1, especially at low frequencies, implying a severe problem of white noise amplification, which is the most serious issue with the SD beamformer.

Hence, one of the most important aspects in practice is how to compromise between $\mathcal{W}[\mathbf{h}(\omega)]$ and $\mathcal{D}[\mathbf{h}(\omega)]$. Ideally, we would like $\mathcal{D}[\mathbf{h}(\omega)]$ to be as large as possible with $\mathcal{W}[\mathbf{h}(\omega)] \geq 1$. To achieve this goal, the authors in [1, 3] proposed to maximize the DF, subject to a constraint on the WNG. Using the distortionless constraint, we find the robust superdirective beamformer:

$$\mathbf{h}_{\text{R},\epsilon}(\omega) = \frac{[\epsilon\mathbf{I}_M + \mathbf{\Gamma}_{\text{d}}(\omega)]^{-1}\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)[\epsilon\mathbf{I}_M + \mathbf{\Gamma}_{\text{d}}(\omega)]^{-1}\mathbf{d}(\omega)}, \tag{9}$$

where $\epsilon \geq 0$ is a Lagrange multiplier. Note that (9) is a regularized (or robust) version of (8), where ϵ can be seen as the regularization parameter. This parameter aims to compromise between supergain and white noise amplification. A small ϵ leads to a large DF and a low WNG, while a large ϵ yields low DF and large WNG. Two interesting cases of (9) are $\mathbf{h}_{\text{R},0}(\omega) = \mathbf{h}_{\text{SD}}(\omega)$ and $\mathbf{h}_{\text{R},\infty}(\omega) = \mathbf{h}_{\text{DS}}(\omega)$. While $\mathbf{h}_{\text{R},\epsilon}(\omega)$ has some control on white noise amplification, it is certainly not easy to find a closed-form expression for ϵ, given a desired value of the WNG.

4. NEW NOISE FIELD AND PROPOSED BEAMFORMER

We assume that the sensed signal is corrupted both by some additive diffuse noise and by some additive white noise.

Therefore, the input SNR is now

$$\text{iSNR}(\omega) = \frac{\text{tr}\left[\phi_X(\omega)\,\mathbf{d}(\omega)\,\mathbf{d}^H(\omega)\right]}{\text{tr}\left[\phi_{\text{d}}(\omega)\,\mathbf{\Gamma}_{\text{d}}(\omega) + \phi_{\text{w}}(\omega)\,\mathbf{I}_M\right]} = \\ = \frac{\phi_X(\omega)}{\phi_{\text{d}}(\omega) + \phi_{\text{w}}(\omega)},$$

where $\text{tr}[\cdot]$ denotes the trace of a square matrix, and $\phi_{\text{d}}(\omega)$ and $\phi_{\text{w}}(\omega)$ are the variances of the diffuse and white noises, respectively. We deduce that the output SNR is

$$\text{oSNR}[\mathbf{h}(\omega)] = \\ \frac{\phi_X(\omega)\left|\mathbf{h}^H(\omega)\,\mathbf{d}(\omega)\right|^2}{\phi_{\text{d}}(\omega)\,\mathbf{h}^H(\omega)\,\mathbf{\Gamma}_{\text{d}}(\omega)\,\mathbf{h}(\omega) + \phi_{\text{w}}(\omega)\,\mathbf{h}^H(\omega)\,\mathbf{h}(\omega)},$$

As a result, the gain in SNR is

$$\mathcal{G}[\mathbf{h}(\omega)] = \\ \frac{\left|\mathbf{h}^H(\omega)\,\mathbf{d}(\omega)\right|^2}{[1-\alpha(\omega)]\,\mathbf{h}^H(\omega)\,\mathbf{\Gamma}_{\text{d}}(\omega)\,\mathbf{h}(\omega) + \alpha(\omega)\mathbf{h}^H(\omega)\,\mathbf{h}(\omega)}, \tag{10}$$

where $\alpha(\omega) = \frac{\phi_{\text{w}}(\omega)}{\phi_{\text{d}}(\omega)+\phi_{\text{w}}(\omega)}$, with $0 \le \alpha(\omega) \le 1$. It is easy to check that the beamformer that maximizes $\mathcal{G}[\mathbf{h}(\omega)]$ is

