Optimal sensor placement for signal extraction

Fatemeh Ghayyem

UNIVERSITÉ GRENOBLE ALPES

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE ALPES

Spécialité : **SIGNAL IMAGE PAROLE TELECOMS**

Arrêté ministériel : 7 août 2006

Présentée par
Fatemeh GHAYYEM

Thèse dirigée par **Christian JUTTEN, Professeur des Universités, Université Grenoble Alpes** et
co-encadrée par **Bertrand RIVET, Université Grenoble Alpes**

préparée au sein du
Laboratoire Grenoble Images Parole Signal Automatique
dans **l'École Doctorale Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**

Positionnement optimal de capteurs pour l'estimation du signal

Optimal sensor placement for signal extraction

Thèse soutenue publiquement le **21 October 2020**,
devant le jury composé de :

Jocelyn CHANUSSOT
Professeur des Universités, Université Grenoble Alpes, Président du jury
Saïd MOUSSAOUI
Professeur des Universités, Ecole Centrale de Nantes, Rapporteur
Pierre CHAINAIS
Professeur des Universités, Ecole Centrale de Lille, Rapporteur
Nadège THIRION-MOREAU
Professeur des Universités, Université du Sud Toulon , Examinatrice
David BRIE
Professeur des Universités, Université de Lorraine, Examinateur
Rodrigo CABRAL FARIAS
Maître de conférences, Université Côte d'Azur, invité

UNIVERSITÉ DE GRENOBLE ALPES

ÉCOLE DOCTORALE EEATS

Electronique, Electrotechnique, Automatique, Traitement du Signal

T H È S E

pour obtenir le titre de

docteur en sciences

de l'Université de Grenoble Alpes

Mention : Signal, Image, Parole, Télécoms

Présentée et soutenue par

Fatemeh GHAYYEM

Positionnement optimal de capteurs pour l'estimation du signal

Thèse dirigée par Christian JUTTEN et Bertrand RIVET

préparée au laboratoire Grenoble Image Parole Signal Automatique (GIPSA-lab)

21 Octobre 2020

Jury :

Président :	Jocelyn CHANUSSOT	-	Université Grenoble Alpes
Rapporteurs :	Saïd MOUSSAOUI	-	Ecole Centrale de Nantes
	Pierre CHAINAIS	-	Ecole Centrale de Lille
Examinateur :	Nadège THIRION-MOREAU	-	Université du Sud Toulon
	David BRIE	-	Université de Lorraine
Directeur :	Christian JUTTEN	-	Université Grenoble Alpes
Co-encadrant :	Bertrand RIVET	-	Université Grenoble Alpes
Invité :	Rodrigo CABRAL FARIAS	-	Université Côte d'Azur

Acknowledgment

First of all, I would like to express my sincere thanks to my esteemed supervisors, Prof. Christian JUTTEN, Dr. Bertrand RIVET, and Dr. Rodrigo CABRAL FARIAS for all their kind support and encouragement. During this Ph.D. project, I learned a lot from them, both scientifically and ethically. Also, my especial thanks goes to Prof. Massoud Babaie-Zadeh for his valuable and compassionate guidance.

Next, I wish to acknowledge the jury members, Prof. Jocelyn CHANUSSOT, Prof. Saïd MOUSSAOUI, Prof. Pierre CHAINAIS, Prof. Nadège THIRION-MOREAU, and Prof. David BRIE for their precious time and concerns in evaluating my Ph.D. thesis.

I would like to thank my beloved husband Mostafa, who is always an inspiration for me to pursue my scientific goals, and believe in my capabilities. Also, I thank my brilliant parents and sisters, who always warm my heart with their encouragement and presence next me.

In the end, I appreciate my adorable friends in GIPSA-LAB, who created a kind and happy atmosphere with so many unforgettable and sweet memories.

This Ph.D. thesis is dedicated to my wonderful husband, Mostafa.

Table des matières

Table des figures

Liste des tableaux

Table des sigles et acronymes

BLUE	*best linear unbiased estimators*
cdf	*cumulative density function*
FI	*Fisher information*
FPR	*false positive rate*
GP	*Gaussian process*
Hrm	*harmonic mean*
MI	*mutual information*
pdf	*probability density function*
Rcl	*recal*
SNR	*signal to noise ratio*
Spc	*specificity*

Mathematical notations and tools

To keep the consistency of the notation throughout the thesis, some of the common mathematical notations are explained here.

The $M \times M$ identity matrix is denoted by $\mathbf{I}_M$. Considering the elements of the identity matrix to be I_{ij}, then, $I_{ij} = 1$ if $i = j$, otherwise, $I_{ij} = 0$. Whenever the dimensions of the identity matrix are not concerned, it is simply denoted by $\mathbf{I}$.

Vectors are denoted by lower case bold letters such as $\mathbf{x}$, and all vectors are assumed to be column vectors. Uppercase bold letters, such as $\mathbf{X}$, denote matrices. The i^{th} element of a vector is indicated by lowercase, subscript-i *e.g.* x_i, and the $(i,\ j)^{th}$ element of a matrix is represented by uppercase, subscript-ij *e.g.* X_{ij}. The superscript $[.]^T$ denotes the transpose of a matrix or vector, so that $\mathbf{x}^T$ is a row vector. The notation $[x_1, \ldots, x_M]$ denotes a row vector with M elements, and the corresponding column vector is $\mathbf{x} = [x_1, ..., x_M]^T$.

We use typeface serif to denote deterministic variables. Also, random variables are written in *italicized* typeface sans serif. The realisations of the random variables are the same as random variables except that they are not italicized. The notations are summarized in the following table :

Definition	Notation	Example
deterministic, scalar	lowercase, serif	x
deterministic, vector	lowercase, bold, serif	$\mathbf{x}$
deterministic, matrix	uppercase, bold, serif	$\mathbf{X}$
random variable, scalar	lowercase, italic, sans serif	x
random variable, vector	lowercase, bold, italic, sans serif	$\boldsymbol{\mathsf{x}}$
random variable, matrix	uppercase, bold, italic, sans serif	$\boldsymbol{\mathsf{X}}$
random variable realization, scalar	lowercase, sans serif	x
random variable realization, vector	lowercase, sans bold, serif	$\mathbf{x}$
random variable realization, matrix	uppercase, sans bold, serif	$\mathbf{X}$

TABLE 1: Table of notations.

