# Frequency-Invariant Sensor Selection for MVDR Beamforming in Wireless Acoustic Sensor Networks

Jie Zhang, *Member, IEEE*, Guanghui Zhang, Lirong Dai

***Abstract*—Wireless acoustic sensor network (WASN) has a wide range of applications in internet of things, where signal estimation is one of the network design objectives. Due to the existence of ambient noises, the recorded audio signals are inevitably corrupted, resulting in a low signal-to-noise ratio (SNR), which triggers the necessity of signal enhancement. As using all sensor measurements brings a large amount of data transmissions and computational cost, the narrowband sensor selection was proposed to choose an informative subset of sensors to perform noise reduction in the audio context. However, the resulting frequency-dependent selection status has to be switched across frequencies. In order to avoid the complicated switching operations, we consider frequency-invariant sensor selection in this work. We propose to minimize the total power consumption over the WASN by constraining the broadband SNR, which can be solved using broadband semi-definite optimization (BroadOpt) or narrowband voting (NaVo) approaches. In order to further reduce the time complexity, we propose two near-optimal greedy methods, including gradient removal (GradR) and weighted input SNR removal (SnrR). As comparison, we also show a broadband energy removal (EnergyR) method. The greedy methods remove one sensor at each iteration from the complete network until the performance constraint is not satisfied. Numerical results using a simulated large-scale WASN show that the greedy methods can achieve a comparable performance compared to the optimization based counterparts, while the corresponding time complexity is much lower. In general, the sensors around the target source and the fusion center are more likely to be selected.**

***Index Terms*—Sensor selection, MVDR beamforming, noise reduction, convex optimization, greedy removal, energy efficiency, internet of audio things, wireless acoustic sensor networks.**

## I. Introduction

RECENTLY, wireless portable devices (e.g., smartphone, laptop) become often-used in our daily life owing to the advanced micro-electronics. Usually, each device has a small microphone array equipped, which can then monitor the acoustic scene. In principle, we are surrounded by a wireless acoustic sensor network (WASN), or the so-called internet of audio things (IoAuT) [1], which is an emerging research field positioned at the intersection of the internet of things, audio processing, signal estimation, artificial intelligence, and human-machine interaction. The WASN refers to the networks of computing devices embedded in physical objects (i.e., audio things) dedicated to the production, reception, analysis, and understanding of audio in distributed environments. The wireless nodes are connected by an infrastructure that enables multidirectional communication, both locally and remotely [2]. Regarding the network topology, the WASN can be organized in a centralized fashion by using a fusion center (FC) or in a distributed fashion [3]. The information exchange occurs between two neighbouring nodes or between nodes and the FC, and the user can thus control the devices remotely. WASNs have a broad range of applications for e.g., amateur drone surveillance [4], binaural hearing aids [5], source localization [6], vehicle classification [7], environmental noise monitoring in smart cities [8], to list a few.

In practice, the recorded audio signals are degraded inevitably due to the existence of noises (e.g., competing speakers, ambient noise, sensor self-noise, reverberation), which heavily affects the speech quality as well as the speech interaction performance. In order to improve the performance, usually speech enhancement (or noise reduction) is involved as a front-end speech processing step [9]. Compared to the conventional microphone array based speech enhancement systems [10], the utilization of WASNs for speech processing can potentially bring several benefits. For instance, as the wireless devices can be distributed anywhere, they might be very close to the target source, resulting in high-quality recordings. For hearing-aid applications, although only a small microphone array is equipped locally, if external wireless devices can share their measurements, more data is then available at binaural hearing aids, leading to a performance improvement [5].

### A. Motivation and related works

In general, the more sensor measurements that are included for noise reduction in the context of WASN or conventional microphone array [10], [11], the better the performance, but the higher the power consumption and computational complexity. As the multi-microphone recordings are highly correlated, some sensor measurements might be even redundant. In case a sensor is distant from the target speaker, the corresponding audio stream is of low quality in terms of signal-to-noise ratio (SNR), which will have a marginal contribution to noise reduction. In addition, the power consumption has to be taken into account, particularly in WASNs, as each sensor is driven with a limited amount of battery resource and the energy consumption directly affects the network lifetime. Intuitively,

This work is supported by the National Natural Science Foundation of China (62101523), Hefei Municipal Natural Science Foundation (2022012) and Fundamental Research Funds for the Central Universities. (Corresponding author: Guanghui Zhang)

J. Zhang and L. Dai are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China (USTC), 230026 Hefei, China. He is also with State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, 100190 Beijing, China. (e-mail: jzhang6@ustc.edu.cn, lrdai@ustc.edu.cn).

G. Zhang is with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong and Centre for Advances in Reliability and Safety (CAiRS), Pak Shek Kok, NT, Hong Kong. (e-mail: ghzhang@link.cuhk.edu.hk)

the sensors that have a marginal contribution can be excluded from the complete network, such that the power consumption can be saved at the cost of a tiny performance sacrifice. As both the noise reduction performance and energy consumption are related to the number of sensors, which thus somehow controls the trade-off between two metrics, the challenge in the context of noise reduction over WASNs that has to be addressed becomes: *given an expected noise reduction performance, how to optimally select the most informative subset of sensors from a large-scale WASN?*

It was shown in [12] that sensor selection is effective for reducing the cardinality of sensors for computation at the cost of a controllable performance sacrifice. Mathematically, it was formulated by optimizing a certain performance measure and constraining the number of the selected sensors, or in the other way around. In principle, this is a combinatorial optimization problem. In order to perform sensor selection efficiently, convex relaxation techniques [13] or greedy heuristics (e.g., submodularity) [14] can be leveraged. In literature, sensor selection techniques have already been applied into, e.g., state estimation [15], spectrum sensing [16], field estimation [17], time-delay of arrival based source localization [18], [19], target tracking [20], [21], speech enhancement [11], [22], [23], robotic systems [24]. Note that in the context of communications (e.g., multiple input multiple output (MIMO) cognitive radio networks), sensor selection might be called antenna selection [25]–[28] (or more generally resource allocation [29]), yet the optimization criteria are similar. By only incorporating the selected sensor subset, the resource consumption can thus be saved significantly compared to blindly using all sensor measurements, as many non-informative sensors are excluded.

In order to reconstruct an energy-efficient WASN as well as to achieve a desired noise reduction performance, in [23] a sparsity-promoting sensor selection based minimum variance distortionless response (MVDR) beamformer was proposed. The optimal sensor subset is solved by considering a semi-definite programming (SDP) problem, which is derived from minimizing the total power consumption in terms of the selection status of sensors and constraining the narrowband output noise power. The MVDR beamformer is then designed using the selected sensor subset, such that the number of sensors used for data transmission and computation is reduced. We call this formulation as narrowband sensor selection (NSS) in this work. The NSS problem was further extended in [11] by taking the effects of the estimation error of relative acoustic transfer function (RTF) into account. Since the NSS problem is considered separately for each frequency bin, the sensor selection results across different frequencies might be different. Such a frequency-variant heterogeneous selection method might cause many switching operations for sensors. For example, in case a sensor is selected in one frequency bin, while it is not selected in another frequency bin, we have to switch off this sensor for energy saving, as keeping a sensor activated also consumes a certain amount of battery resource. Also, the switching-on operations might occur similarly. From the perspective of network complexity, these switching operations would bring some extra message passing on the basis of certain communication protocols. To avoid complex switching operations, in this work we therefore propose a frequency-invariant sensor selection (FISS) approach for MVDR beamforming based noise reduction in large-scale WASNs.

### *B. Contribution*

The proposed FISS problem minimizes the total transmission power over the network subject to a constraint on the broadband output SNR, which is a joint optimization problem in terms of the spatial filter and the selection variable. By constraining the broadband performance, the proposed FISS approach therefore has only one optimization problem and the resulting sensor selection solution is unique. However, given $F$ frequency bins, the NSS method in [23] separately considers $F$ optimization problems and each has an independent solution. Therefore, compared to [23] the contribution of this work is that two broadband sensor selection methods together with two low-complexity approaches are proposed.

At first, we show that the MVDR beamformer is a solution to the FISS problem, which is thus simplified by substituting the MVDR filter as an optimization problem in terms of the selection variable. Due to the fact that the broadband SNR constraint is non-convex, we separate it into multiple constraints on the narrowband output SNR, such that the FISS problem can be relaxed as minimizing the total power consumption subject to multiple constraints on the narrowband SNR. We show that each narrowband constraint can be formulated as a linear matrix inequality (LMI), such that an SDP surrogate is obtained, which is called *BroadOpt* method in this work. In case all narrowband constraints are satisfied, it is sufficient that the expected global broadband SNR can be achieved.

Second, as the proposed BroadOpt method involves multiple LMI constraints, which heavily affects the computational efficiency, we consider to optimize multiple local SDP problems, and each is constrained by the narrowband SNR. Due to the fact that the selection solutions vary across frequencies, in order to resolve a unique status we design a narrowband voting (*NaVo*) approach, which can be regarded as an extension of the NSS method in [23] by averaging the frequency-dependent selections. Note that both BroadOpt and NaVo have a cubic time complexity in terms of the number of sensors.

Third, in order to reduce the computational complexity, we propose two low-complexity greedy approaches: *GradR* and *SnrR*. One is based on the use of the gradient of FISS problem, and the other is on the weighted input SNR. As the gradient somehow measures the contribution of a sensor to minimizing the objective function, we initialize the selected subset using the complete network and remove the sensor that has the smallest gradient with respect to the current selected subset. Alternatively, the input broadband SNR weighted by the individual transmission cost is further used to measure the contribution to the considered energy-aware noise reduction performance. Based on the weighted SNR, we can thus design an *SnrR* method, which removes the sensor that has the smallest weighted SNR at each iteration. The proposed greedy sensor selection approaches follow a similar removal criterion in [30], where however the power consumption is not taken into account. As comparison, we show that the weighted broadband input energy can also be utilized to perform sensor selection, which is termed as *EnergyR* in this work.

Numerical simulations using a large-scale WASN show that the proposed methods can satisfy the performance requirement with a much smaller sensor subset compared to the complete network, which reveals that in practice many sensors are non-informative for energy-aware noise reduction. The sensors that are close to the target source and the FC are more informative, as they are beneficial for enhancing the target signal and saving the power consumption, respectively. It is shown that the proposed greedy approaches have an obvious superiority in computational efficiency over the optimization based methods.

### C. Outline and Notation

The remainder of this paper is structured as follows. Section II introduces the signal model and problem description. In Section III, we present the proposed broadband FISS methods, i.e., *BroadOpt* and *NaVo*. In Section IV, we propose three greedy sensor removal approaches, i.e., *GradR*, *SnrR* and *EnergyR*. Section V presents the experimental results using a simulated WASN. Finally, Section VI concludes this work.