$$\mathbf{h}_\alpha(\omega) = \frac{\mathbf{\Gamma}_{\text{d},\alpha}^{-1}(\omega)\,\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)\,\mathbf{\Gamma}_{\text{d},\alpha}^{-1}(\omega)\,\mathbf{d}(\omega)}, \tag{11}$$

where $\mathbf{\Gamma}_{\text{d},\alpha}(\omega) = [1-\alpha(\omega)]\,\mathbf{\Gamma}_{\text{d}}(\omega) + \alpha(\omega)\mathbf{I}_M$. Then, the maximum gain in SNR is

$$\mathcal{G}[\mathbf{h}_\alpha(\omega)] = \mathbf{d}^H(\omega)\,\mathbf{\Gamma}_{\text{d},\alpha}^{-1}(\omega)\,\mathbf{d}(\omega)\,. \tag{12}$$

The problem is that $\phi_{\text{d}}(\omega)$ and $\phi_{\text{w}}(\omega)$ are not known in practice. In fact, we can express (11) as (9) with a frequency dependent regularizer $\epsilon(\omega) = \alpha(\omega)/(1-\alpha(\omega))$, showing that our beamformer is equivalent to (9). However, our robust superdirective beamformer (11) is preferred for two reasons. First, $\alpha(\omega)$ varies only from 0 to 1 while ϵ in (9) varies from 0 to ∞. The second reason is the simple dependence between the gain and $\alpha(\omega)$, which allows us to efficiently find the appropriate $\alpha(\omega)$ values, as will be shown later. Finding the value of $\alpha(\omega)$ that corresponds to a fixed gain of $\mathcal{G}_0$ ($M \le \mathcal{G}_0 \le M^2$) can be expressed using the following optimization problem:

$$\min_{\alpha}\left|\mathbf{d}^H(\omega)\,\mathbf{\Gamma}_{\text{d},\alpha}^{-1}(\omega)\,\mathbf{d}(\omega) - \mathcal{G}_0\right| \ \text{s. t. } 0 \le \alpha \le 1\,. \tag{13}$$

From simulations not presented here due to space limitation, it can be seen that the gain is continuous and has a single minimum point in the range $\alpha \in [0,1]$, denoted here as $\alpha_{\min}(\omega)$. The gain will monotonically decrease in the range $[0,\,\alpha_{\min}(\omega)]$ and monotonically increase in the range $[\alpha_{\min}(\omega),\,1]$. This property enables us to calculate α simply by conducting a binary-like search for each monotonic section. This method is described in Algorithm 1, which numerically solves (13), i.e., finds α for which the beamformer's SNR gain is closest to $\mathcal{G}_0$ for each frequency independently.

Algorithm 1 MAS - Minimize and Search

Input: Desired gain $\mathcal{G}_0$, and tolerance; **Output**: Optimal regularization α

1: Find $\alpha_{\min}$ that minimizes the gain (e.g., using gradient descent).
2: Divide the range $[0,1]$ into 2 sections in which the gain is monotonic: $[0,\alpha_{\min}]$ and $[\alpha_{\min},1]$.
3: For each section, apply the following continuous binary search:
4: Divide the section into 2 sub-sections.
5: Calculate the gain $\mathcal{G}_k$ in the middle of each sub-section.
6: Choose the gain $\mathcal{G}_k$ and its respective sub-section for which $|\mathcal{G}_k - \mathcal{G}_0|$ is minimal
7: **if** $|\mathcal{G}_k - \mathcal{G}_0| \le$ tolerance **then**
8: $\quad\alpha \leftarrow$(middle of chosen sub-section) and stop.
9: **else**
10: $\quad$update range to be the chosen sub-section and go back to 4
11: **end if**
12: Compare results from $[0,\alpha_{\min}]$ and $[\alpha_{\min},1]$ and choose the best result.

