



HAL
open science

Robust sensor placement for signal extraction

Fateme Ghayem, Bertrand Rivet, Rodrigo Cabral Farias, Christian Jutten

► **To cite this version:**

Fateme Ghayem, Bertrand Rivet, Rodrigo Cabral Farias, Christian Jutten. Robust sensor placement for signal extraction. 2020. hal-03325710v1

HAL Id: hal-03325710

<https://hal.science/hal-03325710v1>

Preprint submitted on 9 Nov 2020 (v1), last revised 25 Aug 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust sensor placement for signal extraction

Fateme Ghayem, *Student Member, IEEE*, Bertrand Rivet, Rodrigo Cabral Farias,
and Christian Jutten, *Fellow, IEEE*

Abstract—This paper proposes an efficient algorithm for robust sensor placement with the purpose of recovering a source signal from noisy measurements. To model uncertainty on the spatially-variant sensors gain and on the spatially correlated noise, we assume that both are realizations of Gaussian processes. Since the signal to noise ratio (SNR) is also uncertain in this context, to achieve a robust signal extraction, we propose a new placement criterion based on the maximization of the probability that the SNR exceeds a given threshold. This criterion can be easily evaluated using the Gaussian process assumption. Moreover, to reduce the computational complexity of the joint maximization of the criterion with respect to all sensor positions, we suggest a sequential maximization approach, where the sensor positions are chosen one at a time. Finally, we present numerical results showing the superior robustness of the proposed approach when compared to standard sensor placement criteria aimed at interpolating the spatial gain and to a recently proposed criterion aimed at maximizing the average SNR.

Index Terms—sensor placement, source extraction, signal to noise ratio, Gaussian processes.

I. INTRODUCTION

MANY signal processing problems can be cast from a generic setting where a source signal propagates through a given structure to the sensors. Under this setting, we can be interested either in (i) characterizing the source signal, or (ii) the structure, or even (iii) the resulting field of signals in some regions of the structure. In all these cases, signals are recorded by multiple sensors located at different positions within or on the structure. Due to price, energy or ergonomic constraints (*e.g.* when the structure is the human body), the number of sensors is often limited and it becomes crucial to place a few sensors at positions which contain the maximum information. This problem corresponds to optimal sensor placement and it is faced in a great number of applications ranging from infrastructure monitoring [1]–[4] and robotics [5] to biomedical signal processing [6], [7].

The way to tackle the problem of optimal sensor placement differs from one application to another, it mainly depends on which of three aspects mentioned above we want to focus on. In this paper, we study this problem aiming at the first aspect, that is, to extract a source signal from a set of noisy measurements collected from a limited number of sensors.

F. Ghayem, B. Rivet and C. Jutten are with the GIPSA-Lab, Department of Images and Signals, University of Grenoble Alpes, France. C. Jutten is also with Institut Universitaire de France, Paris. (e-mail: fateme.ghayem@gmail.com; bertrand.rivet@gipsa-lab.grenoble-inp.fr; christian.jutten@gipsa-lab.grenoble-inp.fr).

R. Cabral Farias is with Univ. Côte d’Azur, CNRS, I3S, 06900 Sophia-Antipolis, France (e-mail: cabral@i3s.unice.fr).

This work has been supported by the Center for International Scientific Studies and Collaboration (CISSC), and the European project ERC-2012AdG-320684-CHESS.

In what follows we describe briefly the model we consider for the recorded signals. All signals are supposed to be real values. Scalars are denoted by lower case letters z , vectors by lower case boldface letters \mathbf{z} and matrices by upper case boldface letters \mathbf{Z} . To retrieve the source of interest $s(t)$ at time t , a noisy measurement $y(\mathbf{x}, t)$ is recorded by a sensor positioned at $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^D$. The noisy measurement is related to $s(t)$ as follows:

$$y(\mathbf{x}, t) = a(\mathbf{x})s(t) + n(\mathbf{x}, t), \quad (1)$$

where $a(\mathbf{x})$ is the *spatial gain* between the source $s(t)$ and the sensor, and $n(\mathbf{x}, t)$ is the corresponding spatially correlated/time uncorrelated additive noise at time t . Now, if M sensors are used at locations $\mathbf{X}_M = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$, then the set of noisy measurements related to each sensor can be denoted by the vector $\mathbf{y}(\mathbf{X}_M, t) = [y(\mathbf{x}_1, t), y(\mathbf{x}_2, t), \dots, y(\mathbf{x}_M, t)]^T$, where \cdot^T is the transpose operator. We also denote $\mathbf{n}(\mathbf{X}_M, t) = [n(\mathbf{x}_1, t), n(\mathbf{x}_2, t), \dots, n(\mathbf{x}_M, t)]^T$ and $\mathbf{a}(\mathbf{X}_M) = [a(\mathbf{x}_1), a(\mathbf{x}_2), \dots, a(\mathbf{x}_M)]^T$ the corresponding noise and spatial gain vectors, respectively. One can then write the measurement model (1) in a more compact form

$$\mathbf{y}(\mathbf{X}_M, t) = \mathbf{a}(\mathbf{X}_M)s(t) + \mathbf{n}(\mathbf{X}_M, t). \quad (2)$$

The main goal is to find the optimal set of locations \mathbf{X}_M , such that the best estimation of the source $s(t)$ can be obtained from the measurements $\mathbf{y}(\mathbf{X}_M, t)$. We further assume that $a(\mathbf{x})$ is not fully known, either because we have measured it only in a few positions, or because the available prior information on it is vague. To model this uncertainty, we take a Bayesian perspective and we assume that $a(\mathbf{x})$ is a realization of a Gaussian process (GP). Such uncertainty modeling approach has been extensively used in function interpolation, where it is named kriging [8], [9], in the design of computer experiments [10] and in global optimization [11]. It has also been used for sensor placement in [12], where the increment of the mutual information (MI) is proposed as a placement criterion. In that work, the increment in MI is maximized using a near-optimal greedy approach. New sensors are added one by one starting with an empty set, in such a way that, at each iteration, the selected location adds maximum information.

The approach in [12] focuses on interpolating a field of signals *e.g.* room temperature. As any other kriging based criterion for interpolation [13], [14], this approach could be used in our setting if the goal was to estimate $a(\mathbf{x})$. But as previously stated, this is not our aim, and the two goals may not be equivalent. Consider a simple example where $a(\mathbf{x})$ has similar uncertainty everywhere in \mathcal{X} and that $n(\mathbf{x}, t)$ has a short correlation length. In this case, interpolation-based criteria will spread the sensors over \mathcal{X} , while a criterion targeting

the extraction of $s(t)$ will spread them across positions which the predicted value of $|a(\mathbf{x})|$ is higher.

The goal we pursue in this paper is similar to what has been recently studied in [15]. In that paper, by considering a GP model for $\mathbf{a}(\mathbf{x})$ and $n(\mathbf{x}, t)$ and using a linear source extraction approach, the output signal-to-noise-ratio (SNR) of the estimated source signal is derived as a function of $\mathbf{a}(\mathbf{x})$ and of the sensor locations \mathbf{X}_M . Since $\mathbf{a}(\mathbf{x})$ is stochastic, the SNR is stochastic too. So, in [15] the average SNR is suggested as the criterion to solve the problem of optimal sensor placement for source extraction. The suggested criterion significantly outperforms kriging-based methods in terms of the SNR. Although the average SNR takes into account the uncertainty in the gains, it may not lead to a robust sensor placement criterion, since increased levels of uncertainty on the gains affect not only the average SNR but also the dispersion of its distribution. As a result, the criterion can suggest some locations which have high values for the average SNR, but also high dispersion, leading to an actual value of the SNR that can be much lower than its average. In this case, not only the selected sensor positions may not improve the SNR, but they may decrease the SNR.

In this paper, we propose to extend the idea in [15] to take into account the robustness issue presented above. A more general criterion, based on the probability that the SNR exceeds a fixed threshold is put forward. Since the evaluation of this criterion requires the probability density function (pdf) of the SNR, we indicate how this pdf can be evaluated using the GP assumption on the spatial gain. Moreover, when we choose the sensor positions in a sequential fashion, we show that the criterion can then be evaluated analytically. Simulation results obtained with the sequential approach show that the threshold in the probability-based criterion can be adjusted to a trade-off between having a maximum increase in the SNR and being robust to a possible decrease in the SNR due to uncertainty. Also, the results confirm the superior robustness of our approach compared to other criteria from the literature.

The rest of the paper is organized as follows. In Section II we derive the criterion based on the SNR and we show how to evaluate the required pdf. Section III presents how the criterion becomes simpler with the sequential approach and in Section IV, we briefly describe two kriging-based existing criteria. In Section V, the simulation results are presented. Finally, section VI concludes the paper and discusses perspectives.

II. OPTIMAL SENSOR PLACEMENT FOR SOURCE EXTRACTION

In this section, we are going to solve the problem of optimal sensor placement for source extraction using the noisy sensor measurements model given by (2). We first provide a brief review of the source extraction problem. Then, we present the proposed criterion.

A. Linear signal extraction and signal to noise ratio

Let us consider the observation model in (2). By using a linear estimator described $\mathbf{f} \in \mathbb{R}^M$, the extracted source signal

$\hat{s}(t)$ is the following [15]:

$$\hat{s}(t) = \mathbf{f}^T \mathbf{y}(\mathbf{X}_M, t) = \mathbf{f}^T \mathbf{a}(\mathbf{X}_M) s(t) + \mathbf{f}^T \mathbf{n}(\mathbf{X}_M, t). \quad (3)$$

For simplicity, in the rest of this paper, the dependence on the set of sensor locations \mathbf{X}_M will be denoted with a subscript $(\cdot)_M$. For example, a function $\mathbf{P}(\mathbf{X}_M)$ will be denoted \mathbf{P}_M . From (3), the SNR of the estimation of $s(t)$ can be derived as

$$\text{SNR}(\mathbf{f}|\mathbf{X}_M) = \frac{\mathbb{E}_t[(\mathbf{f}^T \mathbf{a}_M s(t))^2]}{\mathbb{E}_t[(\mathbf{f}^T \mathbf{n}_M(t))^2]} = \frac{\sigma_s^2 \mathbf{f}^T \mathbf{a}_M \mathbf{a}_M^T \mathbf{f}}{\mathbf{f}^T \mathbf{C}_{MM}^n \mathbf{f}}, \quad (4)$$

where $\mathbb{E}_t[\cdot]$ stands for the expectation over time, and the signal time samples are temporally zero-mean, independent and identically distributed (iid) with $\sigma_s^2 = \mathbb{E}[s(t)^2]$ and $\mathbf{C}_{MM}^n = \mathbb{E}[\mathbf{n}_M(t) \mathbf{n}_M^T(t)]$. From (4), it can be seen that the SNR is a function of both the extraction vector \mathbf{f} and sensor locations \mathbf{X}_M . By considering the SNR as a function of \mathbf{f} and maximizing the output SNR over \mathbf{f} , assuming that the sensor positions \mathbf{X}_M are fixed and known, the best linear estimator \mathbf{f}^* classically is given by

$$\mathbf{f}^*|\mathbf{X}_M = \mathbf{R}_{MM} \mathbf{a}_M, \quad (5)$$

where $\mathbf{R}_{MM} \triangleq (\mathbf{C}_{MM}^n)^{-1}$. By conditioning on \mathbf{X}_M we note that the maximization over \mathbf{f} is done under the assumption that the sensor positions \mathbf{X}_M are known. For simplicity, in the rest of this paper, we denote the best extractor just by \mathbf{f}^* . By replacing (5) in (4), the corresponding SNR is

$$\text{SNR}(\mathbf{f}^*|\mathbf{X}_M) = \sigma_s^2 \mathbf{a}_M^T \mathbf{R}_{MM} \mathbf{a}_M. \quad (6)$$

Now, by observing the resulting output SNR in (6) as a function of sensor locations \mathbf{X}_M and maximizing it over \mathbf{X}_M , we end up with the optimal solution for sensor placement in terms of the best SNR. This solution implicitly assumes a perfect knowledge of spatial gains $a(\mathbf{x}_i)$, $i \in \{1, 2, \dots, M\}$. However, such a perfect knowledge is not always available. To overcome this issue a GP with mean $m^a(\mathbf{x})$ and covariance function $C^a(\mathbf{x}, \mathbf{x}')$ is considered to model $a(\mathbf{x})$ as $\hat{a}(\mathbf{x})$:

$$\hat{a}(\mathbf{x}) \sim \mathcal{GP}\left(m^a(\mathbf{x}), C^a(\mathbf{x}, \mathbf{x}')\right). \quad (7)$$

Consequently, the randomness of $a(\mathbf{x})$ leads to a stochastic SNR as well, which is estimated as:

$$\widehat{\text{SNR}}(\hat{\mathbf{f}}|\mathbf{X}_M) = \sigma_s^2 \hat{\mathbf{a}}_M^T \mathbf{R}_{MM} \hat{\mathbf{a}}_M, \quad (8)$$

where $\hat{\mathbf{f}} = \mathbf{R}_{MM} \hat{\mathbf{a}}_M$.