*Notation:* The notation used in this paper is as follows: Upper (lower) bold face letters are used for matrices (column vectors). $(\cdot)^T$ or $(\cdot)^H$ denotes vector/matrix transposition or conjugate transposition. $\mathbb{E}(\cdot)$ denotes the mathematical expectation operation. $\mathrm{diag}(\cdot)$ refers to a block diagonal matrix with the elements in its argument on the main diagonal. $\mathbf{O}$ and $\mathbf{1}$ denote an all-zeros square matrix and an all-ones column vector of proper size, respectively. $\mathbf{I}_N$ is an identity matrix of size $N$. $\mathbf{e}_k$ denotes an $M$-dimensional column vector with the $k$th element equal to one and zeros elsewhere. $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix. $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$.

## II. Signal model and problem formulation

### A. Signal model

Given a WASN consisting of $M$ spatially distributed acoustic sensor nodes, we assume that the FC is one of the nodes and each node is equipped with a single microphone without loss of generality. In the short-time Fourier transform (STFT) domain, let $l$ and $\omega$, respectively, denote the frame index and the angular frequency index. The noisy signal $Y_k(\omega, l)$ recorded by the $k$th microphone is written as

$$Y_k(\omega, l) = X_k(\omega, l) + N_k(\omega, l),\ k = 1, \dots, M, \tag{1}$$

where $X_k(\omega, l)$ denotes the target signal component at microphone $k$, and $N_k(\omega, l)$ the corresponding additive noise component, which might include coherent noise sources (e.g., competing speakers) and incoherent noises (e.g., late reverberations, sensor self-noise). For the single target source case, the signal component can be written as

$$X_k(\omega, l) = a_k(\omega) S(\omega, l), \tag{2}$$

where $a_k(\omega)$ denotes the acoustic transfer function (ATF) of the target source with respect to the $k$th microphone node, and $S(\omega, l)$ the target signal at the source position. Without loss of generality, we take the FC as the first node and assign it as the reference (a more general and sophisticated reference microphone selection method can be found in [31]), such that the RTF can be defined as

$$h_k(\omega) = a_k(\omega)/a_1(\omega), \tag{3}$$

which is the normalized ATF with respect to the reference. Both the ATF and RTF are time-invariant under the assumption that the target source keeps static during the observation period. The introduction of RTF is due to the fact that in practice the RTF can be estimated using covariance subtraction or covariance whitening method [32]. The ATF is a scaled version of RTF, and the scaling factor makes the estimation of ATF ambiguous. More importantly, the utilization of RTF does not degrade the beamforming performance. In case of the utilization of ATF, the classic MVDR beamformer returns an estimate of the original target source, i.e., $\hat{S}(\omega, l)$; in case of the usage of RTF, we can obtain an estimated signal component at the reference microphone, i.e., $\hat{X}_1(\omega, l)$. In both cases, the output SNRs are equal. For notational brevity, we will omit the time-frame index in the sequel.

Let $\mathbf{y}(\omega) = [Y_1(\omega, l), Y_2(\omega, l), \dots, Y_M(\omega, l)]^T$, which stacks the sensor measurements for each time-frequency bin. Similarly, we define vectors $\mathbf{n}(\omega)$, $\mathbf{a}(\omega)$, $\mathbf{h}(\omega)$ and $\mathbf{x}(\omega)$ for stacking the noise components, ATFs, RTFs and signal components, respectively, such that the signal model can be compactly given by

$$\mathbf{y}(\omega) = \mathbf{a}(\omega) S(\omega) + \mathbf{n}(\omega) = \mathbf{h}(\omega) X_1(\omega) + \mathbf{n}(\omega). \tag{4}$$

Further, we define the noise and noisy covariance matrices as

$$\mathbf{\Phi}_{\mathbf{nn}}(\omega) = \mathbb{E}\{\mathbf{n}(\omega)\mathbf{n}(\omega)^H\}, \quad \mathbf{\Phi}_{\mathbf{yy}}(\omega) = \mathbb{E}\{\mathbf{y}(\omega)\mathbf{y}(\omega)^H\},$$

where the expectation is taken over time frames. Applying a voice activity detector (VAD), the microphone signal can be categorized as speech-absent and speech-present frames, and during these two periods the noise and noisy covariance matrices can be estimated using the average smoothing technique. We assume that the signal component and noise components are mutually uncorrelated, such that $\mathbf{\Phi}_{\mathbf{yy}}(\omega) = \mathbf{\Phi}_{\mathbf{xx}}(\omega) + \mathbf{\Phi}_{\mathbf{nn}}(\omega)$ and $\mathbf{\Phi}_{\mathbf{xx}}(\omega)$ can be estimated by

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{xx}}(\omega) &\triangleq \sigma_S^2(\omega)\mathbf{a}(\omega)\mathbf{a}(\omega)^H \triangleq \sigma_{X_1}^2(\omega)\mathbf{h}(\omega)\mathbf{h}(\omega)^H && (5) \\ &= \mathbf{\Phi}_{\mathbf{yy}}(\omega) - \mathbf{\Phi}_{\mathbf{nn}}(\omega), && (6) \end{aligned}$$

where $\sigma_S^2(\omega) = \mathbb{E}\{|S(\omega, l)|^2\}$ and $\sigma_{X_1}^2(\omega) = \mathbb{E}\{|X_1(\omega, l)|^2\}$ denote the power spectral densities (PSD) of the target signal at the source position and at the reference position, respectively.

In this work, we aim at estimating the target signal using a subset of sensor measurements, which is chosen by optimizing the transmission energy between the selected sensors and the FC subject to an expected estimation performance bound.

### B. Problem description

In order to formulate the FISS problem, it is necessary to define a selection vector

$$\mathbf{p} = [p_1, p_2, \dots, p_M]^T \in \{0, 1\}^M, \tag{7}$$

where $p_k = 1, \forall k$ indicates that the $k$th microphone is selected, and otherwise unselected. With the vector $\mathbf{p}$, we can find the selected subset of sensors as $\mathcal{S}_{\text{in}} = \{k|p_k = 1\}$ and the unselected subset as $\mathcal{S}_{\text{ex}} = \{k|p_k = 0\}$. The subscripts “in” and “ex” represent inclusion and exclusion, respectively. Further, we use $K = |\mathcal{S}_{\text{in}}|$ to represent the number of the selected sensors, resulting in $|\mathcal{S}_{\text{ex}}| = M - K$. Throughout this work, in order to simplify the sensor switching operations, we will consider $\mathbf{p}$ as frequency-invariant, i.e., the selection status of any sensor keeps the same across frequencies.

Given the selected sensors, we can design a spatial frequency-dependent filter $\mathbf{w}_{\mathbf{p}}(\omega) \in \mathbb{C}^K$, e.g., MVDR, and the target signal is then estimated via beamforming as

$$\hat{X}_1(\omega) = \mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{y}_{\mathbf{p}}(\omega), \tag{8}$$

where $\mathbf{y}_{\mathbf{p}}(\omega)$ contains the sensor measurements of the selected subset. The joint frequency and selection dependent narrowband output SNR can then be calculated as

$$\text{oSNR}_{\mathbf{p}}(\omega) = \frac{\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{xx},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)}{\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)}, \tag{9}$$

where $\mathbf{\Phi}_{\mathbf{xx},\mathbf{p}}(\omega)$ and $\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)$ denote the speech covariance matrix and the noise covariance matrix of the selected sensor measurements.

In a large-scale WASN, the energy consumption should be taken into account for the design of data processing algorithms, as it directly influences the network lifetime and the efficiency of data aggregation. Let $C_k, k \in \mathcal{M} = \{1, 2, \ldots, M\}$ denote the energy consumption of sensor $k$, which is composed of two operations: 1) the power for keeping it active, and 2) the power for transmitting its measurements to the FC, which was shown to be proportional to the squared transmission distance [33]–[35]. In order to improve the energy efficiency, in this work we therefore intend to minimize the total energy consumption over the WASN subject to a constraint on the desired noise reduction performance, which can be formulated as the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{p}\in\{0,1\}^M, \mathbf{w}_{\mathbf{p}}} \quad & \sum_{k=1}^{M} p_k C_k \\ \text{s.t.} \quad & \overline{\text{oSNR}}_{\mathbf{p}} \geq \alpha \overline{\text{oSNR}}_{\max} \\ & \mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{h}_{\mathbf{p}}(\omega) = 1, \end{aligned} \tag{10}$$

where $\overline{\text{oSNR}}_{\max}$ denotes the maximum output broadband SNR, which represents the case when all sensors are used[1], and $0 < \alpha \leq 1$ controls the expected output SNR. In practice, usually $\overline{\text{oSNR}}_{\max}$ is unavailable, as the total number of sensors might be even unknown in large-scale WASNs. In this case, $\alpha \overline{\text{oSNR}}_{\max}$ can be assigned by users as an arbitrary desired performance (e.g., 40 dB). In (10), $\overline{\text{oSNR}}_{\mathbf{p}}$ denotes the output broadband SNR using the selected subset of sensors that is determined by $\mathbf{p}$, which is given by

$$\overline{\text{oSNR}}_{\mathbf{p}} = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \frac{\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{xx},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)}{\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)}, \tag{11}$$

with $\Omega$ the number of frequency bins. Note that the equality $\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{h}_{\mathbf{p}}(\omega) = 1$ is included as a distortionless constraint on the target source, since

$$\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{xx},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega) = \sigma_{X_1}^2(\omega), \ \forall \omega, \tag{12}$$

where $\mathbf{h}_{\mathbf{p}}(\omega)$ denotes the RTF with respect to the selected sensors. It is clear that (10) is a non-convex combinatorial optimization problem, because of the non-linear selection operation and the Boolean constraints on $\mathbf{p}$. The difference between the proposed problem formulation in (10) and [23] is twofold: 1) the noise reduction performance is constrained on the broadband output SNR rather than the narrowband output noise power, and 2) the selection variable $\mathbf{p}$ is frequency-invariant rather than frequency-dependent.

## III. Broadband sensor selection

In this section, we will theoretically analyze the proposed general sensor selection problem in (10) for noise reduction in WASNs and propose two optimization-based approaches.