The expectation of a function $f(\boldsymbol{\mathsf{x}}^{(1)}, \boldsymbol{\mathsf{x}}^{(2)}, \ldots, \boldsymbol{\mathsf{x}}^{(K)})$ with respect to a random variable $\boldsymbol{\mathsf{x}}^{(i)}$ where $i \in \{1, 2, \ldots, K\}$, is denoted by $\mathbb{E}_{\boldsymbol{\mathsf{x}}^{(i)}}\left\{f(\boldsymbol{\mathsf{x}}^{(1)}, \boldsymbol{\mathsf{x}}^{(2)}, \ldots, \boldsymbol{\mathsf{x}}^{(K)})\right\}$. If the distribution of a random variable $\boldsymbol{\mathsf{X}}$ is conditioned on another variable $\boldsymbol{\mathsf{Z}}$, then, the conditional expectation of the function $f(\boldsymbol{\mathsf{X}})$ with respect to $\boldsymbol{\mathsf{X}}$ will be written as $\mathbb{E}_{\boldsymbol{\mathsf{X}}|\boldsymbol{\mathsf{Z}}}\{f(\boldsymbol{\mathsf{X}})\}$. Also, the variance of a function $f(\boldsymbol{\mathsf{x}})$ is denoted by $\mathrm{var}[f(\boldsymbol{\mathsf{x}})]$, and the covariance for two vector variables $\boldsymbol{\mathsf{x}}$ and $\boldsymbol{\mathsf{y}}$ is denoted and defined as $\mathrm{cov}[\boldsymbol{\mathsf{x}}, \boldsymbol{\mathsf{y}}] = \mathbb{E}[(\boldsymbol{\mathsf{x}} - \mathbb{E}[\boldsymbol{\mathsf{x}}])(\boldsymbol{\mathsf{y}} - \mathbb{E}[\boldsymbol{\mathsf{y}}])^T]$. Moreover, for simplicity, we denote $\mathrm{cov}[\boldsymbol{\mathsf{x}}, \boldsymbol{\mathsf{x}}]$ as $\mathrm{cov}[\boldsymbol{\mathsf{x}}]$.

Chapitre 1

Introduction

Sommaire

Optimal sensor placement plays an important role in a variety of domains, such as industry, medicine, wireless communications, aerospace, biomedical engineering, civil engineering, environmental studies, and robotics. In such applications, one is usually dealing with data acquisition for the purpose of monitoring a spatial phenomenon. Due to price, energy or ergonomic constraints (*e.g.* when the structure is the human body), the number of sensors is often limited. Therefore, an important task is to find the best spatial positions to install the sensors, such that the maximum information can be collected to provide the best estimation of the phenomenon. This problem corresponds to optimal sensor placement and it is faced in a great number of applications ranging from infrastructure monitoring [Ber+05] ; [Kra+08] ; [Dan91] ; [MZ05] and robotics [CL04] to biomedical signal processing [Her+00] ; [SJ+18] ; [VJV04], among the large variety of examples. Note that most of the time, the sensor recordings not only contain the information about the desired phenomenon, but also a mixture of several phenomena among which, only one is of interest.

This thesis focuses on the optimal sensor placement problem with the purpose of signal extraction in a noisy setting. Since we focus on the best estimation of an underlying signal, our proposed approaches are focused on the maximization of the output signal-to-noise ratio (SNR) of the estimated signal.

In this chapter, we first discuss about the concept of optimal sensors placement for source extraction. Then, the challenges that we want to address in this thesis as well as the corresponding contributions will be briefly mentioned. Finally, we will have an overview of the thesis.

1.1 A preface to optimal sensor placement for signal extraction

Many signal processing problems can be cast from a generic setting where a source signal attenuates through a given structure to the sensors. For example, a heating system (*e.g.* a

radiator) produces the source thermal signal which attenuates through a room, and a sensor receives the attenuated signal and records the temperature at its installed position. To describe a model for the measurement $y(\mathbf{x}, t)$ recorded at time t by a sensor positioned at $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^D$, where D denotes the spatial dimension, consider that the source signal $s(t)$ is attenuated through the space described by $a(\mathbf{x})$ as the *spatial gain* between the source and the sensor. Then, the noisy measurement is related to the source signal $s(t)$ as follows :

$$y(\mathbf{x}, t) = a(\mathbf{x})s(t) + n(\mathbf{x}, t), \tag{1.1}$$

where $n(\mathbf{x}, t)$ is the corresponding spatially correlated/time uncorrelated additive noise. Spatially correlated noise means that, at an instant $t = t_0$, for any pair of positions $\mathbf{x}_i$ and $\mathbf{x}_j$ the following correlation is non-zero :

$$\mathrm{Corr}\big(n(\mathbf{x}_i, t_0), n(\mathbf{x}_j, t_0)\big) = \frac{\mathbb{E}_{\mathbf{x}}\Big\{ \big[n(\mathbf{x}_i, t_0) - m^n(\mathbf{x}_i, t_0)\big] . \big[n(\mathbf{x}_j, t_0) - m^n(\mathbf{x}_j, t_0)\big] \Big\}}{\sqrt{\mathbb{E}_{\mathbf{x}}\big\{n(\mathbf{x}_i, t_0)^2\big\} - m^n(\mathbf{x}_i, t_0)^2}.\sqrt{\mathbb{E}_{\mathbf{x}}\big\{n(\mathbf{x}_j, t_0)^2\big\} - m^n(\mathbf{x}_j, t_0)^2}} \neq 0,$$

where $m^n(\mathbf{x}_i, t_0) = \mathbb{E}_{\mathbf{x}}\big\{n(\mathbf{x}_i, t_0)\big\}$ is the spatial mean of the noise at position $\mathbf{x}_i$. Also, time uncorrelated noise means that the correlation between two time instances of the noise is equal to zero :

$$\mathrm{Corr}\big(n(\mathbf{x}, t_i), n(\mathbf{x}, t_j)\big) = \frac{\mathbb{E}_{t}\Big\{ \big[n(\mathbf{x}, t_i) - m^n(\mathbf{x}, t_i)\big] . \big[n(\mathbf{x}, t_j) - m^n(\mathbf{x}, t_j)\big] \Big\}}{\sqrt{\mathbb{E}_{t}\big\{n(\mathbf{x}, t_i)^2\big\} - m^n(\mathbf{x}, t_i)^2} \ . \ \sqrt{\mathbb{E}_{t}\big\{n(\mathbf{x}, t_j)^2\big\} - m^n(\mathbf{x}, t_j)^2}} = 0,$$

where $m^n(\mathbf{x}, t_i) = \mathbb{E}_{t}\big\{n(\mathbf{x}, t_i)\big\}$ is the temporal mean of the noise at time t_i.

We note that for many electrical signals, we can assume that the propagation is very fast (with respect to the distances between sensors and sources), so the quasi-static approximation of Maxwell law applies and we can neglect the propagation times. It is the case for ECG, EEG, MEG signals, and we can use the time-invariant linear instantaneous model in (1.1). However, it is no longer true for acoustical signals where we have to take into account the propagation delay, and more generally the filtering between sensors and sources. Thus, the following linear convolutive mixture model should be used to represent the sensor measurements :

$$y(\mathbf{x}, t) = a(\mathbf{x}, t) * s(t) + n(\mathbf{x}, t). \tag{1.2}$$

The main situation we considered in this thesis is fetal ECG for which the linear instantaneous model is relevant. This would be also the case for EEG or MEG signals. If we consider PCG or audio sources, it is no longer the case, and extension to a model where $a(\mathbf{x}, t)$ is a filter becomes mandatory.