Remark 1. *If we have a prior information on $\mathbf{a}_K = [a(\mathbf{x}_1), a(\mathbf{x}_2), \dots, a(\mathbf{x}_K)]^T$, for example if we have previously placed the sensors at these positions and measured the gains, then the prior (7) can be updated to include this information, simply by conditioning the GP on the measurements of the gains. The conditioned GP is still a GP but with conditioned mean and covariance functions [16].*

To target the SNR in (6), we need to use a quantitative statistical measure of the SNR distribution which is robust against the uncertainty in the spatial gain model (7). This will be presented in the next part.

B. Proposed criterion

Since the pdf of a random variable contains all the information about it, here, by having the pdf of the SNR, we are able to define different efficient statistical criteria. Therefore, first, we attempt to derive the pdf of the SNR given in (8). Then, based on the obtained pdf, we present our robust criterion for sensor placement.

To simplify notation, we define the random variable $W_M \triangleq (1/\sigma_s^2)\text{SNR}(\mathbf{f}^*|\mathbf{X}_M) = \mathbf{a}_M^T \mathbf{R}_{MM} \mathbf{a}_M$. Now, in order to find the distribution of W_M , we have the following proposition:

Proposition 1. (Distribution of W_M) *If the spatial gain $a(\mathbf{x})$ follows the GP model $a(\mathbf{x}) \sim \mathcal{GP}(m^a(\mathbf{x}), C^a(\mathbf{x}, \mathbf{x}'))$, for any positive integer number M , then W_M is the weighted sum of M independent random variables as follows:*

$$W_M = \sum_{i=1}^M d_i y_i^2, \quad (9)$$

where y_i 's are independent normally distributed random variables $y_i \sim \mathcal{N}(m_{y_i}, 1)$, with $m_{y_i} = \mathbf{u}_i^T (\mathbf{C}_{MM}^a)^{-\frac{1}{2}} \mathbf{m}_M^a$, and, d_i 's and \mathbf{u}_i 's are respectively the eigenvalues and eigenvectors of the matrix \mathbf{A} defined as follows:

$$\mathbf{A} \triangleq (\mathbf{C}_{MM}^a)^{\frac{1}{2}} \mathbf{R}_{MM} (\mathbf{C}_{MM}^a)^{\frac{1}{2}}, \quad (10)$$

and $\mathbf{m}_M^a = [m^a(\mathbf{x}_1), m^a(\mathbf{x}_2), \dots, m^a(\mathbf{x}_M)]$.

Proof: Let vector \mathbf{w} be a normally distributed random vector with zero mean and an identity covariance matrix \mathbf{I}_M : $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I}_M)$. Now, we define a random variable y_i as follows:

$$y_i \triangleq \mathbf{u}_i^T (\mathbf{w} + \mathbf{m}'_M), \quad (11)$$

where

$$\mathbf{m}'_M = (\mathbf{C}_{MM}^a)^{-\frac{1}{2}} \mathbf{m}_M^a. \quad (12)$$

Note that this definition of the random variable y_i provides an independence between $d_i y_i^2$ and $d_j y_j^2$ for $i \neq j$ which will be studied in Remark 2.

Therefore, the squared form of y_i becomes as follows:

$$\begin{aligned} y_i^2 &= (\mathbf{w} + \mathbf{m}'_M)^T \mathbf{u}_i \mathbf{u}_i^T (\mathbf{w} + \mathbf{m}'_M) \\ &= \mathbf{w}^T \mathbf{u}_i \mathbf{u}_i^T \mathbf{w} + 2(\mathbf{m}'_M)^T \mathbf{u}_i \mathbf{u}_i^T \mathbf{w} + (\mathbf{m}'_M)^T \mathbf{u}_i \mathbf{u}_i^T (\mathbf{m}'_M). \end{aligned}$$

According to the above equation, we have the following:

$$\begin{aligned} \sum_{i=1}^M d_i y_i^2 &= \mathbf{w}^T \left[\sum_{i=1}^M (d_i \mathbf{u}_i \mathbf{u}_i^T) \right] \mathbf{w} \\ &+ 2(\mathbf{m}'_M)^T \left[\sum_{i=1}^M (d_i \mathbf{u}_i \mathbf{u}_i^T) \right] \mathbf{w} + (\mathbf{m}'_M)^T \left[\sum_{i=1}^M (d_i \mathbf{u}_i \mathbf{u}_i^T) \right] \mathbf{m}'_M. \end{aligned}$$

Since d_i 's and \mathbf{u}_i 's are the eigenvalues and eigenvectors of \mathbf{A} , it yields $\sum_{i=1}^M (d_i \mathbf{u}_i \mathbf{u}_i^T) = \mathbf{A}$. So, we have

$$\sum_{i=1}^M d_i y_i^2 = \mathbf{w}^T \mathbf{A} \mathbf{w} + 2(\mathbf{m}'_M)^T \mathbf{A} \mathbf{w} + (\mathbf{m}'_M)^T \mathbf{A} (\mathbf{m}'_M). \quad (13)$$

By replacing (12) and (10) in (13), we obtain the following:

$$\sum_{i=1}^M d_i y_i^2 = (\mathbf{m}_M^a + (\mathbf{C}_{MM}^a)^{\frac{1}{2}} \mathbf{w})^T \mathbf{R}_{MM} (\mathbf{m}_M^a + (\mathbf{C}_{MM}^a)^{\frac{1}{2}} \mathbf{w}). \quad (14)$$

Since \mathbf{w} is a normally distributed random vector with zero mean and an identity covariance matrix, and also \mathbf{m}_M^a and \mathbf{C}_{MM}^a are the mean vector and covariance matrix of the model of the spatial gain respectively, we conclude that $\mathbf{m}_M^a + (\mathbf{C}_{MM}^a)^{\frac{1}{2}} \mathbf{w}$ is equal to \mathbf{a}_M . So, we have the following:

$$\sum_{i=1}^M d_i y_i^2 = \mathbf{a}_M^T \mathbf{R}_{MM} \mathbf{a}_M = (1/\sigma_s^2) \text{SNR}(\mathbf{f}^*|\mathbf{X}_M) = W_M. \quad \blacksquare$$

Remark 2. *Considering (11), it is straightforward to show that for any $i \neq j \in \{1, 2, \dots, M\}$, y_i and y_j are uncorrelated. On the other hand, y_i and y_j are jointly normally distributed. Therefore, y_i and y_j are independent. Moreover, as long as the Jacobian of a non-linear transformation $h: \mathbb{R}^M \rightarrow \mathbb{R}^{M'}$ is diagonal, the non-linear functions will not affect the independence of the random variables. Therefore, by defining $h(y_i) \triangleq d_i y_i^2$, since the Jacobian matrix of the function h over the independent random variables $y_i, i \in \{1, 2, \dots, M\}$ is diagonal, it can be concluded that for any $i \neq j \in \{1, 2, \dots, M\}$, $\gamma_i \triangleq d_i y_i^2$ and $\gamma_j \triangleq d_j y_j^2$ are also independent.*

Remark 3. *Since y_i is a normally distributed scalar with non-zero mean m_{y_i} and variance one, its squared form y_i^2 follows a noncentral chi-squared distribution with the number of degrees of freedom $k_i = 1$, and non-centrality parameter $\lambda_i = m_{y_i}^2$. So, the pdf of the random variable $v_i \triangleq y_i^2$ becomes*

$$g_{V_i}(v_i; k_i, \lambda_i) = \frac{1}{2} \exp^{-\frac{(v_i + \lambda_i)}{2}} \left(\frac{v_i}{\lambda_i} \right)^{\left(\frac{k_i}{4} - \frac{1}{2}\right)} I_{\frac{k_i}{2} - 1}(\sqrt{\lambda_i v_i}), \quad (15)$$

where $I(\cdot)$ is the modified Bessel function of the first kind. Now, by defining $\gamma_i = d_i v_i$ to be the scaled form of the random variable v_i with the positive scale factor d_i , the distribution of γ_i becomes as follows:

$$g_{\Gamma_i}(\gamma_i) = \frac{1}{d_i} g_{V_i}\left(\frac{\gamma_i}{d_i}; k_i, \lambda_i\right), \quad (16)$$

denoting $g_{V_i}(\cdot)$ the pdf of the random variable $v_i = y_i^2$ with noncentral chi-squared distribution defined in (15).

From Remark 2 and Remark 3, it can be concluded that the SNR is the sum of M independent random variables γ_i each having a pdf defined in (16). Therefore, due to independence, the pdf of the SNR is given by the convolution product, denoted by $*$, between the pdf of M random variables γ_i :

$$\begin{aligned} g_{W_M}(w_M) &= g_{\Gamma_1}(w_M) * g_{\Gamma_2}(w_M) * \dots * g_{\Gamma_M}(w_M) \\ &= \frac{1}{\prod_{i=1}^M d_i} g_{V_1}\left(\frac{w_M}{d_1}\right) * g_{V_2}\left(\frac{w_M}{d_2}\right) * \dots * g_{V_M}\left(\frac{w_M}{d_M}\right). \end{aligned} \quad (17)$$

Note that, for simplicity, in the above, the parameters k_i and λ_i are omitted, and we presented $g_{V_i}(v_i; k_i, \lambda_i)$ by $g_{V_i}(v_i)$.

Now, by knowing the pdf of the SNR, it is possible to define different criteria. To have an efficient criterion, it is necessary to consider two important properties: first, the criterion has to suggest positions providing maximum output SNR. Second, the criterion should be robust against the uncertainty on the

gains, that is, it should avoid positions that have a non-negligible probability of generating a small SNR. So, to achieve these two goals, we propose to search for a set of positions that maximizes the probability of the SNR to be greater than a threshold denoted θ . This leads to the following problem:

$$\hat{\mathbf{X}}_M = \underset{\mathbf{X}_M}{\operatorname{argmax}} J_P(\mathbf{X}_M, \theta), \quad (18)$$

where the criterion $J_P(\mathbf{X}_M, \theta)$ is based on the cumulative distribution function (cdf) of the SNR:

$$J_P(\mathbf{X}_M, \theta) = Pr(W_M \geq \theta) = 1 - G_{W_M}(\theta), \quad (19)$$

where $G_{W_M}(\cdot)$ represents the cdf of the SNR/σ_s^2 (i.e. W_M) conditioned to the sensors at positions \mathbf{X}_M . By knowing the pdf of W_M , from (17) the cdf becomes as follows:

$$G_{W_M}(\theta; \mathbf{k}, \Lambda) = \frac{1}{\prod_{i=1}^M d_i} \times G_{V_1}\left(\frac{\theta}{d_1}; k_1, \lambda_1\right) * G_{V_2}\left(\frac{\theta}{d_2}; k_2, \lambda_2\right) * \dots * G_{V_M}\left(\frac{\theta}{d_M}; k_M, \lambda_M\right),$$

where $\mathbf{k} = \{k_1, k_2, \dots, k_M\}$, $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$, and $G_{V_1}(\cdot)$ is the corresponding cdf of $g_{V_1}(\cdot)$. Note that from Remark 3 we also have:

$$G_{V_1}(v_1; k_1, \lambda_1) = 1 - Q_{\frac{k_1}{2}}(\sqrt{\lambda_1}, \sqrt{v_1}), \quad (20)$$

where λ_i and k_i are defined in (15), and Q is the Marcum-Q-function [17].