### *A. Dimensionality reduction for unknowns using MVDR*

First of all, in order to observe the optimality of (10) on the selection variable $\mathbf{p}$ and the spatial filter $\mathbf{w}_{\mathbf{p}}(\omega)$, we consider the Lagrangian function of (10), which is given by

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \mathbf{w}_{\mathbf{p}}(1), \ldots, \mathbf{w}_{\mathbf{p}}(\Omega), \lambda, \mu_1, \ldots, \mu_\Omega) &= \sum_{k=1}^{M} p_k C_k \\ + \lambda \left( \alpha \text{oSNR}_{\max} - \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \frac{\sigma_{X_1}^2(\omega)}{\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)} \right) & \\ + \sum_{\omega=1}^{\Omega} \mu_\omega \left( \mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{h}_{\mathbf{p}}(\omega) - 1 \right), & \end{aligned}$$

where $\lambda \geq 0$ and $\mu_\omega \geq 0, \omega = 1, \ldots, \Omega$ are the Lagrange multipliers associated with the inequality and equality constraints, respectively. Note that both $\lambda$ and $\mu_\omega$ are real. The gradient with respect to the conjugate of $\mathbf{w}_{\mathbf{p}}(\omega)$ is given by

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_{\mathbf{p}}^*(\omega)} = \frac{\lambda \sigma_{X_1}^2(\omega) \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)}{\Omega \left(\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)\right)^2} + \mu_\omega \mathbf{h}_{\mathbf{p}}(\omega). \tag{13}$$

Setting $\frac{\partial \mathcal{L}}{\partial \mathbf{w}_{\mathbf{p}}^*(\omega)}$ to zero, it can be seen that

$$\begin{aligned} \mathbf{w}_{\mathbf{p}}(\omega) = &- \frac{\Omega \left(\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)\right)^2 \mu_\omega}{\lambda \sigma_{X_1}^2(\omega)} \\ &\times \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega), \forall \omega. \end{aligned} \tag{14}$$

Substituting $\mathbf{w}_{\mathbf{p}}(\omega)$ from (14) into the equality constraint $\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{h}_{\mathbf{p}}(\omega) = 1$, we can obtain

$$\frac{\mu_\omega}{\lambda} = \frac{-\frac{\sigma_{X_1}^2(\omega)}{\Omega}}{\left(\mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)\right)^2 \mathbf{h}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)}.$$

Plugging the expression of $\frac{\mu_\omega}{\lambda}$ back into (14), it holds that

$$\mathbf{w}_{\mathbf{p}}(\omega) = \frac{\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)}{\mathbf{h}_{\mathbf{p}}(\omega)^H \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)}, \forall \omega, \tag{15}$$

[1] For multi-microphone noise reduction approaches, the performance is generally positively related to the number of microphones, as the more the microphones, the better the performance. In case all sensors are involved, the optimal noise reduction performance is then obtained.

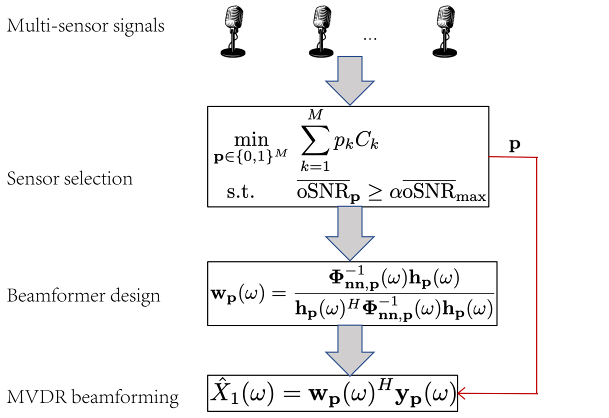

Multi-sensor signals
...
Sensor selection
$\min_{\mathbf{p}\in\{0,1\}^M} \sum_{k=1}^{M} p_k C_k$
s.t. $\overline{\text{oSNR}}_{\mathbf{p}} \geq \alpha \overline{\text{oSNR}}_{\max}$
$\mathbf{p}$
Beamformer design
$\mathbf{w}_{\mathbf{p}}(\omega) = \frac{\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)}{\mathbf{h}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)}$
MVDR beamforming
$\hat{X}_1(\omega) = \mathbf{w}_{\mathbf{p}}(\omega)^H\mathbf{y}_{\mathbf{p}}(\omega)$

Figure 1. The system model of the considered FISS based MVDR beamforming for noise reduction in large-scale WASNs.

which is exactly the close-form solution of the well-known MVDR optimization problem given by [36]–[38]

$$\begin{aligned}\mathbf{w}_{\mathbf{p}}(\omega) =& \arg\min_{\mathbf{w}_{\mathbf{p}}(\omega)} \mathbf{w}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)\mathbf{w}_{\mathbf{p}}(\omega)\\ &\text{s.t.} \quad \mathbf{w}_{\mathbf{p}}(\omega)^H\mathbf{h}_{\mathbf{p}}(\omega) = 1.\end{aligned} \tag{16}$$

Substituting the MVDR beamformer from (15) to the sensor selection dependent broadband output SNR in (11), we obtain

$$\overline{\text{oSNR}}_{\mathbf{p}} = \frac{1}{\Omega}\sum_{\omega=1}^{\Omega} \sigma_{X_1}^2(\omega)\mathbf{h}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega). \tag{17}$$

Hence, in order to avoid optimizing the beamformer variable, we can plug the MVDR beamformer from (15) into (10), leading to a simplified sensor selection problem:

$$\begin{aligned}&\min_{\mathbf{p}\in\{0,1\}^M} \sum_{k=1}^{M} p_k C_k\\ &\text{s.t.} \ \frac{1}{\Omega}\sum_{\omega=1}^{\Omega} \sigma_{X_1}^2(\omega)\mathbf{h}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega) \geq \beta,\end{aligned} \tag{18}$$

where $\beta = \alpha\overline{\text{oSNR}}_{\max}$ is defined for national conciseness. As such, we can get rid of optimizing the sensor selection variables and spatial filter jointly. The dimensionality reduction in this section will not cause any sub-optimality for (10), as it follows the standard KKT conditions [39]. On the other hand, it implies that the sensor selection based MVDR beamforming can be optimally separated into two steps: 1) sparse sensor selection, and 2) MVDR beamformer design using the selected sensors. The system model of the proposed FISS based MVDR beamforming for noise reduction is shown in Fig. 1. Note that indeed (18) is still a non-convex (combinatorial) optimization problem in nature. One simply method to solve (18) is exhaustive search, i.e., evaluating the performances of all $\binom{M}{K}$ choices in case $K$ sensors need to be selected from $M$ sensors, but evidently this is intractable unless both $M$ and $K$ are small. Next, we will focus on resolving the selection variable $\mathbf{p}$ from (18), which can then be fed into (15) for the MVDR beamformer design.

## *B. Linearization*

Second, in order to more clearly observe the dependence of the broadband output SNR on sensor selection, we will linearize $\overline{\text{oSNR}}_{\mathbf{p}}$ in terms of $\mathbf{p}$ in this section. Let $\text{diag}(\mathbf{p})$ denote a diagonal matrix, where the diagonal entries are given by $\mathbf{p}$, such that a selection matrix $\mathbf{\Sigma}_{\mathbf{p}} \in \{0,1\}^{K\times M}$ can be defined by removing the all-zero rows of $\text{diag}(\mathbf{p})$. By definition, for the selection matrix two properties hold, i.e.,

$$\mathbf{\Sigma}_{\mathbf{p}}\mathbf{\Sigma}_{\mathbf{p}}^T = \mathbf{I}_K, \quad \mathbf{\Sigma}_{\mathbf{p}}^T\mathbf{\Sigma}_{\mathbf{p}} = \text{diag}(\mathbf{p}). \tag{19}$$

Applying the selection matrix $\mathbf{\Sigma}_{\mathbf{p}}$, the selected sensor measurements can be constructed by

$$\mathbf{y}_{\mathbf{p}}(\omega) = \mathbf{\Sigma}_{\mathbf{p}}\mathbf{y}(\omega) = \mathbf{\Sigma}_{\mathbf{p}}\mathbf{x}(\omega) + \mathbf{\Sigma}_{\mathbf{p}}\mathbf{n}(\omega) \in \mathbb{C}^K. \tag{20}$$

Similarly, the RTF, the speech and noise covariance matrices for the selection sensors are, respectively, given by

$$\begin{aligned}\mathbf{h}_{\mathbf{p}}(\omega) &= \mathbf{\Sigma}_{\mathbf{p}}\mathbf{h}(\omega) \in \mathbb{C}^K,\\ \mathbf{\Phi}_{\mathbf{xx},\mathbf{p}}(\omega) &= \mathbf{\Sigma}_{\mathbf{p}}\mathbf{\Phi}_{\mathbf{xx}}(\omega)\mathbf{\Sigma}_{\mathbf{p}}^T \in \mathbb{C}^{K\times K},\\ \mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega) &= \mathbf{\Sigma}_{\mathbf{p}}\mathbf{\Phi}_{\mathbf{nn}}(\omega)\mathbf{\Sigma}_{\mathbf{p}}^T \in \mathbb{C}^{K\times K}.\end{aligned}$$

In [23], a linearization strategy was introduced for decomposing $\mathbf{h}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)$, where the noise covariance matrix $\mathbf{\Phi}_{\mathbf{nn}}(\omega)$ is decomposed as

$$\mathbf{\Phi}_{\mathbf{nn}}(\omega) = \eta_\omega\mathbf{I}_M + \mathbf{G}_\omega, \omega = 1, \ldots, \Omega, \tag{21}$$

where the constant $\eta_\omega$ is positive and $\mathbf{G}_\omega$ is a positive definite matrix. This trick is introduced for the calculation of the inverse of $\mathbf{\Phi}_{\mathbf{nn}}(\omega)$, which was similarly used in [18]–[20]. This decomposition can be easily found as long as $\eta_\omega$ is slightly smaller than the minimum eigenvalue of $\mathbf{\Phi}_{\mathbf{nn}}(\omega)$, which is always positive definite due to the presence of coherence and incoherence noises. With $\eta_\omega$ and $\mathbf{G}_\omega$ at hand, we can re-write $\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega)$ as

$$\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}(\omega) = \mathbf{\Sigma}_{\mathbf{p}}\mathbf{\Phi}_{\mathbf{nn}}(\omega)\mathbf{\Sigma}_{\mathbf{p}}^T = \eta_\omega\mathbf{I}_K + \mathbf{\Sigma}_{\mathbf{p}}\mathbf{G}_\omega\mathbf{\Sigma}_{\mathbf{p}}^T. \tag{22}$$

Consequently, $\mathbf{h}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega)$ can be derived as

$$\mathbf{h}_{\mathbf{p}}(\omega)^H\mathbf{\Phi}_{\mathbf{nn},\mathbf{p}}^{-1}(\omega)\mathbf{h}_{\mathbf{p}}(\omega) = \mathbf{h}(\omega)^H\mathbf{Q}(\omega)\mathbf{h}(\omega), \tag{23}$$

where $\mathbf{Q}(\omega)$ is given by

$$\mathbf{Q}(\omega) = \mathbf{\Sigma}_{\mathbf{p}}^T\left(\eta_\omega\mathbf{I}_K + \mathbf{\Sigma}_{\mathbf{p}}\mathbf{G}_\omega\mathbf{\Sigma}_{\mathbf{p}}^T\right)^{-1}\mathbf{\Sigma}_{\mathbf{p}} \tag{24}$$

Using the matrix inversion lemma [40]

$$\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^T = \mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1}\mathbf{A},$$

we can reformulate the matrix $\mathbf{Q}$ as

$$\mathbf{Q}(\omega) = \mathbf{G}_\omega^{-1} - \mathbf{G}_\omega^{-1}\left(\mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p})\right)^{-1}\mathbf{G}_\omega^{-1}. \tag{25}$$

As a result, the narrowband output SNR can be derived as

$$\begin{aligned}\text{oSNR}_{\mathbf{p}}(\omega) &= \sigma_{X_1}^2(\omega)\mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\mathbf{h}(\omega)\\ &- \mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\left(\mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p})\right)^{-1}\mathbf{G}_\omega^{-1}\mathbf{h}(\omega),\end{aligned} \tag{26}$$

where $\mathbf{p}$ appears only in one place instead of in three places as in (18), and which is thus called linearization.