This approach can be used to constrain and optimize other gain properties as well. Instead of fixing the SNR gain, many applications require maximizing it while fixing the WNG or the DF. Since both the WNG and DF are monotonic in $\alpha \in [0,1]$, as can also be seen in simulations, Algorithm 1 can be used here as well. The computational complexity of the binary-like search is $\mathcal{O}\left\{|\omega|\log_2\left[\left(M^2-M\right)/\sigma\right]\right\}$, where $\sigma > 0$ is the acceptable tolerance from the desired gain. This is the only step necessary for a fixed WNG/DF. When fixing the SNR gain, we need to add the complexity of finding the initial minimum point, e.g., using gradient descent method with exact line search which requires $\mathcal{O}\left\{\log(1/\epsilon)\right\}$ iterations to converge up to tolerance $\epsilon > 0$ [14].

5. SIMULATION RESULTS

We simulated the proposed robust superdirective beamformer (11) for several different gain values, where the regularization parameter $\alpha(\omega)$ was found using Algorithm 1. All the presented simulations were performed for a linear microphone array, with $M = 8$ microphones and $\delta = 1$ cm. However, the results are general, and can be repeated for other configurations. In Figure 1(a)–(c) we show the SNR gain alongside the DF and WNG of the proposed beamformer, when set to a constant desired gain level. It can be seen that the gain value can be set as desired within the appropriate range. Although the algorithm converges for every frequency in the range, the desired gain is not always reached at very low frequencies.

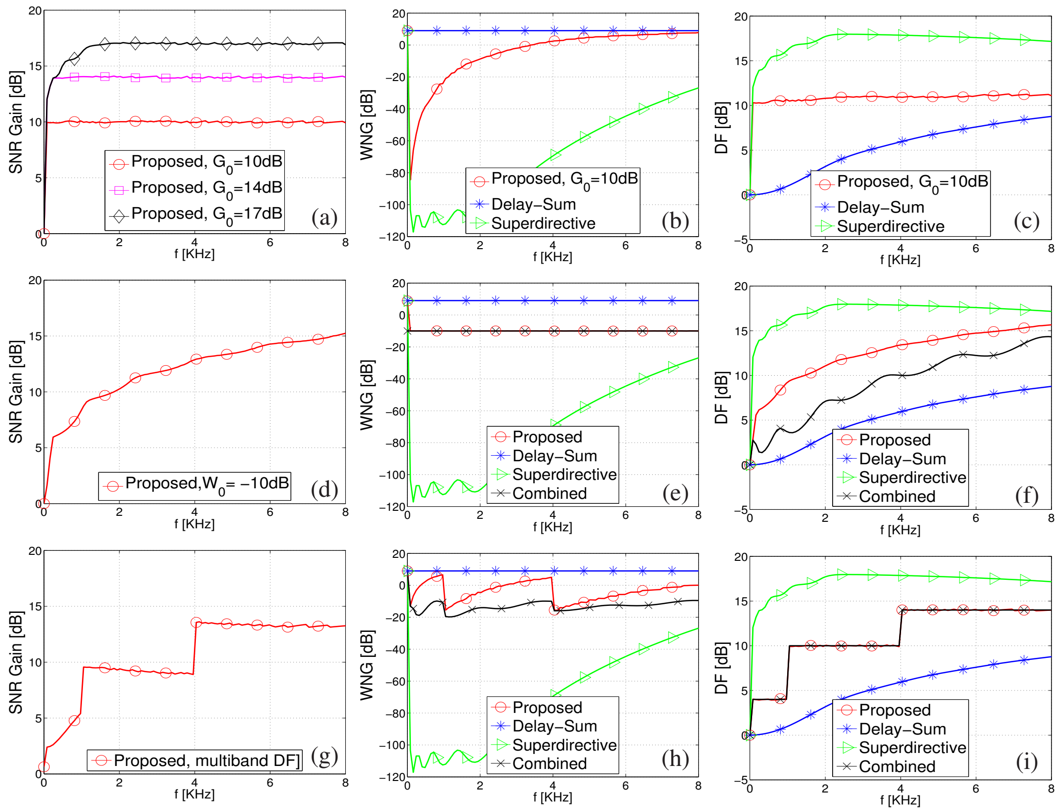

Fig. 1: Array gains of the proposed beamformer for three cases; fixed SNR gain (a)-(c), fixed WNG (d)-(f) and fixed DF in multi-bands (g)-(i). All three cases are compared to the DS and superdirective beamformers, the latter two are compared also to the combined beamformer with $\epsilon = 10^{-4}$. (a) SNR gain, (b) WNG and (c) DF for fixed SNR gain. (d) SNR gain, (e) WNG and (f) DF with desired WNG set to -10 dB. (h) SNR gain, (g) WNG and (i) DF with desired DF gain set to 4, 10, and 14 dB in multi-bands.