Remark 4. (Links with maximum likelihood) It is worth mentioning that if we consider a target value α of SNR/σ_s^2 , then the maximum likelihood can also be used as a sensor placement criterion. In this case, we look for the set \mathbf{X}_M maximizing

$$J_P(\mathbf{X}_M, \alpha) = g_{W_M}(\alpha). \quad (21)$$

However, this criterion does not have an intuitive interpretation regarding robustness and maximization of SNR as (19).

The advantage of criterion (19) is that the free parameter θ can be used to control the risk we want to take when placing new sensors. In other words, by increasing θ to sufficiently high values, the upper tail of the output SNR will be compared for different positions, forcing the criterion to be high for positions leading to a high SNR but which are conservative. In this case, the positions leading to very high average SNR but with large dispersion will be discarded. On the contrary, by reducing the parameter θ to a sufficiently low value, most of the positions will have similar criterion values, and, as a consequence, risky sensor positions leading to a very high average SNR values will not be discarded. So, unlike maximum likelihood, with this criterion, depending on the application, we can make a trade-off between achieving a sufficiently high SNR and reducing the risk that the true output SNR will be much lower than expected.

III. SEQUENTIAL SENSOR PLACEMENT

In this section, we first discuss about the computational complexity of the sensor placement problem for M sensors. Two solutions that reduce the complexity complexity are then presented. One solution is based on a greedy method for breaking the underlying maximization problem into smaller problems. The second solution follows the same greedy approach, but it exploits information on the spatial gain that may be available when the smaller sensor placement problems are solved sequentially.

A. Computational cost

In practice, to optimize the criterion (19), a grid search is used. We are thus looking for the best subset \mathbf{X}_M from a total spatial grid of T positions \mathbf{X}_T to place M sensors such that our proposed criterion in (19) is maximized. The optimal solution to this problem requires a combinatorial search over $\frac{T!}{M!(T-M)!}$ possibilities. This search has a high computational cost, especially if the grid is very tight, i.e. when T is large. Therefore, it is essential to provide a less costly solution. In the following, we present a straightforward way to tackle this problem.

B. Greedy method

One approach to reduce the computational complexity of the combinatorial search is to use a greedy method. In this approach, we start from an empty set and iteratively add $N < M$ sensors such that at each selection, the underlying criterion becomes maximum for the totally selected positions up to each iteration. Assume that K sensors have already been located at \mathbf{X}_K and we want to add the next N new sensors at \mathbf{X}_N where N is smaller than the total number of sensors M . Among all the possibilities for the N new sensors, we search for the subset \mathbf{X}_N such that $\mathbf{X}_{K+N} = \{\mathbf{X}_K \cup \mathbf{X}_N\}$ maximizes (19). Since at each iteration $N < M$, the computational complexity will be reduced compared with a direct search for \mathbf{X}_M . However, such complexity reduction comes at the expense of a possible suboptimality of the chosen sensor positions.

C. Sequential approach

When breaking the larger search for the M optimal positions into a sequential search for N positions, we may be interested in using the extraction system with the new sensors, prior to choosing and placing the next N sensors. In such a case, it is possible to use the available measurements not only to extract the signal but also to obtain new information on the spatial gains. If such information can be retrieved, then it can be used to reduce the uncertainty on $\hat{a}(\mathbf{x})$ and, as a consequence, to reduce the uncertainty on the SNR for the next N sensors to be placed. Note that this sequential approach is different from the greedy method, whose only objective is to iteratively maximize the criterion and not to retrieve any other useful information that may reduce the uncertainty.

In practice, information on the spatial gains may be obtained either if a ground truth signal is available [7], or with blind

signal processing techniques such as independent component analysis (ICA) [18] or sparse component analysis (SCA) [19]. In the following, we suppose that one of these approaches has been used and that the new information available on the gain is equivalent to noisy measurements of it, denoted \mathbf{z}_K , obtained at the previously placed sensor positions \mathbf{X}_K :

$$\mathbf{z}_K = \mathbf{a}_K + \mathbf{v}_K, \quad (22)$$

where $\mathbf{v}_K(\mathbf{x})$ is a Gaussian process independent of $\mathbf{a}(\mathbf{x})$ and $\mathbf{n}(\mathbf{x}, t)$, and with the covariance function $C^v(\mathbf{x}, \mathbf{x}')$.

Once we placed the K sensors and measured \mathbf{z}_K , we can update the criterion for placing the new N sensors by conditioning the gain on \mathbf{z}_K . The gain vector can then be written as $\hat{\mathbf{a}}_M^T = [\hat{\mathbf{a}}_K^T \hat{\mathbf{a}}_N^T]$, where $\hat{\mathbf{a}}_K$ are the gains for which we have measurements, $\hat{\mathbf{a}}_N$ is the random vector corresponding to the spatial gains at candidate positions \mathbf{X}_N , and both are conditioned on \mathbf{z}_K . As mentioned in Remark 1, since both the gains and \mathbf{z}_K are Gaussian, the conditioned gains are also Gaussian with mean and covariance given by [16]

$$\begin{aligned} \mathbf{m}_{M|K} &= \mathbb{E}[\hat{\mathbf{a}}_M | \mathbf{z}_K] \\ &= \mathbf{m}_M + \mathbf{C}_{MK}^a [\mathbf{C}_{KK}^a + \mathbf{C}_{KK}^v]^{-1} (\mathbf{z}_K - \mathbf{m}_K), \end{aligned} \quad (23)$$

and

$$\begin{aligned} \mathbf{C}_{M|K}^a &= \mathbb{E} \left[(\hat{\mathbf{a}}_M - \mathbf{m}_{M|K}) (\hat{\mathbf{a}}_M - \mathbf{m}_{M|K})^T \right] \\ &= \mathbf{C}_{MM}^a - \mathbf{C}_{MK}^a [\mathbf{C}_{KK}^a + \mathbf{C}_{KK}^v]^{-1} (\mathbf{C}_{MK}^a)^T, \end{aligned} \quad (24)$$

where \mathbf{C}_{MK}^a , \mathbf{C}_{KK}^a are blocks of the partitioned covariance matrix $\mathbf{C}_{MM}^a = \begin{bmatrix} \mathbf{C}_{MK}^a & \mathbf{C}_{MN}^a \\ \mathbf{C}_{KN}^a & \mathbf{C}_{NN}^a \end{bmatrix}$.

The estimated SNR is similar to (8) $\widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_N, \mathbf{z}_K)$, but, since \mathbf{X}_K is fixed, it is now conditioned on \mathbf{z}_K and a function of \mathbf{X}_N only. As $\hat{\mathbf{a}}_M$ is Gaussian, the distribution of $W_{M|K} \triangleq (1/\sigma_s^2) \widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_N, \mathbf{z}_K)$ can be obtained in a similar way as presented in Subsection II-B. With the distribution of $W_{M|K}$, we can modify the criterion (19) to have the following robust sequential sensor placement criterion:

$$J_P(\mathbf{X}_N, \theta | \mathbf{z}_K) = \Pr(W_{M|K} \geq \theta) = 1 - G_{W_{M|K}}(\theta). \quad (25)$$

1) *Perfect gain information*: if the information on the gain is assumed to be perfect, that is, $\mathbf{z}_K = \mathbf{a}_K$, then the vector of gains can be written as $\hat{\mathbf{a}}_M^T = [\mathbf{a}_K^T \hat{\mathbf{a}}_N^T]$, where \mathbf{a}_K are now deterministic and $\hat{\mathbf{a}}_N$ is a random vector conditioned on the observed \mathbf{a}_K . To write the conditional SNR, we partition $\mathbf{R}_{MM} = \begin{bmatrix} \mathbf{R}_{KK} & \mathbf{R}_{KN} \\ \mathbf{R}_{KN}^T & \mathbf{R}_{NN} \end{bmatrix}$, and then we have:

$$\begin{aligned} \frac{1}{\sigma_s^2} \widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_M, \mathbf{a}_K) &= [\mathbf{a}_K^T \hat{\mathbf{a}}_N^T] \begin{bmatrix} \mathbf{R}_{KK} & \mathbf{R}_{KN} \\ \mathbf{R}_{KN}^T & \mathbf{R}_{NN} \end{bmatrix} \begin{bmatrix} \mathbf{a}_K \\ \hat{\mathbf{a}}_N \end{bmatrix} \\ &= \mathbf{a}_K^T \mathbf{R}_{KK} \mathbf{a}_K + 2\mathbf{a}_K^T \mathbf{R}_{KN} \hat{\mathbf{a}}_N + \hat{\mathbf{a}}_N^T \mathbf{R}_{NN} \hat{\mathbf{a}}_N. \end{aligned} \quad (26)$$

The conditioned random vector $\hat{\mathbf{a}}_N$ is also Gaussian, its mean $\mathbf{m}_{N|K}$ and covariance $\mathbf{C}_{N|K}^a$ can be obtained respectively from (23) and (24) by setting the noise covariance equal to zero, i.e. $\mathbf{C}_{KK}^v = \mathbf{0}$. They are given by

$$\mathbf{m}_{N|K} = \mathbb{E}[\hat{\mathbf{a}}_N] = \mathbf{m}_N + (\mathbf{C}_{KN}^a)^T (\mathbf{C}_{KK}^a)^{-1} (\mathbf{a}_K - \mathbf{m}_K). \quad (27)$$

and

$$\begin{aligned} \mathbf{C}_{N|K}^a &= \mathbb{E} \left[(\hat{\mathbf{a}}_N - \mathbf{m}_{N|K}) (\hat{\mathbf{a}}_N - \mathbf{m}_{N|K})^T \right] \\ &= \mathbf{C}_{NN}^a - \mathbf{C}_{NK}^a (\mathbf{C}_{KK}^a)^{-1} (\mathbf{C}_{NK}^a)^T \end{aligned} \quad (28)$$

Assuming that \mathbf{R}_{NN} is invertible and by factorizing (26), we have

$$\begin{aligned} \widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_N, \mathbf{a}_K) &= \sigma_s^2 \times \\ & \left[(\hat{\mathbf{a}}_N + \mathbf{R}_{NN}^{-1} \mathbf{R}_{NK} \mathbf{a}_K)^T \mathbf{R}_{NN} (\hat{\mathbf{a}}_N + \mathbf{R}_{NN}^{-1} \mathbf{R}_{NK} \mathbf{a}_K) \right. \\ & \left. + \mathbf{a}_K^T (\mathbf{R}_{KK} - \mathbf{R}_{KN} \mathbf{R}_{NN}^{-1} \mathbf{R}_{KN}^T) \mathbf{a}_K \right]. \end{aligned} \quad (29)$$

Note that the first term is a quadratic form of a Gaussian vector, while the second term is deterministic. Therefore, the random vector

$$\begin{aligned} W_{N|K} &\triangleq (1/\sigma_s^2) \widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_N, \mathbf{a}_K) \\ &= \mathbf{a}_K^T (\mathbf{R}_{KK} - \mathbf{R}_{KN} \mathbf{R}_{NN}^{-1} \mathbf{R}_{KN}^T) \mathbf{a}_K \end{aligned} \quad (30)$$

has a distribution that can be obtained again as described in Subsection II-B. Similarly to (19), a criterion $J_P(\mathbf{X}_N, \theta | \mathbf{a}_K)$ can also be defined.