### C. Broadband optimization (BroadOpt)

From the previous analysis, it is clear that even though the narrowband SNR can be linearized in terms of $\mathbf{p}$, the broadband output SNR as the averaged narrowband SNRs across frequencies will still include $\mathbf{p}$ for $\Omega$ times explicitly, which will still cause non-convexity. In order to satisfy the global constraint in (18) on the broadband SNR, in this section we will consider local constraints instead. Due to the fact that in case it holds that

$$\text{oSNR}_{\mathbf{p}}(\omega) \geq \beta_\omega, \omega = 1, \dots, \Omega \tag{27}$$

where $\beta_\omega = \alpha \text{oSNR}_{\max}(\omega)$ with $\text{oSNR}_{\max}(\omega)$ denoting the maximum narrowband output SNR at frequency $\omega$, it is for sure that the global constraint in (18) is satisfied, i.e.,

$$\overline{\text{oSNR}_{\mathbf{p}}} = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \text{oSNR}_{\mathbf{p}}(\omega) \geq \beta. \tag{28}$$

Using (26), the local constraints can then be re-written as

$$\begin{aligned} &\sigma_{X_1}^2(\omega)\mathbf{h}(\omega)^H \mathbf{G}_\omega^{-1}\mathbf{h}(\omega) - \beta_\omega \geq \\ &\quad \mathbf{h}(\omega)^H \mathbf{G}_\omega^{-1} \left(\mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p})\right)^{-1} \mathbf{G}_\omega^{-1}\mathbf{h}(\omega), \end{aligned} \tag{29}$$

which is called *local constraint* for frequency $\omega$ and can be further reformulated as an LMI using the Schur complement [39]

$$\begin{bmatrix} \mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p}) & \mathbf{G}_\omega^{-1}\mathbf{h}(\omega) \\ \mathbf{h}(\omega)^H \mathbf{G}_\omega^{-1} & Z_\omega - \beta_\omega \end{bmatrix} \succeq \mathbf{O}_{M+1}, \tag{30}$$

where $Z_\omega = \sigma_{X_1}^2(\omega)\mathbf{h}(\omega)^H \mathbf{G}_\omega^{-1}\mathbf{h}(\omega)$ is introduced for notational brevity, and $\mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p})$ is positive definite.

In addition, for the Boolean constraint, we use continuous surrogates for convex relaxation, i.e., relaxing $p_k \in \{0,1\}$ to $0 \leq p_k \leq 1, \forall k$. Altogether, the simplified MVDR sensor selection problem in (18) can be relaxed as

$$\begin{aligned} \min_{\mathbf{p}} \; & \sum_{k=1}^{M} p_k C_k \\ \text{s.t.} \; & \begin{bmatrix} \mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p}) & \mathbf{G}_\omega^{-1}\mathbf{h}(\omega) \\ \mathbf{h}(\omega)^H \mathbf{G}_\omega^{-1} & Z_\omega - \beta_\omega \end{bmatrix} \succeq \mathbf{O}_{M+1}, \forall \omega, \\ & 0 \leq p_k \leq 1, \forall k, \end{aligned} \tag{31}$$

which is an SDP problem and can be efficiently solved using the interior-point method [39]. The Boolean selection variable can then be resolved by randomized rounding or deterministic rounding techniques [23]. Note that (31) includes $\Omega$ LMI constraints, which might cause a rather time-consuming problem for SDP solvers. The time-complexity of the proposed BroadOpt method is thus $\mathcal{O}(\Omega M^3)$.

### D. Narrowband voting approach using local constraints

In order to avoid the high computational complexity problem in the *broadband local* method in Sec. III-C, we propose a narrowband voting approach in this section. Let $\mathbf{p}(\omega)$ denote the selection variable for frequency $\omega$. Based on (18), we consider the NSS problem for each frequency bin, e.g.,

$$\min_{\mathbf{p}(\omega) \in \{0,1\}^M} \sum_{k=1}^{M} p_k(\omega) C_k \quad \text{s.t. oSNR}_{\mathbf{p}}(\omega) \geq \beta_\omega, \tag{32}$$

which is equivalent to the NSS formulation in [23], since it can be shown that constraining the narrowband output SNR reduces to bounding the output noise power in the context of MVDR beamforming. Based on the analysis in previous sections, (32) can be reformulated as an SDP problem, i.e.,

$$\begin{aligned} \min_{\mathbf{p}(\omega)} \; & \sum_{k=1}^{M} p_k(\omega) C_k \\ \text{s.t.} \; & \begin{bmatrix} \mathbf{G}_\omega^{-1} + \eta_\omega^{-1}\text{diag}(\mathbf{p}(\omega)) & \mathbf{G}_\omega^{-1}\mathbf{h}(\omega) \\ \mathbf{h}(\omega)^H \mathbf{G}_\omega^{-1} & Z_\omega - \beta_\omega \end{bmatrix} \succeq \mathbf{O}_{M+1}, \\ & 0 \leq p_k(\omega) \leq 1, \forall k. \end{aligned} \tag{33}$$

As such, the global constraint on the broadband output SNR is split over frequencies, and we obtain $\Omega$ narrowband SDP-based sensor selection problems. However, at each frequency we only need to solve a much lower-complexity SDP problem as compared to (31). In principle, the selection status $\mathbf{p}(\omega)$ obtained by (33) might be different across frequencies.

**Algorithm 1:** Proposed narrowband voting approach

```
Initialize: $p_{\max} = 1$, $p_{\min} = 0$;
For $\omega = 1, \dots, \Omega$ do
    $\mathbf{p}(\omega) = \text{optimize}(33)$;
Broadband averaging: $\mathbf{p} = \frac{1}{\Omega}\sum_{\omega=1}^{\Omega} \mathbf{p}(\omega)$;
Bisection based threshold determination:
    For $\kappa = 1, \dots, K$ do
        $\text{th} = 0.5(p_{\min} + p_{\max})$;
        $\mathcal{S}_{\text{in}} = \{k | p_k \geq \text{th}\}$;
        If $\overline{\text{oSNR}_{\mathbf{p}}} \geq \beta$, $p_{\min} = \text{th}$;
        else $p_{\max} = \text{th}$;
Beamforming: $\hat{X}_1(\omega) = \mathbf{w}_{\mathbf{p}}(\omega)^H \mathbf{y}_{\mathbf{p}}(\omega)$;
Return $\hat{X}_1(\omega)$, $\mathcal{S}_{\text{in}}$.
```

In order to refine the narrowband selection results and obtain a frequency-invariant solution, we design a narrowband voting approach in this section. Let

$$\mathbf{p} = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \mathbf{p}(\omega) \in [0,1], \tag{34}$$

represent the normalized selection status that contains the selection probabilities of sensors. Using the probabilities in $\mathbf{p}$, we can design a bisection algorithm to find a probability threshold. The sensors whose probabilities are larger than the threshold are considered to be selected; otherwise unselected. Such a threshold can be easily determined within several iterations based on the global performance constraint. Therefore, this procedure functions as a deterministic rounding. The proposed narrowband voting (NaVo) approach is summarized in Algorithm 1. Note that for each frequency bin, the time complexity of (33) is of the order of $\mathcal{O}(M^3)$, so the time complexity of NaVo is $\mathcal{O}(\Omega M^3)$, which is the same as the broadband optimization method. It is clear that compared to the NSS method in [23], the proposed NaVo approach is extended by adding an extra normalization step, such that an FISS solution can be obtained.

## IV. Low-complexity frequency-invariant methods

In order to avoid the high computational complexity within the SDP-based methods, in this section we will propose several sub-optimal, but low-complexity FISS approaches.

### A. Gradient removal method

In order to minimize the total power consumption subject to a lower bound on the output broadband SNR in (18), we can somehow maximize an unconstrained optimization problem as

$$\max_{\mathbf{p}\in\{0,1\}^M} \quad g(\mathbf{p}) = \frac{\frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\mathrm{oSNR}_{\mathbf{p}}(\omega)}{\sum_{k=1}^{M} p_k C_k}. \tag{35}$$

The maximum can be obtained by increasing the numerator and decreasing the denominator, which approach the lower bound on the output SNR and the minimum of the objective function of (18), respectively. Based on the results from Section III-B, we can write the numerator as

$$\overline{\mathrm{oSNR}}_{\mathbf{p}} = \frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\Big(\sigma_{X_1}^2(\omega)\mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\mathbf{h}(\omega) - f_\omega(\mathbf{p})\Big), \tag{36}$$

where $f_\omega(\mathbf{p})$ is defined for notational brevity as

$$\begin{aligned} f_\omega(\mathbf{p}) &= \mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}\mathbf{G}_\omega^{-1}\mathbf{h}(\omega) \\ &= \mathrm{Tr}\left(\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}\mathbf{G}_\omega^{-1}\mathbf{h}(\omega)\mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\right), \end{aligned}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace operator for matrices. The gradient of $f_\omega(\mathbf{p})$ with respect to $p_k$ can thus be calculated as

$$\begin{aligned} \frac{\partial f_\omega(\mathbf{p})}{\partial p_k} = \mathrm{Tr}\Big(&\frac{\partial\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}}{\partial p_k} \\ &\times\mathbf{G}_\omega^{-1}\mathbf{h}(\omega)\mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\Big), \end{aligned} \tag{37}$$

Using the property of the gradient of inverse [40]

$$\frac{\partial\mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\mathbf{Y}^{-1},$$

we can obtain

$$\begin{aligned} \frac{\partial\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}}{\partial p_k} =& -\eta_\omega^{-1}\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1} \\ &\times\mathcal{I}_k\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}, \end{aligned}$$

where $\mathcal{I}_k$ is an indicator matrix with $k$th diagonal entry equal to one and zeros elsewhere. $\frac{\partial f_\omega(\mathbf{p})}{\partial p_k}$ can thus be derived as

$$\begin{aligned} \frac{\partial f_\omega(\mathbf{p})}{\partial p_k} =& -\eta_\omega^{-1}\mathbf{h}(\omega)^H\mathbf{G}_\omega^{-1}\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1} \\ &\times\mathcal{I}_k\left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}\mathbf{G}_\omega^{-1}\mathbf{h}(\omega) \\ =& -\eta_\omega^{-1}\mathbf{v}(\omega)^H\mathcal{I}_k\mathbf{v}(\omega) = -\eta_\omega^{-1}|v_k(\omega)|^2, \end{aligned} \tag{38}$$

where $v_k$ denotes the $k$th element of vector $\mathbf{v}(\omega)$, given by

$$\mathbf{v}(\omega) = \left(\mathbf{G}_\omega^{-1}+\eta_\omega^{-1}\mathrm{diag}(\mathbf{p})\right)^{-1}\mathbf{G}_\omega^{-1}\mathbf{h}(\omega). \tag{39}$$