Let's consider $\mathbf{X}_{\mathcal{X}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_P]^T \in \mathbb{R}^{P \times D}$ to be candidates for sensor placement with corresponding index set $\mathcal{X} = \{1, 2, \ldots, P\} \in \mathbb{R}^P$. Now, consider the set of positions $\mathbf{X}_{\mathcal{M}} \in \mathbb{R}^{M \times D}$ as a subset of $\mathbf{X}_{\mathcal{X}}$, to be selected positions for sensor placement where $\mathcal{M} \subset \mathcal{X}$ is the set of indexes corresponding to the selected positions with the size $|\mathcal{M}| = M$. Then, the set of noisy measurements related to each sensor can be denoted by the vector

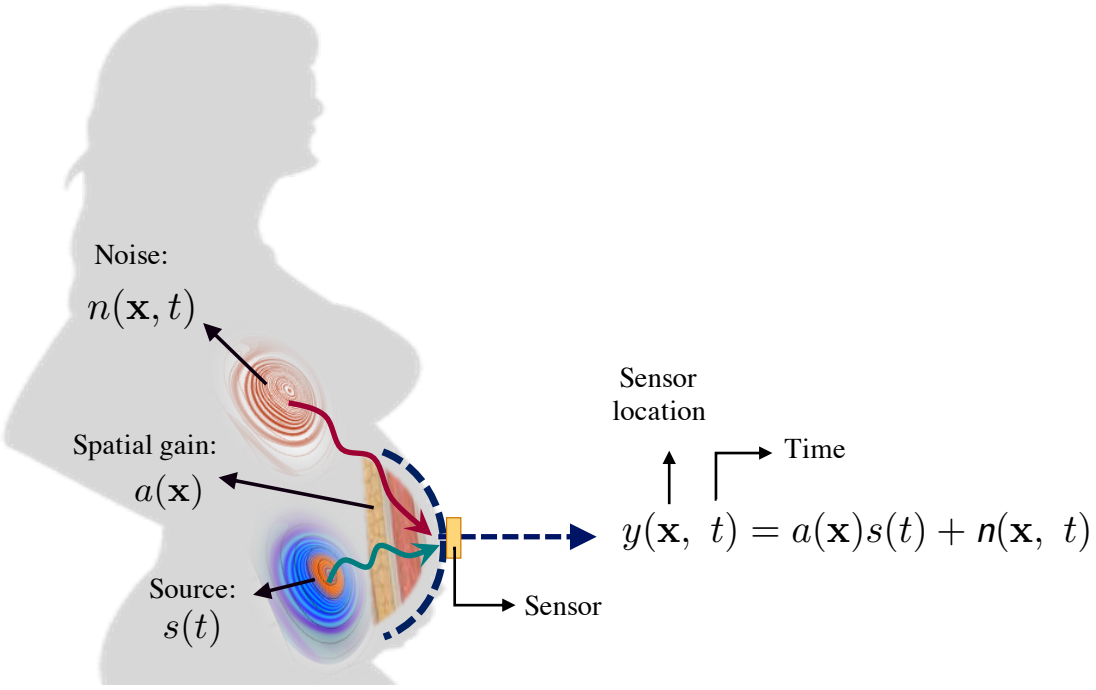


FIGURE 1.1: An example of a pregnant woman to illustrate the model of noisy measurements collected by a set of sensors. The measured signals are the attenuation of a source signal (*e.g.* fetal ECG) through a spatial gain (*e.g.* maternal abdominal tissues). Other source signals such as maternal ECG are considered as the environmental additive noise.

$\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t) = [y(\mathbf{x}_1, t), y(\mathbf{x}_2, t), \ldots, y(\mathbf{x}_M, t)]^T$. Note that in this thesis, all signals are considered to be real-valued. We also define $\mathbf{n}(\mathbf{X}_{\mathcal{M}}, t) = [n(\mathbf{x}_1, t), n(\mathbf{x}_2, t), \ldots, n(\mathbf{x}_M, t)]^T$ and $\mathbf{a}(\mathbf{X}_{\mathcal{M}}) = [a(\mathbf{x}_1), a(\mathbf{x}_2), \ldots, a(\mathbf{x}_M)]^T$ as the corresponding noise and spatial gain vectors, respectively. One can then write the measurement model (1.1) for the set of sensors $\mathbf{X}_{\mathcal{M}}$ in a general compact form as follows :

$$\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t) = \mathbf{a}(\mathbf{X}_{\mathcal{M}})s(t) + \mathbf{n}(\mathbf{X}_{\mathcal{M}}, t). \tag{1.3}$$

In Fig. 1.1 we illustrate an example of a pregnant woman in order to better understand the above model. In this example, the fetal electrocardiogram (ECG) is the source signal $s(t)$. This source signal is attenuated through the maternal abdominal tissues, and the spatial gain between the fetal heart and the sensor located in $\mathbf{x}$ is denoted by $a(\mathbf{x})$. Other signal sources, *e.g.* the maternal ECG signal, are considered as the environmental additive noise signals. Now, let's assume that the attenuated fetal ECG is expanded on the maternal skin surface according to the color-map shown in Fig. 1.1. By installing M sensors at different positions on the maternal skin surface, we have a set of noisy measurements according to (1.3).

Under the above setting, we can be interested in three different topics to be studied :

I) **Characterizing the source signal** $s(t)$ **:** Here, the main purpose is focused on extracting the source signal $s(t)$ from the set of measurements $\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t)$ collected by the

sensors. For instance, in the example of the pregnant woman, we are interested in the extraction of the fetal ECG signal. So, the best set of sensor positions are the positions that provide an estimate of $s(t)$ as accurate as possible.

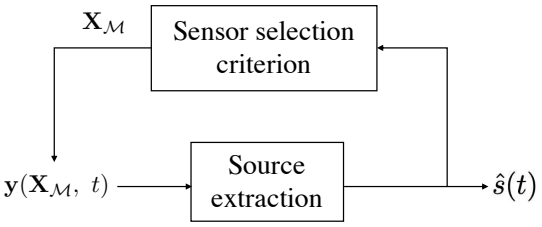
II) **Characterizing the structure $a(\mathbf{x})$:** Finding a proper model for the spatial gain $a(\mathbf{x})$ is another aspect that can be studied using the signal $\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t)$ collected by the sensors. Here, we try to answer this question : using the measurements collected by the sensors, how can we estimate the value of the spatial gain everywhere in the region of interest ? The spatial characteristics of the maternal abdominal tissues are a good example of this subject. So, in this case, the optimal sensor positions are the positions that contain the maximum information to better find a suitable model for the abdominal tissues.

III) **Characterizing the resulting field of signals in the structure $y(\mathbf{x}, t)$:** Here, we are interested in interpolating the attenuated signal through the spatial gain. In other words, we want to interpolate the values of the attenuated signal $y(\mathbf{x}, t)$ at all positions in the region of interest by having a few measurements of the attenuated signal at the sensor positions $y(\mathbf{X}_{\mathcal{M}}, t)$. Therefore, from this point of view, the best sensor positions are the positions that contain the maximum information about the attenuated source signal in the structure.