2) *Adding one sensor at a time under the perfect gain information setting*: a special case for which $J_P(\mathbf{X}_N, \theta | \mathbf{a}_K)$ has a known analytical form is when we add new sensors one by one, i.e. $N = 1$. In this case, $\hat{\mathbf{a}}_N$ is a scalar random variable and both R_{NN} and $\mathbf{C}_{N|K}^a$ are also scalars. Therefore, we can rewrite (29) as follows

$$\begin{aligned} \widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_N, \mathbf{a}_K) &= \sigma_s^2 \times \\ & \left[R_{NN} \left(a_N + \frac{\mathbf{R}_{NK} \mathbf{a}_K}{R_{NN}} \right)^2 + \mathbf{a}_K^T \left(\mathbf{R}_{KK} - \frac{\mathbf{R}_{KN} \mathbf{R}_{KN}^T}{R_{NN}} \right) \mathbf{a}_K \right]. \end{aligned}$$

If we define $q_N \triangleq \left(a_N + \frac{\mathbf{R}_{NK} \mathbf{a}_K}{R_{NN}} \right) / \sqrt{C_{N|K}^a}$, then

$$\begin{aligned} \widehat{\text{SNR}}(\hat{\mathbf{f}} | \mathbf{X}_N, \mathbf{a}_K) &= \sigma_s^2 \times \\ & \left[R_{NN} C_{N|K}^a q_N^2 + \mathbf{a}_K^T \left(\mathbf{R}_{KK} - \frac{\mathbf{R}_{KN} \mathbf{R}_{KN}^T}{R_{NN}} \right) \mathbf{a}_K \right]. \end{aligned} \quad (31)$$

From the Gaussian assumption, we have $q_N \sim \mathcal{N}(m_{q_N}, \sigma_{q_N}^2)$, where $m_{q_N} = \left(m_{N|K} + \frac{\mathbf{R}_{NK} \mathbf{a}_K}{R_{NN}} \right) / \sqrt{C_{N|K}^a}$ and $\sigma_{q_N}^2 = 1$. Therefore, it can be concluded that q_N^2 has a non-central chi-squared distribution as $q_N^2 \sim \chi^2(1, \lambda)$, with degree of freedom $k = 1$ and non-centrality parameter

$$\lambda = m_{q_N} = \left(m_{N|K} + \frac{\mathbf{R}_{NK} \mathbf{a}_K}{R_{NN}} \right) / \sqrt{C_{N|K}^a}. \quad (32)$$

Using expression (31) and (20), we can give an expression for the robust placement criterion

$$J_P(\mathbf{X}_N, \theta | \mathbf{a}_K) = \Pr(W_{N|K} \geq \theta) = 1 - Q_{\frac{1}{2}}(\sqrt{\lambda}, \sqrt{\theta}), \quad (33)$$

where

$$\theta' = \frac{\theta - \mathbf{a}_K^T \left(\mathbf{R}_{KK} - \frac{\mathbf{R}_{KN} \mathbf{R}_{KN}^T}{R_{NN}} \right) \mathbf{a}_K}{R_{NN} C_{N|K}^a}. \quad (34)$$

IV. OTHER SENSOR PLACEMENT CRITERIA

This section contains a brief description of kriging based alternative criteria for sensor placement presented in the literature. We begin the section with a description of classical kriging based criteria for interpolation, more precisely, MI and entropy based criteria. Then, we present a criterion specially tailored for signal extraction, that is based on the expected predicted output SNR. We finish the section with a discussion on the advantages and disadvantages of each approach.

A. Classical kriging criteria

In the classical kriging approaches for sensor placement, the objective is to find sensor locations such that the set of measurements collected by the sensors provides the best estimation of the spatial gain $\hat{\mathbf{a}}(\mathbf{X})$ in the whole space \mathcal{X} . In this regard, several criteria have been proposed. Among them, entropy and MI are two basic tools to define an effective cost function. In what follows, we explain how these two measures can be used as criteria for sensor placement.

1) *Entropy*: Assume that \mathcal{S} is the set of possible positions in the space for sensor placement. Based on the concept of entropy, the most informative subset $\mathcal{A} \subset \mathcal{S}$ with the size $|\mathcal{A}| = M$ to place M sensors, is the one that maximizes the entropy, which implies that the spatial gain at the selected positions are mostly uncertain about each other. Therefore, the following problem needs to be solved [20]:

$$\mathcal{A} = \underset{\mathcal{A} \subset \mathcal{S}; |\mathcal{A}|=M}{\operatorname{argmax}} J_H(\mathbf{X}_{\mathcal{A}}), \quad (35)$$

where $J_H(\mathbf{X}_{\mathcal{A}}) = H(\mathbf{a}_{\mathcal{A}})$ is the entropy of the spatial gain at the set of positions \mathcal{A} . The entropy is defined as follows:

$$H(\mathbf{a}_{\mathcal{A}}) = - \int_{-\infty}^{+\infty} p(\mathbf{a}_{\mathcal{A}}) \log p(\mathbf{a}_{\mathcal{A}}) d\mathbf{a}_{\mathcal{A}}, \quad (36)$$

where $\mathbf{a}_{\mathcal{A}} = \mathbf{a}(\mathbf{X}_{\mathcal{A}})$ is the vector of spatial gain at positions $\mathbf{x}_i \in \mathcal{A}$, $i = \{1, 2, \dots, M\}$ and $p(\cdot)$ is its pdf.

According to [21], solving (35) is NP-hard. Therefore, a greedy approach can be used to find a near optimal solution to this problem [22], [23]. In the greedy approach, we assume that K sensors have already been placed at position $\mathcal{A}_K = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$, and we add a single sensor ($N = 1$) leading to a total number of sensors of $M = K + 1$. Taking into account the Gaussian assumption on the model of the spatial gain as in (7), the conditional entropy is presented as follows:

$$H(a_N | \mathbf{a}_K) = \frac{1}{2} \log C_{N|K}^a + \frac{1}{2} (\log(2\pi) + 1), \quad (37)$$

where, $C_{N|K}^a$ is defined in (28). This criterion measures the uncertainty between the spatial gain at the new position a_N and the spatial gain at the previously selected positions \mathbf{a}_K , and it is preferred to choose the location that maximizes this quantity. Therefore, by defining $J_H(\mathbf{x}_N | \mathbf{a}_K) \triangleq H(a_N | \mathbf{a}_K)$, the next sensor position given by this criterion is the following:

$$\hat{\mathbf{x}}_N = \underset{\mathbf{x}_N \in \mathcal{S} \setminus \mathcal{A}_K}{\operatorname{argmax}} J_H(\mathbf{x}_N | \mathbf{a}_K). \quad (38)$$

The disadvantage of the above criterion is that the decision for sensor placement does not depend on the measurements

collected by the sensors, and it only depends on the prior knowledge on the covariance function of the model of the spatial gain, *i.e.* C^a . Also, it is shown that the criterion attempts to push the sensors far from each other and scatters them in the boundaries of the space which causes the information to be lost and not be able to have a better estimation of the physical phenomenon [24]. This is due to the fact that the criterion just takes into account the entropy of the selected positions and it does not observe the entire space of the interest.

2) *Mutual information (MI)*: Another way to tackle the problem of optimal sensor placement which resolves the weak point of the entropy is to use mutual information. In [12], it is suggested to maximize $MI(\mathbf{a}_{\mathcal{A}}, \mathbf{a}_{\mathcal{S} \setminus \mathcal{A}})$ which is the mutual information between the selected locations for sensor placement \mathcal{A} and the rest of the space $\mathcal{S} \setminus \mathcal{A}$:

$$\mathcal{A} = \underset{\mathcal{A} \subset \mathcal{S}; |\mathcal{A}|=M}{\operatorname{argmax}} J_{MI}(\mathbf{X}_{\mathcal{A}}), \quad (39)$$

where $J_{MI}(\mathbf{X}_{\mathcal{A}}) = H(\mathbf{a}_{\mathcal{S} \setminus \mathcal{A}}) - H(\mathbf{a}_{\mathcal{S} \setminus \mathcal{A}} | \mathbf{a}_{\mathcal{A}})$. In [12], it is shown that this criterion outperforms the entropy criterion in (35). It is also proved that solving (39) is NP-hard, and therefore, a greedy algorithm is suggested to obtain an approximate solution. In this regard, it is proposed to add the sensors one by one, such that the increase of the mutual information is maximized. So, assuming that K sensors have already been located at \mathbf{X}_K , and the new single sensor is supposed to be located at \mathbf{x}_N in such a way that $MI(\mathbf{a}_{K+N}) - MI(\mathbf{a}_K)$ is maximized, then we have the following:

$$\hat{\mathbf{x}}_N = \underset{\mathbf{x}_N \in \mathcal{S} \setminus \mathcal{A}_K}{\operatorname{argmax}} J_{MI}(\mathbf{x}_N | \mathbf{a}_K), \quad (40)$$

where, $J_{MI}(\mathbf{x}_N | \mathbf{a}_K) \triangleq H(a_N | \mathbf{a}_K) - H(a_N | \mathbf{a}_{\mathcal{S} \setminus \mathcal{L}})$, and $\mathcal{L} = \{\mathbf{X}_K \cup \mathbf{x}_N\}$. Under the Gaussian assumption for the spatial gain, the above criterion is simplified as follows:

$$\mathbf{x}_N = \underset{\mathbf{x}_N \in \mathcal{S} \setminus \mathcal{A}_K}{\operatorname{argmax}} \frac{C_{NN}^a - C_{NK}^a (C_{KK}^a)^{-1} C_{KN}^a}{C_{NN}^a - C_{N\mathcal{L}}^a (C_{\mathcal{L}\mathcal{L}}^a)^{-1} C_{\mathcal{L}N}^a}. \quad (41)$$

B. Expected SNR

An alternative solution to deal with the uncertainty on the spatial gain is to consider the expected SNR as a sensor placement criterion [15]:

$$J_E(\mathbf{X}_M) = \mathbb{E}[W_M]. \quad (42)$$

From Section II, we know that W_M is a sum of M independent random variables Γ_i distributed according to (16). Therefore, the criterion above is given by the sum of the means of Γ_i . A simpler way to derive the criterion is to directly evaluate the average of the quadratic form (8) using the GP prior on $\hat{\mathbf{a}}(x)$:

$$\begin{aligned} J_E(\mathbf{X}_M) &= \mathbb{E}[\hat{\mathbf{a}}_M^T \mathbf{R}_{MM} \hat{\mathbf{a}}_M] \\ &= \mathbf{m}_M^T \mathbf{R}_{MM} \mathbf{m}_M + \operatorname{Tr}(\mathbf{R}_{MM} \mathbf{C}_{MM}^a) \\ &= \operatorname{Tr}[\mathbf{R}_{MM} (\mathbf{m}_M \mathbf{m}_M^T + \mathbf{C}_{MM}^a)]. \end{aligned} \quad (43)$$

Note that here both the greedy and sequential approaches can be used as well. By considering that K sensors have already been located at positions \mathbf{X}_K , that gain information \mathbf{z}_K is

available and that we want to add N new sensors at \mathbf{X}_N , the criterion from [15] can be written as follows:

$$J_E(\mathbf{X}_N|\mathbf{z}_K) = \mathbb{E}[\hat{\mathbf{a}}_M^T \mathbf{R}_{MM} \hat{\mathbf{a}}_M | \mathbf{z}_K] \\ = \text{Tr}[\mathbf{R}_{MM}(\mathbf{m}_{M|K} \mathbf{m}_{M|K}^T + \mathbf{C}_{M|K}^a)], \quad (44)$$

where $\mathbf{m}_{M|K}$ and $\mathbf{C}_{M|K}^a$ are, respectively, the conditional mean and covariance given by (23) and (24).