As a consequence, the gradient of (35) with respect to $p_k$ can then be calculated as

$$\begin{aligned} \frac{\partial g(\mathbf{p})}{\partial p_k} &= \frac{-\frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\frac{\partial f_\omega(\mathbf{p})}{\partial p_k}\sum_{k=1}^{M}p_kC_k - C_k\overline{\mathrm{oSNR}}_{\mathbf{p}}}{\left(\sum_{k=1}^{M}p_kC_k\right)^2} \\ &= \left(\sum_{k=1}^{M}p_kC_k\right)^{-2}\left(\frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\frac{|v_k(\omega)|^2}{\eta_\omega}\sum_{k=1}^{M}p_kC_k - C_k\overline{\mathrm{oSNR}}_{\mathbf{p}}\right) \end{aligned} \tag{40}$$

The gradient measures how a function changes along a direction, as the larger the gradient, the faster the function changes in this direction. Motivated by this, we can design a gradient-based sensor selection approach, which is an iterative greedy method in nature. In detail, at the first iteration we initialize the selected subset as $\mathcal{S}_{\mathrm{in}} = \{1,\ldots,M\}$, set $\mathbf{p} = \mathbf{1}_M$, and intend to remove one sensor from $\mathcal{S}_{\mathrm{in}}$. By calculating the gradient using (40), such a sensor can be determined by

$$m = \arg\min\frac{\partial g(\mathbf{p})}{\partial p_k}, \quad k\in\mathcal{S}_{\mathrm{in}}, \tag{41}$$

due to the fact that the vector $\mathbf{p}$ with $p_m = 0$ and ones elsewhere represents the direction that the the objective function of (35) approaches the maximum. In other words, the $m$th sensor has the minimum contribution to increasing $g(\mathbf{p})$. Therefore, we can update the selected subset as

$$\mathcal{S}_{\mathrm{in}} \leftarrow \mathcal{S}_{\mathrm{in}}\backslash m, \quad p_m = 0. \tag{42}$$

Afterwards, the gradients for the remaining $M-1$ sensors in $\mathcal{S}_{\mathrm{in}}$ need to be updated, and a new sensor will be removed similarly. This procedure will be terminated until the performance requirement is not satisfied any more, i.e.,

$$\frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\overline{\mathrm{oSNR}}_{\mathbf{p}}(\omega) \le \beta. \tag{43}$$

It is clear that the proposed gradient removal (GradR) method handles one sensor at each iteration. The most computationally complex operation is computing the gradient, which is dominated by calculating the vector $\mathbf{v}_k(\omega)$ and of the order of $\mathcal{O}(M^2)$. The time complexity of searching the minimum gradient is $\mathcal{O}(M)$ at each iteration. Suppose $Q$ iterations are required for the performance convergence, which is much smaller than $M$, the total time complexity of GradR will be $\mathcal{O}(Q(M^2+M))$.

### B. Weighted broadband input SNR-based approach

In [23], experimental results show that in the narrowband case the sensors that are close to the target source and those close to the FC are more likely to be selected, as they are of high signal quality for improving the noise reduction performance and of low transmission cost for reducing the energy consumption, respectively. Since these two kinds of sensors have a high input SNR and a low individual power consumption, respectively, we can thus use the weighted input SNR to approximately measure the contribution of each sensor to the optimality of (18), which is defined as

$$U_k = \frac{\mathrm{iSNR}_k}{C_k}, \quad k\in\{1,\ldots,M\}, \tag{44}$$

where the broadband input SNR is defined as

$$\begin{aligned} \mathrm{iSNR}_k &= \frac{\sum_{\omega=1}^{\Omega}\sum_l |X_k(\omega,l)|^2}{\sum_{\omega=1}^{\Omega}\sum_l |N_k(\omega,l)|^2} && (45)\\ &= \frac{\sum_{\omega=1}^{\Omega}\mathbf{e}_k^T\mathbf{\Phi}_{\mathbf{xx}}\mathbf{e}_k}{\sum_{\omega=1}^{\Omega}\mathbf{e}_k^T\mathbf{\Phi}_{\mathbf{nn}}\mathbf{e}_k}, && (46)\end{aligned}$$

which can be calculated easily at each sensor locally. Based on the weighted input SNR, we can then design an SNR-based sensor removal (SnrR) approach similarly to the gradient-based method. Due to the independence of complex matrix operations, the proposed SnrR method will be computationally much more efficient. Suppose $Q$ iterations are required for the performance convergence, the time complexity of SnrR is then $\mathcal{O}(QM)$.

### C. Weighted broadband energy-based approach

As the sensors around the target source are more likely to be selected, which have a high input SNR and thus also a high input energy, we can use the weighted broadband input energy to approximate the contribution of each sensor to the optimality of (18), which is defined as

$$E_k = \frac{\mathrm{iEnergy}_k}{C_k}, \quad k \in \{1,\ldots,M\}, \tag{47}$$

where the broadband input energy is defined as

$$\mathrm{iEnergy}_k = \sum_{\omega=1}^{\Omega}\sum_l |Y_k(\omega,l)|^2 \propto \sum_{\omega=1}^{\Omega}\mathbf{e}_k^T\mathbf{\Phi}_{\mathbf{yy}}\mathbf{e}_k, \tag{48}$$

which can also be calculated easily at each sensor locally. Based on the weighted input energy, we can then design an energy-based sensor removal (EnergyR) approach in a similar fashion. Clearly, SnrR and EnergyR have the same computationally complexity. Intuitively, in case there is no coherent interfering sources and the variances of uncorrelated noise across microphone nodes are approximately equal, the EnergyR method would be effective. However, the effectiveness will decrease dramatically in presence of coherent noise sources, particularly in very low SNR environments, as many sensor around the interfering sources might be chosen. Nevertheless, the EnergyR method is presented for comparison. Suppose $Q$ iterations are required for the performance convergence, the time complexity of EnergyR is also $\mathcal{O}(QM)$.

### D. Summary

To this end, we have introduced the proposed broadband methods and three greedy sensor selection approaches. In summary, the implementation details of the proposed low-complexity broadband FISS methods are described in Algorithm 2. The time complexities of the mentioned algorithms are summarized in Table I, where the utility based removal (UtilityR) method [30] is included for comparison. Regardless of the value of $\alpha$, both BroadOpt and NaVo have to solve $M$-dimensional SDP problems, so their time complexities are constant in terms of $\alpha$. However, for the greedy methods in case $\alpha$ becomes larger (i.e., the desired SNR is higher), less sensors need to be removed from the network as $Q = M - K$, which leads to a decrease in the time complexity. Hence, the superiority of the greedy methods in computational efficiency will be more distinct for large $\alpha$-values.

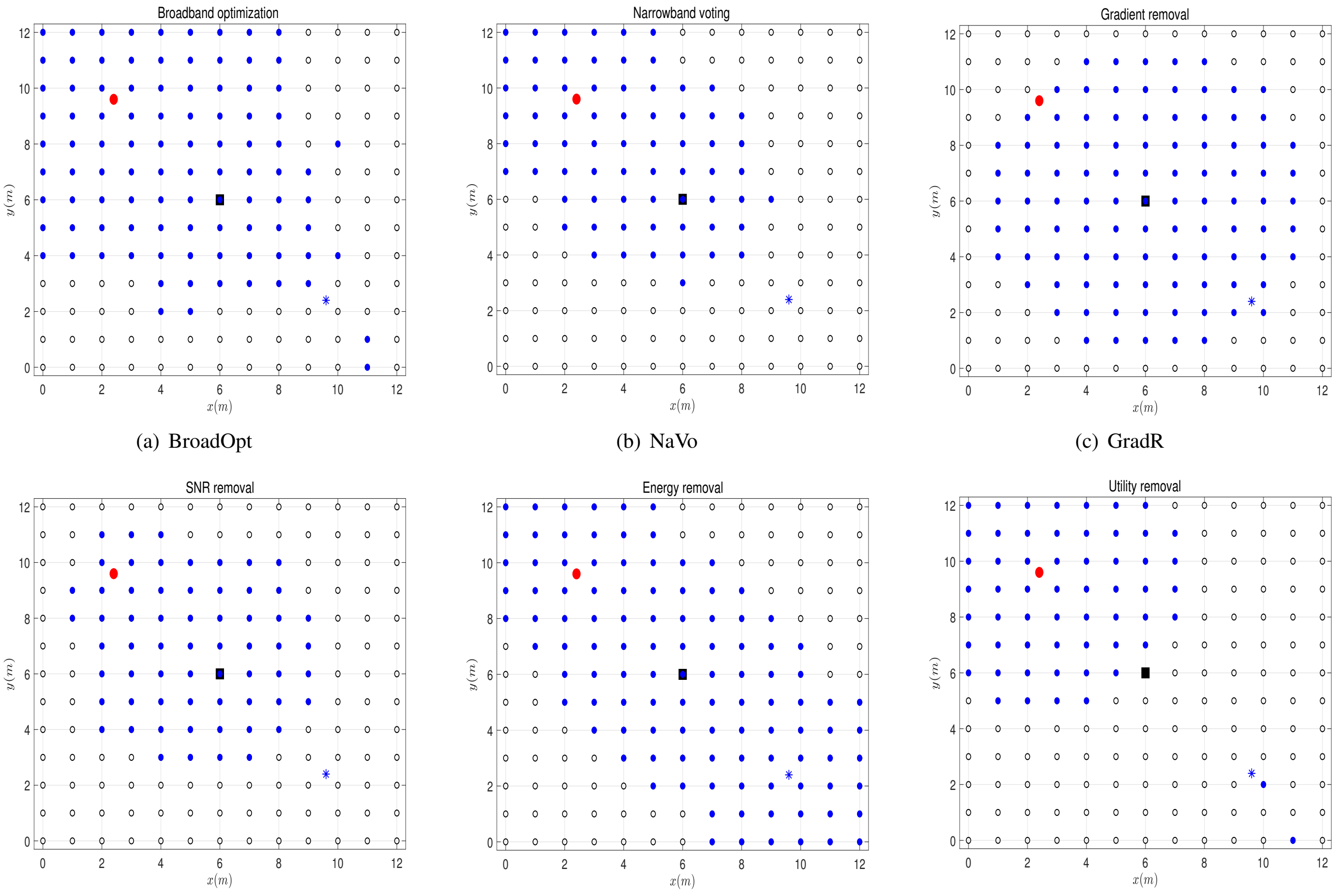
**Algorithm 2:** Proposed low-complexity approaches

```
Initialize: $\mathcal{S}_{\text{in}} = \{1,\ldots,M\}$;
For $k = 1,\ldots,M$ do
    Case 1: Gradient Removal (GradR)
        $m = \arg\min \frac{\partial g(\mathbf{p})}{\partial p_k}, \quad k \in \mathcal{S}_{\text{in}};$
    Case 2: SNR Removal (SnrR)
        $m = \arg\min \frac{\mathrm{iSNR}_k}{C_k}, \quad k \in \mathcal{S}_{\text{in}};$
    Case 3: Energy Removal (EnergyR)
        $m = \arg\min \frac{\mathrm{iEnergy}_k}{C_k}, \quad k \in \mathcal{S}_{\text{in}};$
    Update: $\mathcal{S}_{\text{in}} \leftarrow \mathcal{S}_{\text{in}} \backslash m, \quad p_m = 0$;
    If $\frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\overline{\mathrm{oSNR}_{\mathbf{p}}}(\omega) \leq \beta$, break;
End for
Beamforming: $\hat{X}_1(\omega) = \mathbf{w}_{\mathbf{p}}(\omega)^H\mathbf{y}_{\mathbf{p}}(\omega)$;
Return $\hat{X}_1(\omega)$, $\mathcal{S}_{\text{in}}$.
```