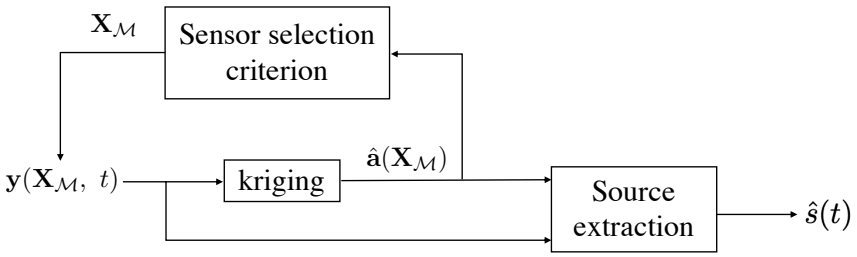
The above three problems are different, and as such, the optimal locations of sensors could be different for each problem. In all these cases, signals are recorded by multiple sensors located at different positions on the structure. The way to tackle the problem of optimal sensor placement differs from one application to another, and it mainly depends on which of three aspects mentioned above we want to focus on. It is important to note that choosing a set of positions for sensor placement may provide useful information concerning one of the above aspects, *e.g.* the source signal. However, this set of measurements may not necessarily contain suitable information about the other aspects, such as estimation of the spatial gain model. So, depending on our main goal, different criteria should be adopted to solve the problem of optimal sensor placement.

In this thesis, we want to accurately estimate the source signal, $s(t)$, from a set of M noisy measurements, $\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t)$, with a limited number M of sensors. To do so, we propose new criteria based on the output SNR on $\hat{s}(t)$, the estimation of the source $s(t)$. Here, our work differs from the classical kriging-based methods [SW87a] ; [Cre90] ; [KSG08], as they focus on reconstructing the spatial gain and not the signal itself. In the classical kriging approach, sensor locations are selected according to criteria such as entropy [SW87a] ; [Cre90] or mutual information [KSG08] ; [VJV04] ; [Sam06] on the spatial gain, while in our proposed criteria we mainly focus on the output SNR of the source signal itself.

Fig. 1.2 schematically presents the differences between the kriging approach and our method. As can be seen in Fig. 1.2a, in our proposed approach we directly focus on our goal which is the extraction of the source signal, and the sensor positions $\mathbf{X}_{\mathcal{M}}$ are selected according to this aim. Meanwhile, from Fig. 1.2b, we see that the main focus of the kriging approaches is on the best estimation of the spatial gain from noisy measurements, and it does not directly consider the best estimation of the source signal. In other words, the classical kriging approach consists of two steps : 1) the positions are selected such that we achieve the best estimation of the spatial gain, and 2) a source extraction method is used to extract the desired source signal



(a) Proposed approach.



(b) Kriging approach.

FIGURE 1.2: Comparison between the kriging and the proposed sensor selection scheme for source extraction. Top : our proposed approach selects sensor positions according to the best extraction of the source signal. Bottom : the classical kriging approaches, select sensor positions according to the best extraction of the spatial gain, and then, a source extraction method is used to recover the source signal from the measurements recorded by the sensors.

from the recordings of the already located sensors. In contrast, in our suggested method, these two steps are combined, *i.e.* the selection of the sensor locations and the extraction of the source signal are done together based on a criterion for estimating the best source signal.

It is very important to emphasize that in many works on this topic such as in [VJV04], it is assumed that we have recorded some signals using a set of sensors, and then, **some of the measurements are selected to be processed** for a certain purpose. However, in this thesis our idea is totally different. Here, from the beginning and **before recording, we want to predict where to put the sensors**. This assumes that we do not have access to the signal *before* the prediction. To better understand the difference between these two cases, we give an example. Consider the example of the pregnant woman in Fig. 1.1, where the aim is to extract the fetal ECG signal. Here, we can design the experiment for source extraction in two different ways : 1) we can use a belt filled with several sensors, and we record the signals in a full spatial grid on the maternal skin surface. Then, according to the recorded signals, we propose an algorithm *to select* the information of M sensors among the total sensors to be processed for source extraction. 2) instead of using a belt containing a lot of sensors, from the beginning, we design an algorithm to *predict* the optimal positions for M sensors, and the decision of the sensor placement is done without using a lot of observations recorded by a set of sensors. In this thesis, we focus on the second approach.

Here, we also mention that in the framework presented in this thesis, we 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 [Kri51] ; [Cre90], in the design of computer experiments [SWN13] and in global optimization [JSW98]. It has also been used for sensor placement in [KSG08], where the increment of the mutual information (MI) is proposed as a placement criterion. In [KSG08], the increment in MI is maximized using a near-optimal greedy approach. New sensors are added one by one starting from an empty set, in such a way that, at each iteration, the selected location adds maximum information.

It is also interesting to notice that our work can be seen as a problem of optimal source extraction, thus being related to the domain of source separation [CJ10b]. Source separation is a well-known problem attempting to separate different sources from a set of noisy measurements. These measurements are a mixture of the different source signals which are recorded by a set of sensors. If the number of sensors is larger than or equal to the number of sources to be retrieved, and assuming noiseless recordings, we are in an over-determined case. In this situation, there exist equivariant source separation methods [Car98]. These methods state that in a (over-)determined condition, the quality of the source retrieval is independent of the spatial gains and thus independent of the sensor positions. This property justifies why previous studies focused on providing the best algorithms to recover the sources but were not interested in choosing the best sensor locations. In the problem we study in this thesis, since the noise spatial correlation matrix is always full rank, the number of underlying sources composing the noise is always greater than the number of sensors. Therefore, such nice equivariance property is lost and the performance may be dependent on the gain $a(\mathbf{x})$ and on the noise, making the choice of sensor locations of great concern. Note that here, we are interested in extracting only a single source signal from a linear mixture with many other sources. So, in our model, we consider all the other sources as an additive noise to the desired source signal. We also stress that in our work, we do not focus on proposing a new algorithm for source extraction itself, but on choosing the sensor positions such that we obtain the best extraction performance with a given source estimator.

1.2 Challenges and contributions

In the concept of optimal sensor placement, the classical approaches focus on interpolating a field of signals, *e.g.* room temperature [KSG08]. As any other kriging-based criterion for interpolation [SW87b] ; [Sac+89], this approach could be used in our setting if our aim was to achieve the second and the third goals we mentioned in the previous section, *i.e.* characterizing the structure of the spatial gain $a(\mathbf{x})$, or characterizing the resulting field of signals in the structure $y(\mathbf{x}, t)$. But as previously stated, this is not our aim, and our goal may not be equivalent to these two goals. Therefore, in this study, we want to choose the sensor locations so that the estimation of a source signal of interest from the recordings leads to the best estimation in the signal-to-noise ratio (SNR) sense.

To achieve this goal, by considering a GP model for $a(\mathbf{x})$ and $n(\mathbf{x}, t)$, and using a linear source extraction approach, we derive the output signal-to-noise-ratio (SNR) of the estimated source signal as a function of $a(\mathbf{x})$ and of the sensor locations $\mathbf{X}_{\mathcal{M}}$. Since $a(\mathbf{x})$ is stochastic, the SNR is stochastic, too. So, based on the probability density function of the SNR, we derive new criteria that are optimized by a grid search using a greedy approach.