C. Discussion on different criteria

In [15] it is shown that maximizing J_E , *i.e.* the expectation of the SNR, outperforms the classical kriging approaches based on the entropy J_H and the mutual information J_{MI} criteria for source estimation. This is clearly due to the fact that the J_E criterion targets the SNR instead of the spatial gain for sensor placement. If we inspect expressions (43) and (44) for the expected SNR criterion, a remarkable difference with respect to J_H and J_{MI} is the presence of the mean of the gain. This difference may make the optimal sensor positions generated with J_E differ substantially from those generated with J_{MI} or J_H . For example, when the uncertainty on the gain is small (σ_a is small), $\mathbf{C}_{M|K}^a$ can be neglected. In this case, to maximize (44), the sensors need to be placed far enough, so that \mathbf{R}_{MM} is close to a diagonal matrix. At the same time the elements of \mathbf{m}_K and $\mathbf{m}_{M|K}$ should have large absolute values, so that the optimal positions will be spread on \mathcal{X} but will be biased to the positions where the gain is large in absolute value. In the same situation, criteria J_H and J_{MI} will simply spread the positions in \mathcal{X} in such a way that the gains on the sensors will be maximally uncorrelated.

It is expected that J_P (33) outperforms J_H and J_{MI} for signal extraction, since it targets directly the SNR. As a downside, J_P is generally more complex to evaluate than other criteria, including J_E , since it requires numerical evaluation of the cdf of a quadratic form of a Gaussian vector. The only case for J_P which is comparable in terms of complexity to the other methods is either when $M = 1$ or $N = 1$. In these cases, J_P is given in closed-form. Another drawback of J_P when compared to J_E is that it relies on the Gaussian assumption on the gain. For J_E , the GP model assumption on $a(x)$ is not necessary, it is only required that $a(\mathbf{x})$ is a stochastic process with mean $m^a(\mathbf{x})$ and covariance $C^a(\mathbf{x}, \mathbf{x}')$.

Although in terms of computational complexity, J_E is simpler than J_P , in terms of robustness against uncertainty it may be inferior. The uncertainty of the SNR is summarized into its mean and by increasing the uncertainty of the spatial gain, the robustness of this criterion is decreased. We can see this behavior by inspecting (43) and (44). When the uncertainty becomes remarkable, the term depending on $\mathbf{C}_{M|K}^a$ will bias the chosen locations to have large variance on their gains, thus generating positions which may lead to an uncertain SNR.

V. NUMERICAL RESULTS

Experiments concerning the robustness of the proposed criterion J_P and of the criteria from the literature J_E , J_H and J_{MI} are presented in this section¹. We first discuss about

¹All the simulations in this section are done in MATLAB-R2018b on operating system macOS version 10.14.3, with processor 3.2 GB Intel Core i5 and memory 8 GB 1600 MHz DDR3.

the concept of robustness and how to measure it. Then, the effect of θ in J_P is shown. Finally, we compare the different criteria in terms of robustness and output SNR.

A. Simulation setup

In the numerical experiments, we consider a one dimension grid $\mathbf{X}_T = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$ in the normalized range $\mathbf{x}_i \in [0, 1]$ where $i = 1, 2, \dots, T$ for possible sensor locations. Note that to keep the consistency of the paper, in the experimental part we represent the scalars with bold lowercase letters. Depending on the smoothness of the signals, the size of the grid T and the number of initial sensors K , as well as their positions \mathbf{X}_K will be changed in different simulation parts. In all cases, we assume that $\mathbf{m}_K = \mathbf{a}_K + \mathbf{b}_K + \bar{\mathbf{u}}_K$, where \mathbf{m}_K is the mean of the spatial gain which will be used as the estimation of the spatial gain, \mathbf{a}_K is the true value of the spatial gain, \mathbf{b}_K is a bias in the spatial gain, and $\bar{\mathbf{u}}_K$ is the uncertainty of the spatial gain. Note that, the sum of the uncertainty and the bias represent the error \mathbf{v}_K in equation (22) *i.e.* $\mathbf{v}_K = \mathbf{b}_K + \bar{\mathbf{u}}_K$. We use GP models $\mathcal{GP}(m(\mathbf{x}), C(\mathbf{x}, \mathbf{x}'))$, with a square exponential covariance function $C(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp(-(\mathbf{x} - \mathbf{x}')^2 / (2\rho^2))$ to produce the spatial gain $a(\mathbf{x})$, the noise $n(\mathbf{x})$, and the uncertainty $\bar{u}(\mathbf{x})$. Here, ρ is a smoothness parameter where a small ρ means fast spatial changes, while a large ρ means smooth changes. Also, the bias $b(\mathbf{x})$ is generated using a scaled GP such that the ratio between the bias and the spatial gain at each position remains intact for each Monte-Carlo realization. Note that when the GP is scaled in this way, it is not a GP anymore. The subscripts $(\cdot)_a$, $(\cdot)_n$, $(\cdot)_b$, and $(\cdot)_u$ refer to the GP parameters of the spatial gain, the noise, the bias and the uncertainty of the spatial gain, respectively. The smoothness parameters ρ_a , ρ_n , ρ_b , and ρ_u , and the variances σ_a , σ_n , σ_b , and σ_u will be changed in different parts of the numerical experiments. We set all the mean functions $m(\mathbf{x})$ to be equal to 0. Note that, in practice, we do not have the actual spatial gain $a(\mathbf{x})$, and in our simulations, as an oracle, we generate randomly one realization from a GP with zero-mean and covariance parameters σ_a and ρ_a .

To implement the proposed algorithm $J_P(\mathbf{X}_N, \theta | \hat{\mathbf{a}}_K)$, it is required to choose a relevant value for the parameter θ . So, we suggest the following model:

$$\theta = \widehat{SNR}(\hat{\mathbf{f}})^{(K)} + \delta \left(\widehat{SNR}(\hat{\mathbf{f}})^{(K+N)} - \widehat{SNR}(\hat{\mathbf{f}})^{(K)} \right). \quad (45)$$

In the above equation, by using (26), $\widehat{SNR}(\hat{\mathbf{f}})^{(K)}$ and $\widehat{SNR}(\hat{\mathbf{f}})^{(K+N)}$ are the mean of the SNR before and after adding the new N sensors, respectively. Setting θ according to the above model helps us to use a meaningful value for this parameter, which is relative to the initial value of the SNR added by a coefficient of the increase in the SNR after adding new sensors. In the rest of the simulation part, we parameterize J_P as a function of δ instead of θ .

B. FR and Failure[%] as measures for the uncertainty effect

Although J_E and J_P are criteria targeting SNR, uncertainty on $a(\mathbf{x})$ may lead to large values of these criteria at positions

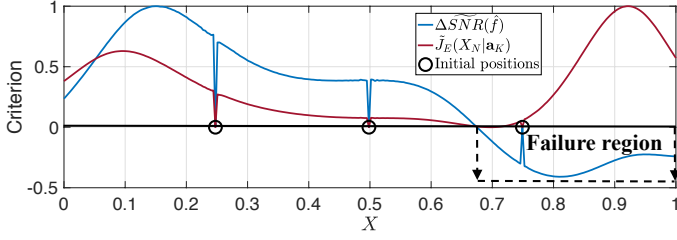


Fig. 1: Failure region is where the true value of the SNR with estimated extraction vector $\hat{\mathbf{f}}$ is smaller than its initial value before adding the new sensors. Here, the mean of the SNR is used as the criterion depicted with a red color. The blue curve represents the variation of the true SNR with estimated $\hat{\mathbf{f}}$. The tilde superscript represents the normalization of the criteria.

where in practice the value of the true output $\text{SNR}(\hat{\mathbf{f}}|\mathbf{X}_M)$ may be decreased. The true SNR is given by (4), where $\mathbf{f} = \hat{\mathbf{f}}$ is the estimated extraction vector. This vector is given by $\hat{\mathbf{f}} = \mathbf{R}_{MM}\mathbf{m}_M$ in the non sequential approach and by $\hat{\mathbf{f}} = \mathbf{R}_{MM}\mathbf{m}_{M|K}$ in the sequential approach. As a consequence, the true output SNR can be rewritten as

$$\text{SNR}(\hat{\mathbf{f}}|\mathbf{X}_M) = \frac{\sigma_s^2 \hat{\mathbf{f}}^T \mathbf{a}_M \mathbf{a}_M^T \hat{\mathbf{f}}}{\hat{\mathbf{f}}^T \mathbf{C}_{MM}^n \hat{\mathbf{f}}}. \quad (46)$$

Note that (46) differs from the estimated SNR, $\widehat{\text{SNR}}$ (8), since this true SNR (46) depends on the true spatial gain \mathbf{a}_M , contrary to $\widehat{\text{SNR}}$ which depends on the estimation of the spatial gain $\hat{\mathbf{a}}_M$. In the rest, the set of positions that deteriorates the true SNR is called *failure region* (FR).

Fig. 1 is an example to understand the notion of FR. In this figure, the size of the grid is 300, and $K = 3$ sensors have already been located at the positions $\mathbf{X}_K = \{0.25, 0.5, 0.75\}$, which are marked by circles. Here, we considered an unbiased situation *i.e.* $b(\mathbf{x}) = 0$, the variances are set to be $\sigma_a = 0.15$, $\sigma_u = 0.15$, $\sigma_n = 0.5$, and the smoothness parameters are $\rho_a = 0.2$, $\rho_u = 0.2$, $\rho_n = 0.1$. The standard deviation of the source signal is $\sigma_s = 2$. We use J_E (44) as the placement criterion. Now, the aim is to find the best location for the 4th sensor. In this figure, the tilde superscript indicates that the function is normalized with its maximum value to be equal to one, and the initial value equal to zero. The normalized variation of SNR, denoted $\widetilde{\Delta \text{SNR}}(\hat{\mathbf{f}})$, is as follows:

$$\widetilde{\Delta \text{SNR}}(\hat{\mathbf{f}}) = \frac{\text{SNR}(\hat{\mathbf{f}}) - \text{SNR}_{init}(\hat{\mathbf{f}})}{\text{SNR}_{max}(\hat{\mathbf{f}}) - \text{SNR}_{init}(\hat{\mathbf{f}})}, \quad (47)$$

where $\text{SNR}_{init}(\hat{\mathbf{f}})$ represents the initial value of the SNR before adding the new sensor, and $\text{SNR}_{max}(\hat{\mathbf{f}})$ is the maximum value of the SNR after adding the new sensor. Note that in this paper, the points of interest in the criteria are the locations of the maxima (and not their amplitudes), which are not affected by the normalization. In Fig. 1 it is seen that, due to the uncertainty, the normalized variation of the true SNR *i.e.* $\widetilde{\Delta \text{SNR}}(\hat{\mathbf{f}})$ can take negative values at some regions, which means that by placing sensors at these positions, not only the SNR will not be improved, but it will even take smaller values compared to the

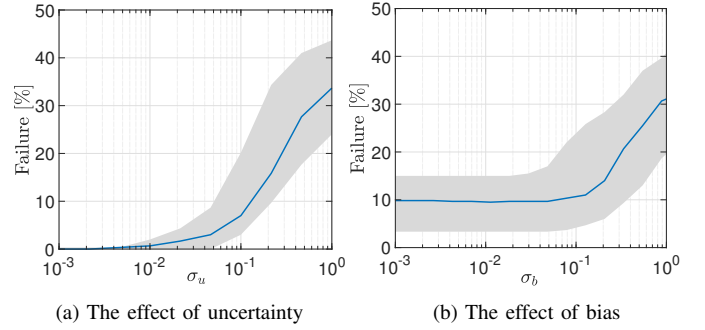


Fig. 2: Failure [%] is a measure to compute the ratio of the sensor placement region which is in the failure region. (a) In this figure, the effect of the uncertainty level on the Failure [%] is depicted. Here, the bias is considered to be equal to 0. (b) The effect of bias on Failure [%] is depicted in this figure. The variance of the uncertainty is set equal to $\sigma_u = 0.01$.

initial SNR. Therefore, since $\text{SNR}(\hat{\mathbf{f}}) < \text{SNR}_{init}(\hat{\mathbf{f}})$ in FR, the numerator of (47) becomes negative, and so, $\widetilde{\Delta \text{SNR}}(\hat{\mathbf{f}})$ becomes negative. In this figure, FR is marked with dashed arrows. As depicted, although the true SNR takes quite smaller value in FR, the suggested criterion takes significant values at these locations. In particular, the criterion takes its maximum value in a point \mathbf{x} inside FR, and so, a failure has happened due to the uncertainty. So, this is an example of the lack of robustness of this criterion.