Table I
THE SUMMARY OF THE TIME-COMPLEXITY OF FISS APPROACHES.

| Method | Time complexity | $\alpha$ ↑ |
|---|---|---|
| BroadOpt | $\mathcal{O}(\Omega M^3)$ | constant |
| NaVo | $\mathcal{O}(\Omega M^3)$ | constant |
| GradR | $\mathcal{O}(Q(M^2+M))$ | decrease |
| SnrR | $\mathcal{O}(QM)$ | decrease |
| EnergyR | $\mathcal{O}(QM)$ | decrease |
| UtilityR [30] | $\mathcal{O}(Q(M^2+M))$ | decrease |

It is worth nothing that the proposed BroadOpt and NaVo methods require the noise statistics of the complete network, which thus belong to the model-based scheme. Compared to the model-driven approaches, the proposed greedy methods are data-driven, which can then be used for adaptive signal estimation and beamforming applications. In practice, the sensor signal statistics may change over time due to the movement of sources or sensors. In such dynamic scenarios, the model-based methods have to track the noise statistics in an online fashion, while the greedy methods only need to update the sensor utilities (i.e., gradient, weighted SNR, weighted energy), which is much more time-efficient than the former. Furthermore, another superiority of the greedy algorithms is that the extension to a forward mode is rather straightforward (i.e., at each iteration the sensor that has the largest contribution to the energy-aware noise reduction can be added to the selected subset), which, however is out the scope of this work, e.g., see [41]. From this perspective, the proposed greedy methods are designed in a backward mode.



(a) BroadOpt

(b) NaVo

(c) GradR

(d) SnrR

(e) EnergyR

(f) UtilityR

Figure 2. Sensor selection examples: (a) BroadOpt, (b) NaVo, (c) GradR, (d) SnrR, (e) EnergyR, and (f) UtilityR [30] for $\alpha = 0.6$ (Circle: sensors, Blue solid circle: selected sensors, Solid square: FC, Red solid dot: target source, Blue star: interfering source).

## V. Experiments

In this section, we will exploit a large-scale WASN to evaluate the performance of proposed FISS based MVDR beamforming approaches from several perspectives.

### A. Experimental setup and benchmark

*Experimental setup:* Fig. 2 shows the typical experimental setting that we use in the simulations, where 169 microphone nodes are uniformly distributed in a 2D room with dimensions (12×12) m$^2$. One target source is placed at (3.6, 9.6) m, and the FC is at (6, 6) m. One coherent interfering source is located at (9.6, 3.6) m. The speech source is originated from the TIMIT database [43], and the noise signal from the NoiseX-92 database [44], respectively. The room impulse responses (RIRs) of directional sources are generated using the image method [42]. The time-domain microphone signals are synthesized by summing: 1) the source component (convolving the source speech signal and its RIR), 2) interference component (convolving the interferer (i.e., a competing speaker) and the corresponding RIR) and 3) the uncorrelated noise component (i.e., microphone self noise). The uncorrelated noise is modeled as a white Gaussian random process. The signal to interferer ratio and the signal to uncorrelated noise ratio are set to be 0 dB and 30 dB, respectively. All the signals are sampled at 16 kHz. The reverberation time is set to be $T_{60}$ = 200 ms. We assume that all sensors are homogeneous (i.e., the power for keeping active is the same for all sensors), such that minimizing the power consumption reduces to minimizing the transmission power, and we can thus initialize $C_k, \forall k$ using the squared distance between sensor $k$ and the FC.

*Comparison method:* In order to observe the validity of the proposed methods, the most intuitive way is to compare with the optimal sensor selection solution, which can only be given by exhaustive searching. In large-scale WASNs (e.g., the considered experimental setup), exhaustive search is evidently intractable. Therefore, we employ the utility removal (UtilityR) method in [30] as the benchmark. Specifically, the UtilityR method can be designed similarly as the proposed greedy methods, which takes the SNR decrease of removing a sensor from the current selected sensor subset as the individual sensor utility and removes the sensor that has the minimum SNR decrease at each iteration. The most time consuming operation at each iteration is the matrix inverse of the noise covariance matrix, which is of the order of $\mathcal{O}(M^2)$, so that the time complexity of UtilityR is same as that of GradR.

[2] We employ such uniform sensor placement in a 2D configuration for the clarity of illustration. The application to a 3D scenario is straightforward, as the image method [42] can also generate RIRs in a 3D room. Note that the wireless sensor nodes can also be distributed randomly.

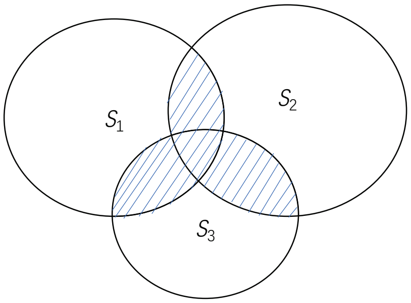


Figure 3. The relationship between the feasible sets of BroadOpt and NaVo.

### B. Performance evaluation in a static scenario

First of all, we evaluate the proposed methods in a static scenario, where we assume that the existing sources keep static and the noise covariance matrix is estimated using sufficiently long noise-only frames. Fig. 2 shows some sensor selection examples using the proposed methods with $\alpha = 0.6$. We can see that the selected sensors of the proposed NaVo method are included as a subset in the selected sensors obtained by BroadOpt. This can be roughly explained using Fig. 3. Suppose the feasible set of the $\omega$th frequency bin of (32) or (31) is denoted by $\mathcal{S}_\omega$, the feasible set of BroadOpt in (32) turns out to be the union set of $\mathcal{S}_1, \cdots, \mathcal{S}_\Omega$ (e.g., $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ when $\Omega = 3$), as all LMI constraints have to be satisfied simultaneously. The feasible set of NaVo in (31) will be the intersection region of $\mathcal{S}_1, \cdots, \mathcal{S}_\Omega$ (e.g., $\mathcal{S}_1 \cap \mathcal{S}_2 \cup \mathcal{S}_1 \cap \mathcal{S}_3 \cup \mathcal{S}_2 \cap \mathcal{S}_3$ when $\Omega = 3$), as these subsets are more likely to be selected across frequencies. As more sensors are selected, the BroadOpt method will cost a larger transmission power compared to NaVo. On the other hand, it is interesting that BroadOpt also selects some sensors close to the interfering source, which is due to the fact that these sensors are beneficial for suppressing the noise source, even though they have a low SNR. Most selected sensors are around the target source and the FC, as they are helpful for enhancing the desired speech signal (with a high input SNR) and for saving the transmission power, respectively.

In Fig. 2, it is clear that the proposed low-complexity approaches are also able to select the sensors that are close to the target source and the FC. The GradR and EnergyR methods additionally choose some sensors around the interfering source, while SnrR fails to select these sensors. This is due to the fact that the SnrR method only takes the input SNR and transmission cost into account. It can be seen that there are much more sensors around the interfering sources are selected by EnergyR, which is due to the fact that in experiments the signal-to-interference ratio is 0 dB. That is, the target source and the interfering source are equally “important” in the sense of energy, while the importances of the sensors around them are very different. It is expected that in case the input SNR increases, the selected sensors of the EnergyR method will assemble around the target source; in case the input SNR is very low, it is expected that the sensors around the interfering source dominate the selected subset. In addition, intuitively the NaVo and SnrR methods obtain the smallest cardinality of the selected sensor subset. Compared to the proposed methods, the UtilityR method fails to select the sensors around the FC, as it does not take the power consumption into account, which might cost a higher transmission power. Note that the results in Fig. 2 could be further optimized, as the proposed BroadOpt and NaVo involves convex relaxation and the low-complexity methods follow a greedy selection procedure.

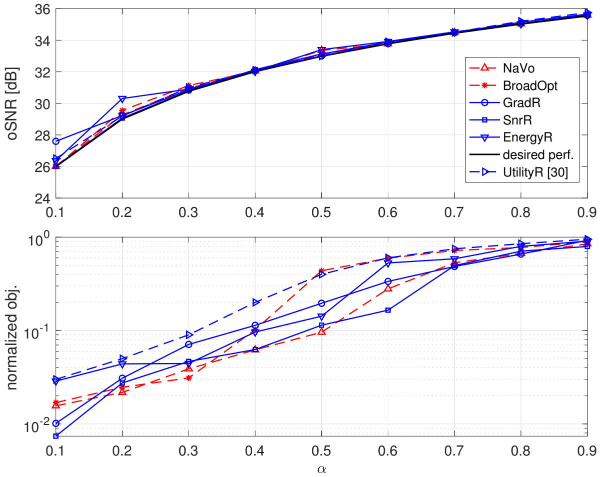


Figure 4. The output SNR and normalized transmission cost in terms of $\alpha$.