The last challenge that we address in this thesis is about the limitations of the greedy approach for the sensor placement problem. In the greedy approach, the sensors are added one by one from a fixed grid of candidate sensor positions. However, two limitations can be observed in this approach : first, the sensor locations are restricted to be on a predefined grid, and second, the greedy approach is sub-optimal. Consequently, to have a good performance, the grid should be fine enough, leading to a high computation cost. To overcome these issues, in this thesis we will present a gradient-based optimization approach to improve the optimization algorithm for optimal sensor placement. We propose a first order optimization-based approach that in contrast to the one-by-one strategy adopted by the greedy method on a grid, optimizes all the sensor positions at once, and does not discretize the search space. On the other hand, since placing two sensors very close to each other may not be feasible, *e.g.*, due to the physical size of the sensors, a regularization term is added to avoid choosing too close sensor positions.

1.3 Overview of the thesis

As mentioned above, the problem of optimal sensor placement for source extraction concerns several challenges which are already discussed along with our main contributions. In the following, we will have a brief overview of the next chapters.

— **Chapter 2.** State of the art

To continue this manuscript, we first begin with a description of the classical kriging-based criteria for interpolation. To do so, Chapter 2 contains a brief description of the kriging-based alternative criteria for sensor placement presented in the literature. This chapter is mostly focused on the criteria that use the concepts of mutual information (MI) and entropy. In this chapter, the greedy approach which is a common method for sensor placement will be explained, too.

— **Chapter 3.** Average signal to noise ratio as a criterion for optimal sensor placement

This chapter details our first proposed method to choose the best sensor locations for source extraction. Our first contribution is a basic method based on the average output SNR of the linearly extracted signal as a quality criterion to select sensor locations. Such a criterion includes the uncertainty on the spatial gain of the source to be extracted, providing a suitable solution for the optimal sensor placement problem. In this chapter, we will also present numerical simulations to show the superior efficiency and accuracy of the proposed method in the source extraction problem compared to the classical sensor placement criteria such as entropy and mutual information. At the end of this chapter, we will discuss about the challenges of the first proposed method.

— **Chapter 4.** Robust sensor placement for signal extraction

The fourth chapter is dedicated to our second contribution which is an improved version of the first method to provide a robust algorithm for sensor placement. Since the SNR is 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. Here, we will see that our first proposed method, which targets the average SNR, can be seen as a special case of the new robust criterion. In this chapter, to reduce the computational cost of evaluating the criterion, we will also propose a sequential approach where new sensor locations are chosen in batches. Then, in the next part, we provide several numerical results to show a consistent superiority of the proposed criterion compared with the classical kriging and the average SNR criteria in terms of the output SNR and robustness against the uncertainty on the model of spatial gain. In the end, we will mention some of the challenges regarding the proposed method.

— **Chapter 5.** Gradient-based optimization approach for optimal sensor placement

In Chapter 5, we present a new optimization approach to solve the optimal sensor placement problem. Unlike the greedy approach, our new optimization method locates all the sensors at once. Thanks to a constraint that we add to the problem, our proposed algorithm is able to control the average distances between the sensors. To solve our constrained problem, we use an alternating optimization penalty method. We apply this optimization method to our first criterion presented in Chapter 3. As this criterion is non-convex, the proposed algorithm should be carefully initialized. For this purpose, we propose to initialize it with the result of the greedy method. After presenting the proposed method, we present experimental results to show the superiority of the proposed method over the greedy approach.

— **Chapter 6.** Conclusions and perspectives

In the last chapter, we summarize the work presented in this manuscript. We also discuss about the perspectives and open problems related to this work.

Chapitre 2

State of the art

Sommaire

In this chapter, we review classical methods for optimal sensor placement. To this end, first, we will talk about Gaussian Processes (GP) as a useful tool for statistical modeling. Then, we will review the kriging approach as a classical interpolation method which is based on GP assumption on an underlying physical phenomenon. Next, optimal design will be presented as a model-based approach which is unbiased with low variance. In this part, we will discuss different criteria relying on the variance of an estimator. In this regard, the concepts of maximum likelihood and Fisher information will be explained, and then, linear estimator will be discussed. Finally, we explain the Bayesian version of the experimental design with two specific cases : entropy and mutual information (MI).

2.1 Gaussian Processes (GP) and kriging

In this section, we talk about the kriging approach which is a classical technique to interpolate the values of an underlying signal based on some observations. The kriging approach is usually based on GP and benefits from its properties obtained from the normal distribution. So, in the following, first, we will talk about GP, and then, we continue with reviewing the kriging technique.

2.1.1 Gaussian Processes

Consider a to be a one-dimensional random variable. If a follows a normal distribution, the probability density function (pdf) of this random variable is denoted as follows :

$$\mathit{a} \sim \mathcal{N}\big(m^a, \sigma_a^2\big), \tag{2.1}$$

where $m^a = \mathbb{E}\{\mathit{a}\}$ is the mean and $\sigma_a^2 = \mathbb{E}\big\{(\mathit{a} - m^a)^2\big\} > 0$ is the variance which parameterizes the distribution. The above one-dimensional normal distribution is also known as the univariate normal distribution, whose pdf is defined as follows :

$$p_{\mathit{A}}\big(a\big) = \frac{1}{\sigma_a\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{a-m^a}{\sigma_a}\right)^2}. \tag{2.2}$$

Joint Gaussian distribution, also known as multivariate normal distribution (MVN), is a general form of the univariate normal distribution defined for a random vector which consists of a set of random variables. If we consider $\boldsymbol{\mathit{a}} = [\mathit{a}_1,\ \mathit{a}_2, \ldots,\ \mathit{a}_P]^T$ to be a set of random variables, then their multivariate normal distribution is denoted as :

$$\boldsymbol{\mathit{a}} \sim \mathcal{MVN}\big(\mathbf{m^a}, \mathbf{C}^a\big), \tag{2.3}$$

where $\mathbf{m^a}$ is the mean vector as follows :

$$\mathbf{m^a} \triangleq \begin{bmatrix} \mathbb{E}\{\mathit{a}_1\} \\ \mathbb{E}\{\mathit{a}_2\} \\ \vdots \\ \mathbb{E}\{\mathit{a}_P\} \end{bmatrix} = \begin{bmatrix} m_{a_1} \\ m_{a_2} \\ \vdots \\ m_{a_P} \end{bmatrix}, \tag{2.4}$$

and $\mathbf{C}^a$ is the covariance matrix :

$$\mathbf{C}^a \triangleq \mathbb{E}\Big\{(\boldsymbol{\mathit{a}} - \mathbf{m^a})(\boldsymbol{\mathit{a}} - \mathbf{m^a})^T\Big\} = \mathbb{E}\Big\{\boldsymbol{\mathit{a}}\,\boldsymbol{\mathit{a}}^T\Big\} - \mathbf{m^a}\mathbf{m^a}^T. \tag{2.5}$$