This is due to the constant regularization (of 10^{-14}), which is added to avoid singularity issues while inverting $\mathbf{\Gamma}_{\mathrm{d}}$ at low frequencies. In those low frequencies, the constant regularization is more dominant than $\alpha(\omega)$, hence the desired gain is not reached. While achieving the fixed SNR gain under the combined noise field, the proposed beamformer also performs well under diffuse noise. However, white noise may still be amplified to intolerable levels.

To continue and improve the WNG, a modified optimization problem can be defined, which maximizes the SNR gain under a constant WNG. As depicted in Figure 1(d)–(f), our approach yields an accurate solution for this scenario as well (using Algorithm 1 from step 4). Furthermore, it can be seen that the proposed beamformer outperforms the combined beamformer [7] with $\epsilon = 10^{-4}$. That is, the proposed beamformer has a higher DF for a fixed WNG given a similar setup. Taking this approach one step further, we design a multi-band fixed beamformer. This way, we can constrain the DF to be piece-wise constant gradually increasing in steps, thus considering the WNG-DF trade-off at each frequency band separately. The proposed approach yields accurate results both for the fixed bands and transition areas, as can be seen in Figure 1(g)–(i). A similar analysis can be done to design a multi-band fixed WNG beamformer.

6. CONCLUSION

We have introduced an optimal robust beamformer and a computationally efficient algorithm for finding its regularization parameter. We showed that our approach facilitates the design of beamformers with fixed SNR gain, beamformers with maximal SNR gain for constant WNG or DF, and multi-band fixed beamformers. The proposed design method enables a fine tuning of the compromise between the DF and robustness against white noise.

Several issues should be further investigated. The proposed beamformer should be tested for various angles of incidence, and not only in the end-fire direction. Also, it may be useful to incorporate additional considerations into the design process, such as side-lobe requirements and performance under other types of noise fields.

7. REFERENCES

[1] H. Cox, R. M. Zeskind, and M. M. Owen, “Robust adaptive beamforming,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer Science & Business Media, 2013.

[3] H. Cox, R. M. Zeskind, and T. Kooij, “Practical supergain,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, no. 3, pp. 393–398, 1986.

[4] J. Li, P. Stoica, and Z. Wang, “On robust capon beamforming and diagonal loading,” *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1702–1715, 2003.

[5] S. A. Vorobyov, A. B. Gershman, and Z. Q. Lou, “Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem,” *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313–324, 2003.

[6] S. Doclo and M. Moonen, “Superdirective beamforming robust against microphone mismatch,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 617–631, 2007.

[7] R. Berkun, I. Cohen, and J. Benesty, “Combined beamformers for robust broadband regularized superdirective beamforming,” *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 23, no. 5, pp. 877–886, 2015.

[8] R. Berkun, I. Cohen, and J. Benesty, “A tunable beamformer for robust superdirective beamforming,” *in Proc. International Workshop on Acoustic Signal Enhancement*, 2016.

[9] G. W. Elko and J. Meyer, “Microphone arrays,” in *Springer Handbook of Speech Processing*, pp. 1021–1041. Springer, 2008.

[10] J. Benesty and C. Jingdong, *Study and Design of Differential Microphone Arrays*, vol. 6, Springer Science & Business Media, 2012.

[11] J. Benesty, C. Jingdong, and Y. Huang, *Microphone Array Signal Processing*, vol. 1, Springer Science & Business Media, 2008.

[12] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[13] R. T. Lacoss, “Data adaptive spectral analysis methods,” *Geophysics*, vol. 36, no. 4, pp. 661–675, 1971.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.