Further, we compare the broadband output SNR (which is the average SNR over the frequency range) and the normalized transmission power of the proposed methods in Fig. 4. The desired noise reduction performance is controlled by scaling the best performance $\overline{\mathrm{oSNR}}_{\max}$, which is obtained by involving all existing sensors. The obtained transmission cost is normalized by $\sum_{k=1}^{M} C_k$, which represents the maximum value of the objective function (i.e., when all sensors are involved). It is clear that the output SNRs of all methods are above the desired performance, meaning that the performance requirement is satisfied. The performance gap of the low-complexity methods with respect to the desired SNR is slightly larger than the optimization based methods. The transmission cost increases in terms of $\alpha$, as more sensors have to be selected for performance when increasing $\alpha$. By taking the energy efficiency into account, the proposed methods consume a lower transmission power compared to UtilityR, yet achieve a comparable noise reduction performance. The power consumption of BroadOpt is generally larger than that of NaVo, as the former involves more sensors, which was shown in Fig. 2. The NaVo and SnrR methods obtain the comparable and smallest transmission cost compared to the rest, because it was shown in Fig. 2 that they have the smallest cardinality of the selected subsets. More importantly, it can be seen that increasing $\alpha$ from 0.1 to 0.9, the SNR gain is only 10 dB, while the power consumption will be raised by 100 times as against when $\alpha = 0.1$. For example, when $\alpha = 0.6$, the SNR loss is around 2 dB, but the power consumption can be saved by 70% using the proposed methods for the considered WASN. This reveals that in practice many sensors are redundant for noise reduction, particularly in large-scale WASNs. The proposed methods can satisfy the performance constraint on the signal quality with a significant saving in transmission cost. Note that in case

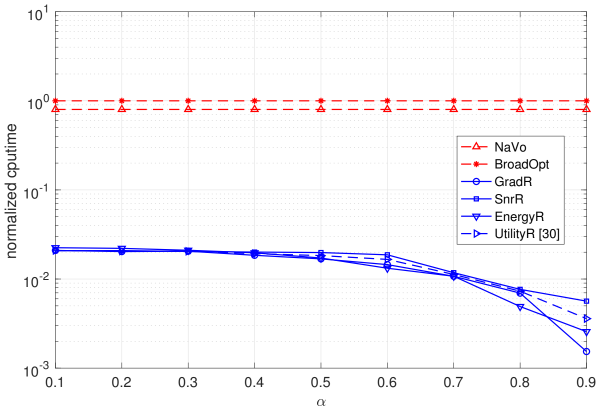


Figure 5. The normalized cputime of comparison methods in terms of $\alpha$.

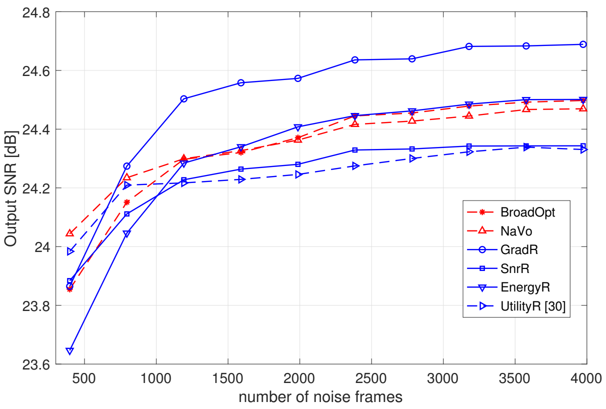


Figure 6. The output SNR in terms of the number of noise frames.

$\alpha = 1$, the performance of all selection methods will become equivalent to the best case $\overline{\text{oSNR}}_{\max}$ = 36.2 dB as all sensors have to be deployed.

We compare the time complexity of the aforementioned broadband sensor selection approaches in Fig. 5. The obtained running times are normalized by the cputime of BroadOpt, which is most time consuming. We can see that the cputimes of BroadOpt and NaVo are almost equal, although the latter is slightly smaller than the former. This is due to the fact that for BroadOpt, there are $\Omega$ LMI constraints, and each is of the order of $\mathcal{O}(M^3)$; for NaVo there are $\Omega$ SDP optimization problems, and each has an $M$-dimensional LMI constraint. As such, they are generally of the same time complexity, i.e., of the order of $\mathcal{O}(\Omega M^3)$. It is clear that the cputimes of both BroadOpt and NaVo are irrelevant to $\alpha$, as for any $\alpha$-value the order of the SDP problem does not change. On the other hand, the proposed low-complexity methods (i.e., GradR, SnrR and EnergyR) have the same time complexity, which is shown to be much smaller than the convex optimization based methods (i.e., BroadOpt and NaVo). More importantly, the cputimes (of GradR, SnrR and EnergyR) decrease with an increase in $\alpha$. This is due to the fact that increasing $\alpha$ means a higher performance that has to be reached, i.e., more sensors have to be included for performing MVDR beamforming. In this case, a smaller amount of sensors has to be excluded from the complete network for the proposed low-complexity methods, i.e., less iterations are required for performance satisfaction.

## *C. Performance evaluation in dynamic scenarios*

As the proposed methods depend on the use of the noise covariance matrix, which has to be estimated using the sample correlation matrix in practice. Given sufficiently long noise-only frames, the sample correlation matrix can be viewed as an unbiased estimator of the noise covariance matrix. However, due to the limited amount of noise data, there exists an estimation error in the noise covariance matrix. This often happens in the case of online continuous speech enhancement, as the microphone signals are received and processed frame by frame. For this dynamic case, we estimate the noise covariance matrix using the average smoothing technique as

$$\mathbf{\Phi}_{\mathbf{nn}}(l) = \mu\mathbf{\Phi}_{\mathbf{nn}}(l-1) + (1-\mu)\mathbf{n}(l)\mathbf{n}(l)^H, \tag{49}$$

where $\mu$ is the forgetting factor and set to be 0.98 in experiments. This smoothing technique is widely used in speech processing algorithms to track the dynamics of background noises (i.e., a competing sound source and Gaussian additive noise). As in the considered experimental setup, the noise component can be seen as short-time stationary, the forgetting factor should be chosen close to 1. It can be found via simulations that in case $\mu > 0.9$, it has a negligible effect on the performance. The inverse of $\mathbf{\Phi}_{\mathbf{nn}}(l)$ is given by

$$\mathbf{\Phi}_{\mathbf{nn}}^{-1}(l) = \mu^{-1}\mathbf{\Phi}_{\mathbf{nn}}^{-1}(l-1) - \frac{\mu^{-2}(1-\mu)\mathbf{\Phi}_{\mathbf{nn}}^{-1}(l-1)\mathbf{n}(l)\mathbf{n}(l)^H\mathbf{\Phi}_{\mathbf{nn}}^{-1}(l-1)}{1+\mu^{-1}(1-\mu)\mathbf{n}(l)^H\mathbf{\Phi}_{\mathbf{nn}}^{-1}(l-1)\mathbf{n}(l)}. \tag{50}$$

The time complexity of (50) is only $\mathcal{O}(M^2)$ in each time frame when $\mathbf{\Phi}_{\mathbf{nn}}^{-1}(l-1)$ in the previous frame is available.

We apply this recursive covariance estimation procedure into the proposed methods, and the output SNRs in terms of the number of noise frames are shown in Fig. 6. In order to make $\mathbf{\Phi}_{\mathbf{nn}}(l)$ always invertible, we force the minimum number of noise frames to be greater than the number of microphones and initialize $\mathbf{\Phi}_{\mathbf{nn}}(1) = 10^{-6}\mathbf{I}_M$, which functions as diagonal loading [45]. It can be seen that for all methods, the performance increases when more noise frames are available, as the estimate of the noise covariance matrix becomes more accurate. Due to the fact that the proposed GradR method selects more sensors than the rest (e.g., see Fig. 2), it achieves the best performance. This observation applies to other comparison methods. The proposed methods obtain a better performance than UtilityR, which means that using the proposed methods can achieve a better robustness against the noise covariance matrix estimation error. Note that although the number of noise frames affects the noise covariance matrix estimation error as well as the selection results of all methods, particularly at the beginning, it is clear from Fig. 6 that the overall performance and sensor statuses will converge with an increase in the number of noise frames.

As the proposed optimization based sensor selection methods (i.e., BroadOpt and NaVo) as well as the MVDR beamformer depend on the RTF vector of the target source, we further validate the effectiveness of the proposed methods

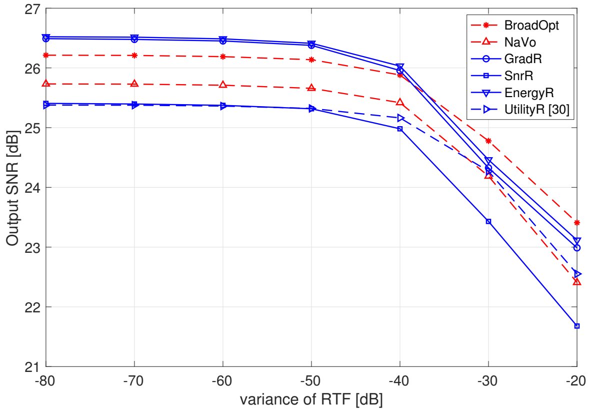


Figure 7. The output SNR in terms of RTF estimation errors, where the average variance of true RTF vectors across frequencies is around -5 dB.

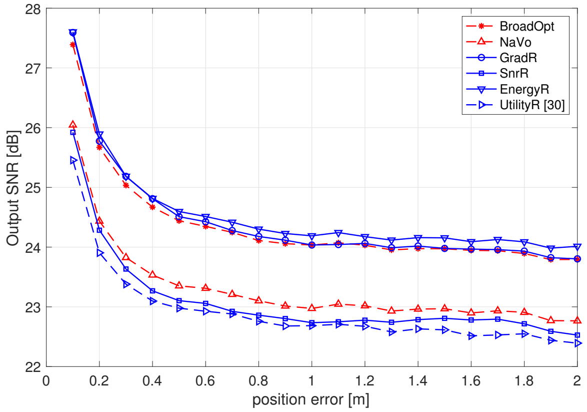


Figure 8. The output SNR in terms of signal position error.

against the RTF estimation error. First, we use the covariance whitening method [32] for RTF estimation and then add an error vector under the assumption that the RTF estimation error is zero-mean Gaussian distributed. The obtained output SNR in terms of the RTF estimation error (in dB) is shown in Fig. 7. It is clear that for all methods the noise reduction performance decreases with an increase in the RTF estimation error. The proposed BroadOpt, GradR and EnergyR methods exhibit a stronger robustness against the RTF estimation errors, due to the fact that more sensors around the target source are selected. The proposed NaVo method cannot work as robust as the proposed BroadOpt approach when the RTF error becomes large, which is due to the fact that NaVo fails to select the sensors close to the interfering source (e.g., see Fig. 2) as they are rather important for noise cancellation, particularly in the RTF mismatch case. Note that the EnergyR method also chooses many sensors around the noise source, which will cost a higher transmission power and is thus less energy efficient. Hence, in the large RTF mismatch case we recommend to apply the proposed BroadOpt method to perform energy-aware noise reduction in large-scale WASNs.

In addition, given the target source position, which can be estimated using sound source localization algorithms, in free sound field the steering vector can also be used for the MVDR beamformer design instead of the RTF. Let the time delay and the propagation distance from the target source to microphone $k$ be denoted by $\tau_k$ and $d_k$, respectively, which can be calculated using the source position in combination with the microphone positions. The steering vector is thus given by $\hat{\mathbf{a}} = \left[d_1^{-1}\exp(-j\omega\tau_1), \ldots, d_M^{-1}\exp(-j\omega\tau_M)\right]^T$. However, in practice there exists an estimation error in the source position, particularly in the presence of ambient noises. For this, we show the output SNRs of comparison methods in terms of the source positioning error in meter in Fig. 8. We assume that the estimated source position is located randomly in the circle centred by the true source position with a radius of the pre-defined error. The results are averaged over 100 realizations. It is clear that with an increase in the positioning error, the noise reduction performance decreases. Compared to Fig. 7, it is more clear that the proposed BroadOpt, GradR and EnergyR approaches show a stronger robustness against the positioning error. The proposed methods obtain a better performance than UtilityR, as the selected sensor subsets are more concentrated to the target source and the FC.