Considering the above mean vector and covariance matrix, the pdf of the multivariate normal distribution is :

$$p_{\boldsymbol{\mathit{A}}}\Big(\mathbf{a}\Big) = \frac{1}{(2\pi)^{P/2}|\mathbf{C}^a|^{1/2}}\ e^{-\frac{1}{2}(\mathbf{a}-\mathbf{m^a})^T[\mathbf{C}^a]^{-1}(\mathbf{a}-\mathbf{m^a})}. \tag{2.6}$$

Now, assume that the random variable $\mathit{a}(\mathbf{x})$ is a function of another variable, *e.g.* space $\mathbf{x} \in \mathbf{X}_\mathcal{P}$, where $\mathbf{X}_\mathcal{P} \in \mathbb{R}^{P\times D}$ is considered to be the discrete spatial grid of size P in dimension D, which can be equal to 1, 2, or 3. If for any subset of the space $\mathbf{X}_\mathcal{M} \subset \mathbf{X}_\mathcal{P}$ with size $|\mathbf{X}_\mathcal{M}| = M$, the set of random variables $\boldsymbol{\mathit{a}}(\mathbf{X}_\mathcal{M}) = [\mathit{a}(\mathbf{x}_1),\ \mathit{a}(\mathbf{x}_2), \ldots,\ \mathit{a}(\mathbf{x}_M)]^T$ has a multivariate normal distribution, then the random variable $\mathit{a}(\mathbf{x})$ is said to be a GP [RW06] :

$$\mathit{a}(\mathbf{x}) \sim \mathcal{GP}\big(m^a(\mathbf{x}), C^a(\mathbf{x}, \mathbf{x}')\big), \tag{2.7}$$

where $m^a(\mathbf{x})$ is the mean function, and $C^a(\mathbf{x}, \mathbf{x}')$ is the covariance function, both being functions of the position $\mathbf{x}$. Now, by having the mean vector :

$$\mathbf{m^a}(\mathbf{X}_\mathcal{P}) \triangleq \begin{bmatrix} \mathbb{E}\{\mathit{a}(\mathbf{x}_1)\} \\ \mathbb{E}\{\mathit{a}(\mathbf{x}_2)\} \\ \vdots \\ \mathbb{E}\{\mathit{a}(\mathbf{x}_P)\} \end{bmatrix} = \begin{bmatrix} m^a(\mathbf{x}_1) \\ m^a(\mathbf{x}_2) \\ \vdots \\ m^a(\mathbf{x}_P) \end{bmatrix}, \tag{2.8}$$

and the covariance matrix :

$$\begin{aligned}\mathbf{C}^a(\mathbf{X}_\mathcal{P},\mathbf{X}_\mathcal{P}) &\triangleq \mathbb{E}\Big\{\big[\boldsymbol{a}(\mathbf{X}_\mathcal{P})-\mathbf{m}^\mathbf{a}(\mathbf{X}_\mathcal{P})\big]\big[\boldsymbol{a}(\mathbf{X}_\mathcal{P})-\mathbf{m}^\mathbf{a}(\mathbf{X}_\mathcal{P})\big]^T\Big\}\\ &= \mathbb{E}\Big\{\boldsymbol{a}(\mathbf{X}_\mathcal{P})\boldsymbol{a}(\mathbf{X}_\mathcal{P})^T\Big\}-\mathbf{m}^\mathbf{a}(\mathbf{X}_\mathcal{P})\mathbf{m}^\mathbf{a}(\mathbf{X}_\mathcal{P})^T,\end{aligned} \tag{2.9}$$

the pdf of the random vector $\boldsymbol{a}(\mathbf{X}_\mathcal{M})$ will be an MVN which depends on the positions $\mathbf{X}_\mathcal{M}$:

$$p_{\boldsymbol{A}}\Big(\mathbf{a}(\mathbf{X}_\mathcal{M})\Big) = \frac{1}{(2\pi)^{\frac{M}{2}}\mid \mathbf{C}^a(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})\mid^{\frac{1}{2}}}\, e^{-\frac{1}{2}\left(\mathbf{a}(\mathbf{X}_\mathcal{M})-\mathbf{m}^\mathbf{a}(\mathbf{X}_\mathcal{M})\right)^T\left(\mathbf{C}^a(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})\right)^{-1}\left(\mathbf{a}(\mathbf{X}_\mathcal{M})-\mathbf{m}^\mathbf{a}(\mathbf{X}_\mathcal{M})\right)}. \tag{2.10}$$

GP has an interesting property which is called the marginalization property. This property says that any subset of the underlying set of random variables has a normal distribution (2.3) regardless of the rest of the set. The random variables within the retained subset are called the marginal variables, and the mean and covariance of the normal distribution which specify the marginal variables will be the corresponding components of the mean vector and of the covariance matrix associated with the primary GP (2.7). The marginalization property of a GP will be explained with more details in the next part, and we will show how it can be used in the kriging approach for interpolating a signal from a few set of observations.

Finally, we note that the mean function can be a non-linear, or a linear parametrized function as follows :

$$m^a(\mathbf{x}) = \sum_{i=1}^{p}\beta_i f_i(\mathbf{x}) \tag{2.11}$$

where β_i's are parameters and $f_i(\mathbf{x})$'s are some fixed functions. Also, the function $C^a(\mathbf{x},\mathbf{x}')$ specifies the covariance matrix of the GP at any two points $\mathbf{x}$ and $\mathbf{x}'$. Different types of kernels are reported in the literature [Ste12] ; [RW06]. Depending on the behavior of the underlying gain function, we use a different kernel. For example, if we want to model a rather smooth gain function with infinitely differentiable sample-paths, we can use the following squared exponential kernel :

$$C^a(\mathbf{x},\mathbf{x}') = \sigma_a^2 \exp\big(-\frac{1}{2}r(\mathbf{x},\mathbf{x}')\big) \tag{2.12}$$

where $r(\mathbf{x},\mathbf{x}') = (\mathbf{x}-\mathbf{x}')^T\mathrm{Diag}\{\boldsymbol{\rho}_a\}^{-1}(\mathbf{x}-\mathbf{x}')$, with $\mathrm{Diag}\{\boldsymbol{\rho}_a\}$ denoting a diagonal matrix with its diagonal entries equal to the elements of $\boldsymbol{\rho}_a$, and $\boldsymbol{\rho}_a = \begin{bmatrix}\rho_a^{(1)} & \cdots & \rho_a^{(D)}\end{bmatrix}^T$ being a vector with correlation length-scales for the different spatial axes. By increasing $\rho_a^{(i)}$, the spatial gain becomes smoother, and by decreasing this parameter, the spatial gain becomes more non-smooth. Parameter σ_a is the standard deviation of the process, and thus, it controls the amplitude of its fluctuations around $m^a(\mathbf{x})$.

2.1.2 Kriging : prediction using observations

In this part, we discuss about the classical kriging approach to predict a physical phenomenon based on some observations. The kriging approach is an interpolation technique relying

on GP. Here, we will explain how the marginalization property of a GP can be useful to interpolate the underlying signal using kriging.