To analyze the affect of the uncertainty and the bias in the model of the spatial gain on the size of the FR, we need to define a quantitative statistical measure. To do so, we can simulate N_{MC} Monte-Carlo realizations of the gain, bias, uncertainty and noise GP, and then count the total number of positions within the FR, here denoted N_{FR} . For a total size of spatial grid N_T , we define the failure rate as follows:

$$\text{Failure} [\%] \triangleq \frac{\text{Size of the FR}}{\text{Total size of the spatial grid}} = \frac{N_{FR}}{N_T}. \quad (48)$$

Fig. 2 is provided to illustrate the effect of uncertainty and bias on Failure[%]. In this figure, three initial sensors are placed at $\mathbf{X}_K = \{0.25, 0.5, 0.75\}$, and we place the 4th sensor using J_E . The parameters are set as $\rho_a = \rho_u = 0.2$, $\rho_n = 0.01$, $\sigma_a = 0.5$ and $\sigma_n = 0.01$. We generate $N_{MC}^a = 10$ realizations for the spatial gain, and for each realization, we consider 50 runs for the bias, uncertainty and noise ($N_{MC}^b = N_{MC}^u = N_{MC}^n = 50$), which leads to a total number of $N_{MC} = 500$ Monte-Carlo (MC) realizations. Then, Failure[%] is evaluated according to (48). In this figure, the effects of the bias and the uncertainty on Failure[%] are studied separately. Firstly, in Fig. 2a, we consider an unbiased situation setting $b(\mathbf{x}) = 0$, and different uncertainty levels in the interval $\sigma_u \in [10^{-3}; 10^0]$ are used. The blue curve is the average of Failure [%] over all N_{MC} realizations, and the gray shadow represents the standard deviation. As it was expected, by increasing σ_u , the average and the variance of Failure[%] increases. This experiment is repeated in Fig. 2b to study the effect of the bias. In this figure we use the previous configuration to set up the parameters, except that σ_u is kept

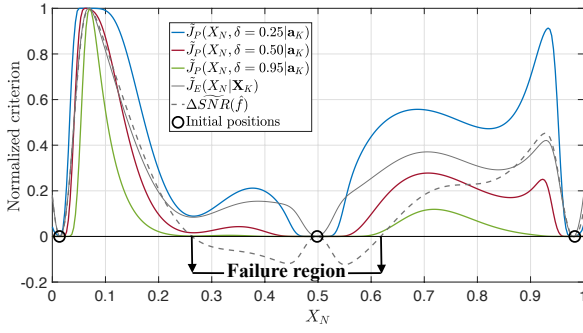


Fig. 3: Effect of parameter δ . The solid curves are the proposed criterion for different δ . The superscript tilde represents the normalization of the function. By increasing δ , the algorithm is more robust against FR. However, other candidates for sensor placement which provide significant increase in the SNR can be ignored. So, it is required to make a trade-off between reducing the risk of being in FR, and increasing the SNR.

fixed to $\sigma_u = 0.01$, and the level of the bias is changed in the range $\sigma_b \in [10^{-3}; 10^0]$. As it can be seen, in average, the effect of the bias is negligible up to $\sigma_b \simeq 0.1$. If the level of bias goes beyond this value, the effect of the bias on Failure[%] becomes significant, which means that the suggested model for the spatial gain is not appropriate, and it is required to apply better methods to provide a good approximation for the spatial gain. Since our proposed method is only focused on the uncertainty of the spatial gain, henceforth we assume that a suitable model is used for the spatial gain and we set $b(x) = 0$.

C. Effect of δ on the criterion J_P

In this part, we study the effect of δ (which controls θ) on the results obtained with J_P . To do so, Fig. 3 is provided. In this figure, the parameters are set as follows: $\rho_a = 0.2$, $\rho_u = 0.2$, $\rho_n = 0.1$, $\sigma_a = 0.15$, $\sigma_u = 0.15$, $\sigma_n = 0.5$, and $\sigma_s = 2$, with the size of the spatial grid equal to 300. Starting with three initial sensors at $\mathbf{X}_K = \{0.05, 0.5, 0.95\}$, we look for the best position for the 4th sensor. The proposed criterion is used as the target function with three different values of δ (0.25, 0.5, and 0.95), which are depicted in Fig. 3 with a blue, red, and green color, respectively. Moreover, $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ which is the output SNR normalized according to (47) is depicted with a dashed curve. The plots for $J_E(\mathbf{x}_N|\mathbf{X}_K)$ and $J_P(\mathbf{x}_N, \delta|\mathbf{X}_K)$ are normalized such that their maximum values be equal to 1, and the initial values equal to 0. The normalized forms are denoted by a superscript tilde. Note that the locations of maxima for the different criteria will not be changed by the normalization.

As depicted in Fig. 3, the proposed criterion behaves differently according to δ . The larger this parameter is, the smaller the values of the criterion within FR are. It is noticeable that by increasing δ , besides avoiding FR, it is probable to avoid some positions with significant increase in the SNR (e.g. for $\mathbf{x} \approx 0.95$). Consequently, high values of δ should be used to avoid locating the new sensor in FR, with the cost of achieving a smaller amount of increase in the SNR. Otherwise, we can

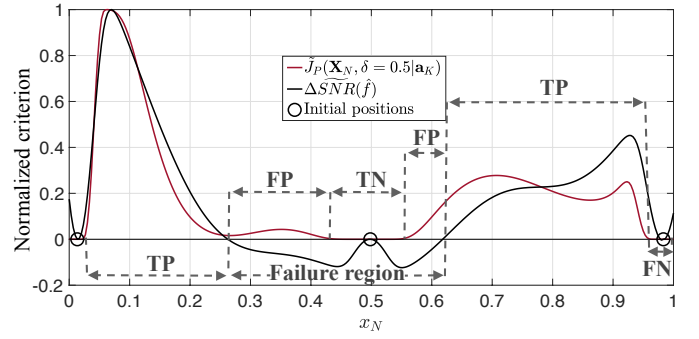


Fig. 4: Region classification: Positive (P) vs. Negative (N), and True (T) vs. False (F). P and N represent the regions where $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ is positive and negative, respectively. T is a situation when the criterion and $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ are both positive, or the criterion is zero and $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ is negative. F is when the criterion is positive while $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ is negative, or the criterion is zero whenever $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ is positive.

decrease δ to keep most of the positions with a significant increase in the SNR, but this leads to an increased risk of having large values of the criterion for positions in FR. So, by choosing an appropriate δ , we can make a trade-off between avoiding positions in the FR and keeping the regions with a high increase in the SNR.

D. Effect of the smoothness of the spatial gain and noise correlation length-scale on robustness

In this part, we study the influence of the ρ_a and ρ_n on the robustness of the criteria. To do so, we first classify different regions of the space according to Fig. 4. The parameter configuration in this figure are as follows: $\rho_a = 0.2$, $\rho_u = 0.2$, $\rho_n = 0.1$, $\sigma_a = 0.15$, $\sigma_u = 0.15$, $\sigma_n = 0.5$, $\sigma_s = 2$, and $\delta = 0.5$. The size of the spatial grid is 300, with three initial sensors at $\mathbf{X}_K = \{0.05, 0.5, 0.95\}$. According to this figure, we name the regions with positive $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ to be positive (P), and the regions with negative $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ to be negative regions (N). Here, $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ denotes the normalized output SNR defined in (47). Accordingly, the true positive (TP) is the size of the region where both $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ and the criterion are positive. The number of positions that $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ is negative but the criterion takes a positive value is the so-called false positive (FP). The notation true negative (TN) corresponds to the size of the region where $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$ is negative, and the criterion is zero. Finally, false negative (FN) is related to the position with positive $\widetilde{\Delta SNR}(\hat{\mathbf{f}})$, and zero value for the underlying criterion. Now, according to this region classification, and under different combinations of the values of ρ_a and ρ_n , we calculate the false positive rate (FPR) of a criteria as follows:

$$FPR = \frac{\text{false positive}}{\text{total number of negatives}} = \frac{FP}{TN + FP}. \quad (49)$$

In above, the denominator is actually the size of the FR. It is noticeable that if the size of the FR is small, although the

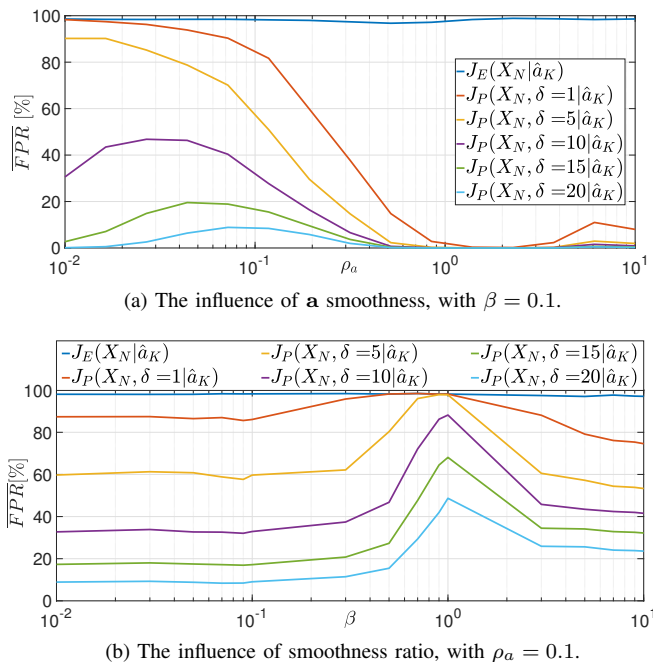


Fig. 5: The effect of the smoothness of the spatial gain and the noise on the robustness. (a) Here, different smoothness values of the spatial gain is considered, and the smoothness ratio between the noise and the spatial gain is set as $\beta = \rho_n/\rho_a = 0.1$. As the spatial gain gets closer to a spatial white noise, $FPR[\%]$ increases. (b) Here, the effect of the noise smoothness is studied. The more similar smoothness degree of the noise and the spatial gain, the larger $FPR[\%]$ is. Compared to the spatial gain, as the noise gets closer to a white noise (decreasing β), or smooth noise (increasing β), $FPR[\%]$ decreases.

FPR gets large, the probability of selecting a position in FR is small. Conversely, if the size of the FR is large, although FPR gets small, the probability of selecting a sensor in FR can be large. To avoid these two marginal cases, according to Fig. 2, we set $\sigma_u \in [0.1, 0.8]$ which provides a moderate Failure[%] between 8% and 30%.