## VI. Conclusions

In this work, we investigated the broadband FISS strategy for MVDR beamforming based noise reduction in WASNs. Motivated by the NSS problem formulation, we proposed to minimize the transmission power over the network subject to a constraint on the broadband output signal quality (e.g., SNR). Compared to the narrowband counterpart, the resulting broadband sensor selection method can thus resolve a frequency-invariant network selection status, which avoids the switching on/off operations for WASNs. For this, we proposed two convex optimization based solvers (i.e., BroadOpt and NaVo). Indeed, the proposed NaVo method divides a large-scale SDP problem in BroadOpt into $M$ sub-problems which are of a much smaller size and uses a narrowband voting strategy for decision making. In order to further reduce the time complexity for the convex solvers, we proposed three greedy approaches (i.e., GradR, SnrR and EnergyR), which remove one sensor from the candidate set at each iteration until the performance requirement is unsatisfied. It was shown that all methods can reach the desired performance with a much smaller amount of sensors compared to the complete network, meaning that the power consumption can be saved significantly. This also reveals that in practice many sensors are non-informative for noise reduction, especially in large-scale WASNs. The sensors around the target source and the FC, and some close to the interfering source are beneficial for energy-aware beamformer designs. In practice, the proposed sensor selection strategies can be applied to effectively remove the network redundancy and construct an energy-efficient WASN. Given an unknown large WASN, one can initialize a small sensor subset around the target source and then increase the subset by gradually adding the sensors that have a larger contribution to the energy-aware noise reduction to the selected subset until the performance bound is satisfied. It was also shown that the proposed low-complexity methods are computationally much more efficient than the SDP-based methods. As in this work the sources keep static, in the future we will consider broadband sensor selection for more dynamic WASNs, e.g., by sensor scheduling.

## REFERENCES

[1] L. Turchet, G. Fazekas, M. Lagrange, H. S. Ghadikolaei, and C. Fischione, "The internet of audio things: State of the art, vision, and challenges," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10233–10249, 2020.

[2] G. Han, L. Liu, J. Jiang, L. Shu, and G. Hancke, "Analysis of energy-efficient connected target coverage algorithms for industrial wireless sensor networks," *IEEE Trans. Industrial Informatics*, vol. 13, no. 1, pp. 135–143, 2017.

[3] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: a signal processing perspective," in *IEEE Symp. Comm. & Vehi. Tech. in the Benelux (SCVT)*, 2011, pp. 1–6.

[4] X. Yue, Y. Liu, J. Wang, H. Song, and H. Cao, "Software defined radio and wireless acoustic networking for amateur drone surveillance," *IEEE Comm. Mag.*, vol. 56, no. 4, pp. 90–97, 2018.

[5] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Spatially correct rate-constrained noise reduction for binaural hearing aids in wireless acoustic sensor networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2731–2742, 2020.

[6] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wire. Comm. & Mobile Comp.*, vol. 2017, pp. 1–24, 2017.

[7] Stavros Ntalampiras, "Moving vehicle classification using wireless acoustic sensor networks," *IEEE Trans. Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 129–138, 2018.

[8] F. Alías and R. M. Alsina-Pagès, "Review of wireless acoustic sensor networks for environmental noise monitoring in smart cities," *J. sensors*, vol. 2019, pp. 1–13, 2019.

[9] Q. Wang, S. Guo, and Ka Fai Cedric Yiu, "Distributed acoustic beamforming with blockchain protection," *IEEE Trans. Industrial Informatics*, vol. 16, no. 11, pp. 7126–7135, 2020.

[10] J. Benesty, S. Makino, and J. Chen, *Speech enhancement*, Springer Science & Business Media, 2005.

[11] J. Zhang, J. Du, and L. Dai, "Sensor selection for relative acoustic transfer function steered linearly-constrained beamformers," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 1220–1232, 2021.

[12] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, 2009.

[13] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, 2015.

[14] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *IEEE Conf. on Decision and Control*, 2010, pp. 2572–2577.

[15] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011.

[16] Y. Selen, H. Tullberg, and J. Kronander, "Sensor selection for cooperative spectrum sensing," in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*. IEEE, 2008, pp. 1–11.

[17] H. Zhang, J. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2383–2395, 2009.

[18] Z. Dai, G. Wang, X. Jin, and X. Lou, "Nearly optimal sensor selection for TDOA-based source localization in wireless sensor networks," *IEEE Trans. Vehicular Technology*, vol. 69, no. 10, pp. 12031–12042, 2020.

[19] Y. Zhao, Z. Li, B. Hao, and J. Shi, "Sensor selection for TDOA-based localization in wireless sensor networks with non-line-of-sight condition," *IEEE Trans. Vehicular Technology*, vol. 68, no. 10, pp. 9935–9950, 2019.

[20] S. Liu, S. P. Chepuri, M. Fardad, E. Masazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3509–3522, 2016.

[21] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 363–375, 2013.

[22] A. Bertrand, "Utility metrics for assessment and subset selection of input variables for linear estimation [tips & tricks]," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 93–99, 2018.

[23] J. Zhang, S. P. Chepuri, R. C. Hendriks, and R. Heusdens, "Microphone subset selection for MVDR beamformer based noise reduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 550–563, 2018.

[24] G. E. Hovland and B. J. McCarragher, "Dynamic sensor selection for robotic systems," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 1997, vol. 1, pp. 272–277.

[25] W. Huang, Y. Huang, Y. Zeng, and L. Yang, "Wideband millimeter wave communication with lens antenna array: Joint beamforming and antenna selection with group sparse optimization," *IEEE Trans. Wireless Communications*, vol. 17, no. 10, pp. 6575–6589, 2018.

[26] H. Li, L. Song, and M. Debbah, "Energy efficiency of large-scale multiple antenna systems with transmit antenna selection," *IEEE Trans. Communications*, vol. 62, no. 2, pp. 638–647, 2014.

[27] F. Shu, Z. Wang, R. Chen, Y. Wu, and J. Wang, "Two high-performance schemes of transmit antenna selection for secure spatial modulation," *IEEE Trans. Vehicular Technology*, vol. 67, no. 9, pp. 8969–8973, 2018.

[28] K. Yang, H. Cui, L. Song, and Y. Li, "Efficient full-duplex relaying with joint antenna-relay selection and self-interference suppression," *IEEE Trans. Wireless Communications*, vol. 14, no. 7, pp. 3991–4005, 2015.

[29] Y. Sun, Dwk Ng, J. Zhu, and R. Schober, "Robust and secure resource allocation for full-duplex MISO multicarrier NOMA systems," *IEEE Trans. Communications*, vol. 66, no. 9, pp. 4119–4137, 2017.

[30] A. Bertrand, J. Szurley, P. Ruckebusch, I. Moerman, and M. Moonen, "Efficient calculation of sensor utility and sensor removal in wireless sensor networks for adaptive signal estimation and beamforming," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5857–5869, 2012.

[31] J. Zhang, H. Chen, L.-R. Dai, and R. C. Hendriks, "A study on reference microphone selection for multi-microphone speech enhancement," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 671–683, 2021.

[32] J. Zhang, R. Heusdens, and R. C. Hendriks, "Relative acoustic transfer function estimation in wireless acoustic sensor networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 10, pp. 1507–1519, 2019.

[33] B. Sklar, *Digital Communications: Fundamentals and Applications*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[34] Q. Wang, M. Hempstead, and W. Yang, "A realistic power consumption model for wireless sensor network devices," in *The 3rd annual IEEE communications society on sensor and ad hoc communications and networks*, 2006, vol. 1, pp. 286–295.

[35] Y. Huang and Y. Hua, "Energy planning for progressive estimation in multihop sensor networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4052–4065, 2009.

[36] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[37] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Signal Process. Mag.*, vol. 5, no. 2, pp. 4–24, 1988.

[38] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, vol. 1, Springer Science & Business Media, 2008.

[39] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[40] K. B. Petersen, M. S. Pedersen, et al., "The matrix cookbook," *Technical University of Denmark*, vol. 7, pp. 15, 2008.

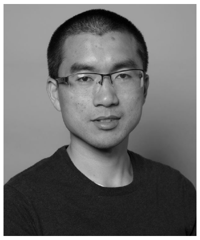
[41] J. Szurley, A. Bertrand, M. Moonen, P. Ruckebusch, and I. Moerman, "Energy aware greedy subset selection for speech enhancement in wireless acoustic sensor networks," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, 2012, pp. 789–793.

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

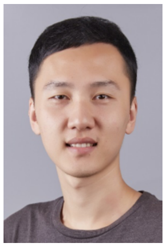
[43] J. S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, vol. 15, pp. 29–50, 1988.

[44] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech communication*, vol. 12, no. 3, pp. 247–251, 1993.

[45] L. Jian, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, 2003.



**Jie Zhang** (*IEEE Member*) received the M.Sc. degree (with honor) from the School of Electronics and Computer Engineering, Shenzhen Graduate School, Peking University, Beijing, China, in 2015, and the Ph.D. degree from the Circuits and Systems Group at the Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, The Netherlands, in 2020. Currently, he is an associate research fellow (assistant professor) in the National Engineering Research Center for Speech and Language Information Processing (NERC-SLIP), University of Science and Technology of China (USTC), Heifei, China. He was the receipt of the Best Student Paper Award from the IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), Sheffield, UK, 2018, and a student stravel grant for IEEE Signal Processing Society. His current research interests include single/multi-microphone speech processing for noise reduction, speech recognition and separation, binaural hearing aids, energy-aware wireless (acoustic) sensor networks.



**Guanghui Zhang** is currently a Research Assistant Professor at the Department of Computer Science, Hong Kong Baptist University. From 2020 to 2021, he worked as a Post-Doctoral Fellow with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, and with the Centre for Advances in Reliability and Safety, The Hong Kong Polytechnic University. Before that, he received the Ph.D. degree in Information Engineering from The Chinese University of Hong Kong in 2020, and the M.S. degree in Electronic Science and Technology from Peking University in 2016.

**Li-Rong Dai** was born in China in 1962. He received the B.S. degree in electrical engineering from Xidian University, Xi'an, China, in 1983, the M.S. degree from the Hefei University of Technology, Hefei, China, in 1986, and the Ph.D. degree in signal and information processing from the University of Science and Technology of China (USTC), Hefei, China, in 1997. In 1993, he joined USTC. He is currently a Professor with the School of Information Science and Technology, USTC. He has authored or coauthored more than 100 papers in the areas of his research interests, which include speech synthesis, speaker and language recognition, speech recognition, digital signal processing, voice search technology, machine learning, and pattern recognition.