To this end, we give an example of the optimal sensor placement problem which this thesis is based on. From Chapter 1, we recall the source extraction problem where we have a source signal $s(t)$ attenuated by passing through a spatial field being characterized by the spatial gain $a(\mathbf{x})$. Assuming $n(\mathbf{x}, t)$ to be an additive noise signal, the measurement

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

is recorded at time t using a sensor at position $\mathbf{x}$. If we consider M sensors located at $\mathbf{X}_{\mathcal{M}} = [\mathbf{x}_1,\ \mathbf{x}_2, \ldots, \mathbf{x}_M]^T$, we obtain a set of noisy measurements

$$\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t) = \mathbf{a}(\mathbf{X}_{\mathcal{M}})s(t) + \mathbf{n}(\mathbf{X}_{\mathcal{M}}, t),$$

and the purpose is to find the best sensor positions $\mathbf{X}_{\mathcal{M}}$ to have the best extraction of the source $s(t)$. To solve this problem, the classical kriging methods first estimate the best M sensor locations such that the best estimation of the spatial gain at the rest of the space is achieved.

To apply the kriging approach to the sensor placement problem, we assume the unknown spatial gain $\mathit{a}(\mathbf{x})$, which is a function of the space $\mathbf{x}$, to be a GP (2.7) with deterministic mean vector $m^a(\mathbf{x})$ and covariance matrix $C^a(\mathbf{x}, \mathbf{x}')$. Consider $\mathbf{X}_{\mathcal{P}} = [\mathbf{x}_1,\ \mathbf{x}_2, \ldots,\ \mathbf{x}_P]^T$ to be the spatial candidates for sensor placement with index set $\mathcal{P} = \{1,\ 2, \ldots, P\}$, and a set of initial K sensors located at positions $\mathbf{X}_{\mathcal{K}}$, where $\mathcal{K} \subset \mathcal{P}$ represents the indexes of the selected positions for sensors placement. Now, we select an arbitrary subset $\mathbf{X}_{\mathcal{N}} \subset \mathbf{X}_{\mathcal{P}\backslash\mathcal{K}}$ from the unobserved positions, denoting the indexes of the unobserved positions by $\mathcal{P}\backslash\mathcal{K}$. Let's $\mathbf{X}_{\mathcal{M}}$ be the union of the initial positions $\mathbf{X}_{\mathcal{K}}$ and the new positions $\mathbf{X}_{\mathcal{N}}$, indicating the total number of positions for sensor placement. Then, the covariance matrix $\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})$ which contains the uncertainty information about the spatial gain at the observed positions and the selected unobserved positions, can be partitioned as follows :

$$\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) = \begin{bmatrix} \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) & \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \\ \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}}) & \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) \end{bmatrix}. \tag{2.13}$$

The mean vector can also be written in blocks as :

$$\mathbf{m}^{\mathbf{a}}(\mathbf{X}_{\mathcal{M}}) = \begin{bmatrix} \mathbf{m}^{\mathbf{a}}(\mathbf{X}_{\mathcal{K}}) \\ \mathbf{m}^{\mathbf{a}}(\mathbf{X}_{\mathcal{N}}) \end{bmatrix}. \tag{2.14}$$

Since $\mathit{a}(\mathbf{x})$ is a GP, any random vector consisting of the spatial gains at a subset of the space has a multivariate normal distribution. So, $\mathbf{a}(\mathbf{X}_{\mathcal{M}})$ has a multivariate normal distribution and if $\mathbf{a}(\mathbf{X}_{\mathcal{K}})$ represents the observations measured by the sensors, then, by using the marginalization property of the GP, the distribution of the spatial gain at $\mathbf{X}_{\mathcal{N}}$ conditioned on the observations $\mathbf{a}(\mathbf{X}_{\mathcal{K}})$ has a multivariate normal distribution as follows :

$$\mathit{a}(\mathbf{X}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}}) \sim \mathcal{MVN}\Big(\mathbf{m}^{\mathbf{a}}(\mathbf{X}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}}),\ \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}})\Big), \tag{2.15}$$

with the following conditional mean vector :

$$\mathbf{m^a}(\mathbf{X}_\mathcal{N} \mid \mathbf{X}_\mathcal{K}) = \mathbf{m^a}(\mathbf{X}_\mathcal{N}) + \mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{K})\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K})^{-1}\big(\mathbf{a}(\mathbf{X}_\mathcal{K}) - \mathbf{m^a}(\mathbf{X}_\mathcal{K})\big), \quad (2.16)$$

and conditional covariance matrix :

$$\mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{N} \mid \mathbf{X}_\mathcal{K}) = \mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{N}) - \mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{K})\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K})^{-1}\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{N}). \quad (2.17)$$

Note that in the case of noisy observations of the spatial gain, *i.e.* $\mathbf{z}(\mathbf{X}_\mathcal{K}) = \mathbf{a}(\mathbf{X}_\mathcal{K}) + \mathbf{b}(\mathbf{X}_\mathcal{K})$, by assuming the additive noise $\mathbf{b}(\mathbf{X}_\mathcal{K})$ to be independent of $\mathbf{a}(\mathbf{X}_\mathcal{K})$ and having a normal distribution with zero mean and covariance matrix $\mathbf{C}^b(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K})$, the above equations remain intact, except that the covariance matrix $\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K})$ will be replaced by $\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K}) + \mathbf{C}^b(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K})$. That is :

$$\begin{aligned}&\mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{N} \mid \mathbf{X}_\mathcal{K}) \\ &= \mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{N}) - \mathbf{C}^a(\mathbf{X}_\mathcal{N}, \mathbf{X}_\mathcal{K})\big[\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K}) + \mathbf{C}^b(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K})\big]^{-1}\mathbf{C}^a(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{N}). \quad (2.18)\end{aligned}$$

2.2 Optimal design

Optimal design is a set of experiments that optimally specifies a statistical model according to a set of observations *e.g.* [Puk87] ; [Atk88] ; [Atk96] ; [BV04]. To estimate the parameters of a statistical model, optimal design reduces the required experiments and optimizes the parameters such that the estimator becomes unbiased, with low variance. The more information the observations carry about the parameters, the better the estimation of the parameters. In what follows, we describe how an optimal design works, and for providing a clear understanding of its link to the optimal sensor placement, we continue with an example of the estimation of the spatial gain.

2.2.1 Maximum Likelihood (ML) estimator and Fisher information

Let's consider $h(\mathbf{x}; \Theta)$ to be a model for representing $a(\mathbf{x})$, *i.e.* the spatial gain at position $\mathbf{x}$. This model depends on a set of parameters $\Theta = [\theta_1,\ \theta_2, \ldots,\ \theta_L]^T$, and by taking a position $\mathbf{x}$ as an input, it provides us with an approximation of the true spatial gain in the output : $\hat{a}(\mathbf{x}) = h(\mathbf{x}; \Theta)$. If we denote $\varepsilon(\mathbf{x}) = a(\mathbf{x}) - \hat{a}(\mathbf{x})$ as the model error, then, we have the following :

$$a(\mathbf{x}) = \hat{a}(\mathbf{x}) + \varepsilon(\mathbf{x}) = h(\mathbf{x}; \Theta) + \varepsilon(\mathbf{x}). \quad (2.19)$$

The block-diagram to illustrate the above equation is given in Fig. 2.1.