Fig. 5a is provided to show the robustness of the proposed criterion $J_P(\mathbf{x}_N, \delta | \mathbf{a}_K)$ in terms of the average $FPR[\%]$ and compares it with the criterion $J_E(\mathbf{x}_N | \mathbf{a}_K)$ for different smoothness conditions ρ_a . In this figure, three initial sensors are used at $X_K = \{0.05, 0.5, 0.95\}$ in a spatial grid of size 300, and we look for the 4th sensor position. The variances $\sigma_s = \sigma_a = \sigma_u = \sigma_n$ are all set to be equal to 1. The smoothness parameter ρ_a takes 15 different values in the interval $[10^{-2}, 10]$ in the logarithmic scale, and for each case, we set $\rho_u = \rho_a$. The smoothness ratio between the spatial gain and the noise is fixed to be $\beta = \rho_n/\rho_a = 0.1$. The total number of $N_{MC} = 500$ realizations of the spatial gain, the mean, the uncertainty and the noise is considered. Note that the x -axis has a logarithmic scale in Fig. 5. $\overline{FPR}[\%]$ represents the average value of $FPR[\%]$ over the whole Monte-Carlo realizations. As depicted in this figure, the proposed criterion is more robust compared to $J_E(\mathbf{x}_N | \mathbf{a}_K)$. This figure shows that the larger the parameter δ is, the more robust the algorithm

is against uncertainty for different ρ_a .

Now, the effect of the noise smoothness is shown in Fig. 5b. In this figure, we set $\rho_a = 0.1$, and the values of the smoothness ratio $\beta = \rho_n/\rho_a$ are sampled in the interval $[10^{-2}, 10]$, thus leading to different values of ρ_n . As it is seen, again the proposed method shows a better performance as the smoothness of the noise is varied. Note that, for the proposed method, $\overline{FPR}[\%]$ takes its maximum value when the smoothness of the noise gets close to the smoothness of the spatial signal *i.e.* $\beta = \rho_n/\rho_a = 1$. It is due to the fact that the smoothness similarity between the noise and the propagated source signal makes it difficult to separate them. It is also interesting to mention that for larger values of β , where the noise tends to be close to a constant function, the $\overline{FPR}[\%]$ starts to decrease. This result shows that whenever the noise is significantly different from the signal in terms of the smoothness, the $\overline{FPR}[\%]$ takes smaller values.

E. Controlling the trade-off between robustness and average SNR maximization

We noted earlier that by increasing the δ parameter, the proposed method becomes more robust against the uncertainty of the spatial gain, and the $FPR[\%]$ decreases. However, making the risk as small as possible, *i.e.* being too much strict against FR, may cause some good local maxima to be ignored, and consequently the algorithm becomes incapable to detect positions that provide a good output SNR. Therefore, it is important to make a trade-off between reducing the $FPR[\%]$ and increasing the output SNR for choosing the parameter δ .

In this section, we study the effect of δ on the output SNR and FPR. To this aim, in Fig. 6 we compare the behaviour of the proposed method in terms of FPR and the improvement of the SNR under different uncertainty conditions and different selections of δ . In this figure, we used a grid of size 100 between 0 and 1 for the sensor positions. An initial sensor is considered in the middle of the grid at $\mathbf{x}_1 = 0.5$, and we added 5 sensors using a greedy approach on the proposed criterion. The parameters are set as $\sigma_a = 1$, $\rho_a = 0.1$, $\sigma_s = 1$, $\sigma_n = 1$ and $\beta = 0.2$. The variance of the uncertainty for the spatial gain varies as follows: $\sigma_u \in \{0, 0.1, 0.2, \dots, 1\}$. We used 10 MC realizations for $a(\mathbf{x})$ and 10 MC realizations for $m_a(\mathbf{x})$. The average FPR denoted \overline{FPR} is depicted as the blue curve with y-axis on the left, and the average improvement in the output SNR, *i.e.* $\Delta SNR(\hat{\mathbf{f}}) = SNR(\hat{\mathbf{f}}) - SNR(\hat{\mathbf{f}})^{(init)}$ is presented in orange with the y-axis on the right. Here, $SNR(\hat{\mathbf{f}})^{(init)}$ denote the true value of the SNR at the initial situation with a unique sensor located at $\mathbf{x} = 0.5$. Note that $SNR(\hat{\mathbf{f}}) - SNR(\hat{\mathbf{f}})^{(init)}$ is the variation of the SNR, which is equivalent to the nominator of (47). This experiment is repeated for three different values of δ , namely 1, and 20.

As Fig. 6 shows, by increasing the uncertainty σ_u , the FPR increases and the output improvement of SNR decreases. Furthermore, we can see that by choosing a small value for δ , the output improvement of SNR takes larger value in average, however, the FPR becomes larger, meaning that the probability of being in the failure region increases. Comparing Fig. 6-a with Fig. 6-b, we see that by increasing δ from 1 to 10,

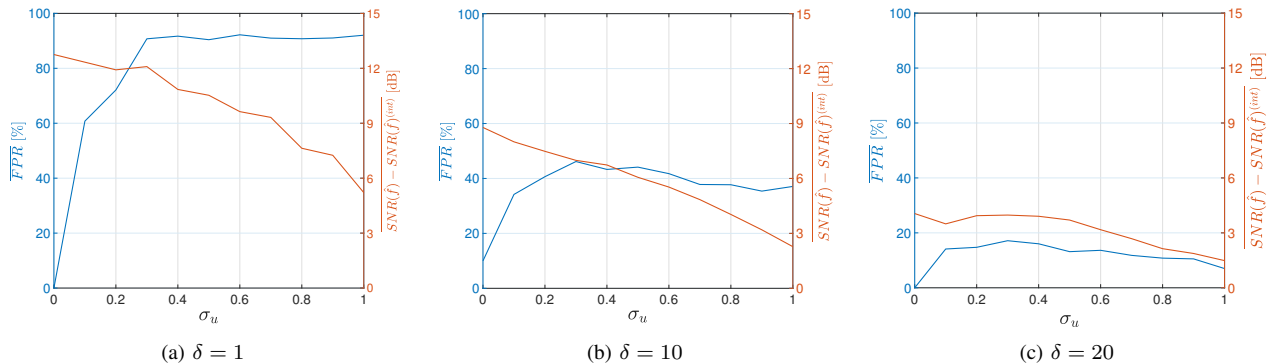


Fig. 6: The trade-off between the robustness and the maximization of the averaged SNR improvement can be controlled by changing the parameter δ . The blue curve with y-axis on the left represents the average $FPR[\%]$ according to different levels of uncertainty, and the orange curve with y-axis on the right depicts the average improvement of the SNR. From left to right, sub-figures (a), (b), and (c) show the results for different values of δ . It is seen that by increasing δ , although the average increase in the SNR gets smaller, $FPR[\%]$ gets smaller too. Depending on the application, and the level of uncertainty σ_u , we need to set δ , such that both $FPR[\%]$ and the average improvement in the SNR are acceptable.

the average FPR becomes smaller for different values of σ_u . However, the average improvement in the SNR decreases. Going through a more strict selection for δ , in Fig. 6-c we see that the average FPR gets quite smaller, providing a safe situation to avoid FR. Nevertheless, the improvement for the output SNR significantly decreases.

To conclude, we can say that, when the uncertainty is low, since the output SNR improvement can be large enough, it is possible to dedicate some SNR to reduce $FPR[\%]$, and so, we can choose a larger δ . In contrast, when σ_u is large, since the output SNR is affected by this uncertainty, we cannot dedicate a large part of the SNR to have a smaller value of $FPR[\%]$. Therefore, it is better not to be too strict, and select a moderate value for δ . Depending on the application, we can make a trade-off between $\Delta SNR(\hat{\mathbf{f}})$ and $FPR[\%]$ by changing the parameter δ : in sensitive applications where decreasing the SNR is not acceptable, we need to choose a large δ , even if the improvement of the output SNR is not large. On the other hand, in applications which are less sensitive to output SNR, we can take a risk of being trapped in the failure region to have a large improvement in the SNR, and, as a consequence, we can choose a smaller δ .

F. Sequential approach

In this part, we compare the proposed method $J_P(\mathbf{x}_N, \delta | \mathbf{X}_K)$ with $J_H(\mathbf{x}_N | \mathbf{X}_K)$, $J_{MI}(\mathbf{x}_N | \mathbf{X}_K)$ and $J_E(\mathbf{x}_N | \mathbf{X}_K)$ by using a sequential approach. This comparison is made in two ways: the average output SNR *i.e.* $SNR(\hat{\mathbf{f}})$, and the average $FPR[\%]$. Here, the overline represents the average value of each measure over the total Monte-Carlo realizations. For setting up the experiments, we assume that $K = 1$ initial sensor is located at $x_1 = 0.5$. Then we add sensors one by one up to $M = 10$. Note that, since in this experiment we want to add up to 10 sensors to show the improvement of the algorithms according to the number of sensors, we chose the smoothness parameter of the spatial gain *i.e.* ρ_a to be smaller than the previous

experiments. In this way, we provide more variation for the spatial gain. Therefore, the source signal cannot be recovered just by using a very few sensors *e.g.* two or three sensors, and to recover the source signal with a desired accuracy, we need to collect more information by using sufficient number of sensors, which is up to 10 sensors in this experiment. Accordingly, the parameters are set as follows: $\rho_a = \rho_u = 0.01$, $\beta = 0.5$, $\sigma_a = \sigma_n = 1$. We considered two different situations for σ_u . One is $\sigma_u = 0.1$ which represents a small value for the level of uncertainty, and the other one is $\sigma_u = 0.8$, which assumes a higher uncertainty. Depending on the level of uncertainty, we set $\delta = 13$, and $\delta = 10$ under the assumption of $\sigma_u = 0.1$ and $\sigma_u = 0.8$, respectively. Note that by reducing δ under higher uncertainty, we attempt to achieve a required minimum improvement in the average SNR, while having the same level of robustness measured by $FPR[\%]$. In this part a number of $N_{MC}^a = 10$ realizations are used for the spatial gain $\mathbf{a}(\mathbf{x})$, and for each realization we consider 10 runs for the bias, uncertainty and noise ($N_{MC}^b = N_{MC}^u = N_{MC}^n = 10$), leading to a total number of $N_{MC} = 100$ Monte-Carlo realizations. For each realization of the spatial gain, we also have a new realization for the additive noise $\mathbf{n}(\mathbf{x})$.

Fig. 7 presents the simulation results for different uncertainty levels. The first column is related to $\sigma_u = 0.1$, and the second column is for $\sigma_u = 0.8$. The first row, shows $\Delta SNR(\hat{\mathbf{f}})$ by adding new sensors, and the second row corresponds to $FPR[\%]$. In all the sub-figures, the blue, red, orange, and green curves show the performances of the proposed criterion $J_P(\mathbf{x}_N, \delta | \hat{\mathbf{a}}_K)$, and other criteria based on the mean of the SNR $J_E(\mathbf{x}_N | \hat{\mathbf{a}}_K)$, the mutual information $J_{MI}(\mathbf{x}_N | \hat{\mathbf{a}}_K)$, and the entropy $J_H(\mathbf{x}_N | \hat{\mathbf{a}}_K)$, respectively.

We begin with a low level of uncertainty $\sigma_u = 0.1$ in the first column. By looking at Fig. 7-a, it is seen that the proposed criterion significantly outperforms the other methods and provides a larger value for $\Delta SNR(\hat{\mathbf{f}})$. By sequentially adding 10 sensors, $J_P(\mathbf{x}_N, \delta | \hat{\mathbf{a}}_K)$ provides $\Delta SNR(\hat{\mathbf{f}}) = 30$ dB, which

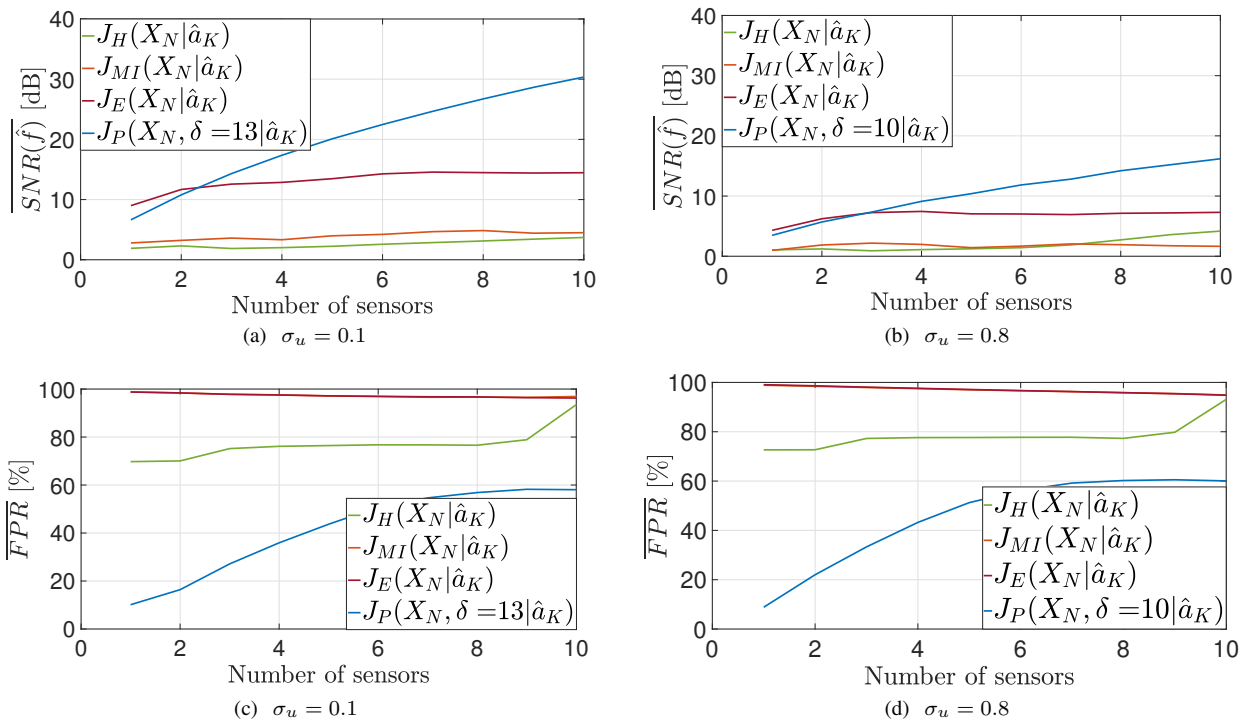


Fig. 7: In this figure, the performance of the proposed method is compared with the prior works by using a sequential approach. Two situations are studied in this figure: $\sigma_u = 0.1$ in the first column, and $\sigma_u = 0.8$ in the second column. In the top, the average output SNR is reported, and in the bottom, the average $FPR[\%]$ is presented. It is seen that both in a low and high uncertainty levels, the proposed method provides a larger SNR, as well as more robustness against the uncertainty.

still can be increased by adding more sensors. The next best criterion is $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$, which gives $\Delta SNR(\hat{\mathbf{f}}) = 15$ dB. It is seen that unlike the proposed criterion, the other criteria cannot provide a better SNR by adding more sensors since their related curves have become almost flat. This is due to the uncertainty, which is not considered in these methods. It is important to mention that by adding a few number of sensors (up to 2-3 in this example), we do not see a significant difference in the performance of the proposed criterion and the one in $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$. It may be due to the fact that for a very few number of sensors, the information is not enough to have a good recovery of the source signal. However, immediately after adding the next sensors, due to the uncertainty, the proposed criterion starts outperforming $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$. Therefore, in this case, to avoid increasing the computational complexity, one can simply start adding primary sensors using $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$. Then, by increasing the number of sensors, it is recommended to use the proposed criterion as a robust method against uncertainty. It is interesting to mention that as claimed in [15], $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ has notably a better performance compared to the classical kriging approaches $J_H(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ and $J_{MI}(\mathbf{x}_N|\hat{\mathbf{a}}_K)$.

To complete the discussion, sub-figure (c) shows the results of $FPR[\%]$. Here, we can see that the proposed method is quite robust against the uncertainty compared to the prior works. In contrast, the criteria based on the MI and the mean of the SNR do not have a robust behavior. Therefore, adding new sensors by using these two criteria cannot be useful due to the high $FPR[\%]$. Note that the curves related to $J_{MI}(\mathbf{x}_N|\hat{\mathbf{a}}_K)$

and $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ are quite close to each other. In sub-figure (c), $J_H(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ seems to perform better than $J_{MI}(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ and $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$, and it provides approximately $FPR[\%] = 70\%$ up to adding 9 sensors, with stepwise increasing changes by adding more sensors. From the blue curve, we can see that the proposed method is significantly more robust against the uncertainty. It is observed that $FPR[\%]$ starts from 10%, and slightly increases by adding the number of sensors up to about 8 sensors. Then, $FPR[\%]$ remains at its maximum value around 60%. This behavior is very promising, which tells us that by increasing the number of sensors, and consequently providing better $SNR(\hat{\mathbf{f}})$ (concluded from sub-figure (a)), we can be hopeful that $FPR[\%]$ will not go beyond 60%. However, depending on the application, this value can still be considered risky enough to avoid adding more new sensors.

To continue, we analyze the second column for $\sigma_u = 0.8$. To compare the performance of the proposed method with the previous criteria, in sub-figures (b) and (d), we seen that the proposed criterion has a superior performance compared to the other methods both in $SNR(\hat{\mathbf{f}})$ and $FPR[\%]$. Sub-figure (b) shows that by adding 10 sensors, the proposed criterion provides $SNR(\hat{\mathbf{f}}) = 16$ dB, which is 9 dB higher than the result achieved by $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$. In this part, by choosing $\delta = 10$, we try to keep the results for $FPR[\%]$ close to sub-figure (c). Comparing sub-figures (a) and (b), it is seen that by increasing the uncertainty level from $\sigma_u = 0.1$ to $\sigma_u = 0.8$, the average output SNR decreases. For instance, considering that 10 sensors are added, for the proposed criterion, $SNR(\hat{\mathbf{f}})$ drops from 30 dB in sub-figure (a) to 16 dB in sub-figure (b).

From the same point of view, the changes of the output SNR for $J_E(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ is from 15 dB to 8 dB, and for $J_{MI}(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ it is from 4 dB to 2 dB. The results for $J_H(\mathbf{x}_N|\hat{\mathbf{a}}_K)$ remain almost unchanged under different level of uncertainties.

VI. CONCLUSIONS AND PERSPECTIVES

In this paper, we studied the problem of optimal sensor placement for signal extraction. We considered a linear signal extraction and targeted the predicted output SNR for sensor placement problem. Since the SNR is a continuous random variable in our setting, we presented a robust criterion for sensor placement based on the maximization of its cdf at a target SNR value.

Depending on the chosen target SNR value in the proposed criterion, we can make a trade-off between an improvement of the SNR and the robustness of the criterion. We also showed that the proposed criterion is derived from the distribution of the SNR so that the previously proposed criterion [15] can be seen as a special case of the present study. To reduce the computational cost of evaluating the criterion, we proposed a sequential approach where new sensor locations are chosen in batches. We presented how to update this sequential version when some information on the the gains of the already placed batches of sensors is available. Numerical results showed a consistent superiority of the proposed criterion compared with classical kriging and the average SNR criteria in terms of the output SNR and robustness against the uncertainty on the model of spatial gain.

There are several future research topics regarding the proposed approach. First, we did not take into account the uncertainty on the hyperparameters of the models of the noise and spatial gain. This will be an important problem to be tackled in further studies. Second, in the sequential approach we have stated that information on the gains of the already placed sensors can be retrieved with ICA or SCA. In future work, such an approach where sensor placement is coupled with blind source separation techniques should be put into test. Next, we mostly focused on one dimensional problems that can be solved by a direct search for the optimal solution in a fine grid. If we consider higher dimensional settings, then searching in a grid will be prohibitively complex. Therefore, it will be necessary to develop a fast global optimization algorithm for maximizing our criterion. As an example, recently, a gradient-based technique is proposed in [25] that can be extended for the new criterion we proposed in this paper. Finally, we mention that, although by adding new sensors we probably have improvement in $\overline{SNR}(\hat{\mathbf{f}})$, at the same time the computational complexity of the algorithm will be increased. Therefore, it is necessary to make a trade-off between increasing the SNR and decreasing the computational complexity to set the number of sensors. Using the Akaike information criterion (AIC) [26] to provide such a balance, we can propose a criterion which contains one term representing the improvement of the SNR and another term to measure the complexity of the algorithm.

REFERENCES

[1] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J. P. Watson, "Sensor placement in municipal water networks," *Journal of Water Resources Planning and Management*, vol. 131, no. 3, May 2005.

[2] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient sensor placement optimization for securing large water distribution networks," *Journal of Water Resources Planning and Management*, vol. 134, no. 6, November 2008.

[3] C. K. Daniel, "Sensor placement for on-orbit modal identification and correlation of large space structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 2, pp. 251–259, 1991.

[4] M. Meo and G. Zumpano, "On the optimal sensor placement techniques for a bridge structure," *Engineering Structures*, vol. 27, no. 10, pp. 1488–1497, August 2005.

[5] S. Y. Chen and Y. F. Li, "Automatic sensor placement for model-based robot vision," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 393–408, February 2004.

[6] H. J. Hermens, B. Frenriks, C. Disselhorst-Klugb, and G. Rau, "Development of recommendations for SEMG sensors and sensor placement procedures," *Journal of Electromyography and Kinesiology*, vol. 10, no. 5, pp. 361–374, October 2000.

[7] A. L. D. Siqueira Junior, A. F. Spirandeli, R. Moraes, and V. Zarzoso, "Respiratory waveform estimation from multiple accelerometers: An optimal sensor number and placement analysis," *IEEE journal of biomedical and health informatics*, 2018.

[8] D. G. Krige, "A statistical approach to some basic mine valuation problems on the witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, no. 6, pp. 119–139, 1951.

[9] N. Cressie, "The origins of kriging," *Mathematical Geology*, vol. 22, no. 3, pp. 239–252, 1990.

[10] T. J. Santner, B. J. Williams, and W. I. Notz, *The design and analysis of computer experiments*, Springer Science & Business Media, 2013.

[11] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[12] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[13] M. C. Shewry and H. P. Wynn, "Maximum entropy sampling," *Journal of applied statistics*, vol. 14, no. 2, pp. 165–170, 1987.

[14] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical science*, pp. 409–423, 1989.

[15] F. Ghayem, B. Rivet, C. Jutten, and R. Cabral Farias, "Optimal sensor placement for signal extraction," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4978–4982, May 2019.

[16] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, vol. 1, MIT press Cambridge, 2006.

[17] A. Nuttall and H. Albert, "Some integrals involving the q-m function," *Transactions on Information Theory*, vol. 21, no. 1, pp. 95 – 96, 1975.

[18] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[19] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, 2001.

[20] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker, "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments," *Journal of the American Statistical Association*, vol. 86, no. 416, pp. 953–963, 1991.

[21] C. Ko, J. Lee, and M. Queyranne, "An exact algorithm for maximum entropy sampling," *Operations Research*, vol. 43, no. 4, pp. 684–691, 1995.

[22] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.

[23] N. A. C. Cressie, "Statistics for spatial data," 1991.

[24] N. Ramakrishnan, C. Bailey-Kellogg, S. Tadepalli, and V.N. Pandey, "Gaussian processes for active data mining of spatial aggregates," *Proceedings of the 2005 SIAM International Conference on Data Mining*, p. 12, 2005.

[25] F. Ghayem, B. Rivet, C. Jutten, and R. Cabral, "Gradient-based algorithm with spatial regularization for optimal sensor placement," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2020.

[26] H. Akaike, "Information theory and an extension of the maximum likelihood principle," *Selected Papers of Hirotugu Akaike*, , no. 1, pp. 199–213, 1998.