Now, our purpose is to estimate the model parameters Θ according to a set of observations recorded by sensors. In the previous chapter, we showed that the set of noisy measurements $\boldsymbol{y}(\mathbf{X}_\mathcal{M}, t) = [y(\mathbf{x}_1, t),\ y(\mathbf{x}_2, t), \ldots, y(\mathbf{x}_M, t)]^T$ collected by M sensors located at $\mathbf{X}_\mathcal{M} = [\mathbf{x}_1,\ \mathbf{x}_2, \ldots, \mathbf{x}_M]^T$ can be described according to the following equation :

$$\boldsymbol{y}(\mathbf{X}_\mathcal{M}, t) = \boldsymbol{a}(\mathbf{X}_\mathcal{M})s(t) + \boldsymbol{n}(\mathbf{X}_\mathcal{M}, t), \quad (2.20)$$

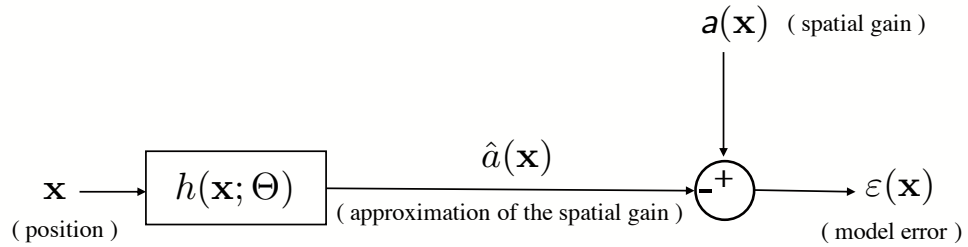


FIGURE 2.1: The block-diagram of equation (2.19) to represent the error model of the spatial gain approximation.

where $s(t)$ is the source signal, $\boldsymbol{a}(\mathbf{X}_{\mathcal{M}}) = [a(\mathbf{x}_1),\ a(\mathbf{x}_2), \ldots, a(\mathbf{x}_M)]^T$ is the spatial gain, and $\boldsymbol{n}(\mathbf{X}_{\mathcal{M}}, t) = [n(\mathbf{x}_1, t),\ n(\mathbf{x}_2, t), \ldots, n(\mathbf{x}_M, t)]^T$ is the noise corresponding to the positions $\mathbf{X}_{\mathcal{M}}$. If we fix the time t at an instant t_0 ($t = t_0$), then we have :

$$\boldsymbol{y}(\mathbf{X}_{\mathcal{M}}, t_0) = \boldsymbol{a}(\mathbf{X}_{\mathcal{M}})s(t_0) + \boldsymbol{n}(\mathbf{X}_{\mathcal{M}}, t_0), \tag{2.21}$$

which now only depends on the sensor positions $\mathbf{X}_{\mathcal{M}}$. So, without any loss of generality, if we record the data at a single instant t_0, we can consider $s(t_0) = 1$ (except for $s(t_0) = 0$). In this case, $s(t)$ is a scaling factor that can be incorporated in the spatial gain $\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})$ by adjusting the parameters accordingly. Then, the above notation can be simplified as :

$$\boldsymbol{y}(\mathbf{X}_{\mathcal{M}}) = \boldsymbol{a}(\mathbf{X}_{\mathcal{M}}) + \boldsymbol{n}(\mathbf{X}_{\mathcal{M}}). \tag{2.22}$$

Now, by replacing (2.19) in the above equation, we have :

$$\boldsymbol{y}(\mathbf{X}_{\mathcal{M}}) = \mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta) + \boldsymbol{\varepsilon}(\mathbf{X}_{\mathcal{M}}) + \boldsymbol{n}(\mathbf{X}_{\mathcal{M}}), \tag{2.23}$$

where $\mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta) = [h(\mathbf{x}_1; \Theta),\ h(\mathbf{x}_2; \Theta), \ldots,\ h(\mathbf{x}_M; \Theta)]^T$, and $\boldsymbol{\varepsilon}(\mathbf{X}_{\mathcal{M}}) = [\varepsilon(\mathbf{x}_1),\ \varepsilon(\mathbf{x}_2), \ldots,\ \varepsilon(\mathbf{x}_M)]^T$. In the above equation, two types of noises are considered : $\boldsymbol{\varepsilon}(\mathbf{X}_{\mathcal{M}})$ is the model error (2.19) and it is assumed to be deterministic, and $\boldsymbol{n}(\mathbf{X}_{\mathcal{M}})$ is the measurement noise assumed to be a random variable. This random assumption can be due to different facts. For instance, if we have an electronic noise, this noise depends on the temperature which is out of our control. Now, by combining the two noises, we end up with the following equation :

$$\boldsymbol{y}(\mathbf{X}_{\mathcal{M}}) = \mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta) + \boldsymbol{e}(\mathbf{X}_{\mathcal{M}}), \tag{2.24}$$

where $\boldsymbol{e}(\mathbf{X}_{\mathcal{M}}) = \boldsymbol{n}(\mathbf{X}_{\mathcal{M}}) + \boldsymbol{\varepsilon}(\mathbf{X}_{\mathcal{M}})$. Then, if we set $p_{\boldsymbol{E}}\Big(\mathbf{e}(\mathbf{X}_{\mathcal{M}});\ \Gamma\Big)$ to be the pdf of the random noise $\boldsymbol{e}(\mathbf{X}_{\mathcal{M}})$ with deterministic set of parameters $\Gamma = [\gamma_1,\ \gamma_2, \ldots, \gamma_l]^T$, the pdf of the measurements $\boldsymbol{y}(\mathbf{X}_{\mathcal{M}})$ will be :

$$p_{\boldsymbol{Y}}\Big(\mathbf{y}(\mathbf{X}_{\mathcal{M}}); \Theta,\ \Gamma\Big) = p_{\boldsymbol{E}}\Big(\mathbf{e}(\mathbf{X}_{\mathcal{M}});\ \Gamma\Big) \left|\frac{\partial \boldsymbol{e}(\mathbf{X}_{\mathcal{M}})}{\partial \boldsymbol{y}(\mathbf{X}_{\mathcal{M}})}\right| \Bigg|_{\mathbf{y}(\mathbf{X}_{\mathcal{M}}) = \mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta) + \mathbf{e}(\mathbf{X}_{\mathcal{M}})}. \tag{2.25}$$

Next, by considering Γ as an implicit parameter characterizing the distribution, and noting $\frac{\partial \boldsymbol{e}(\mathbf{X}_{\mathcal{M}})}{\partial \boldsymbol{y}(\mathbf{X}_{\mathcal{M}})} = 1$, the above pdf can be simplified as follows :

$$p_{\boldsymbol{Y}}\Big(\mathbf{y}(\mathbf{X}_{\mathcal{M}}); \Theta\Big) = p_{\boldsymbol{E}}\Big(\mathbf{y}(\mathbf{X}_{\mathcal{M}}) - \mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta);\ \Gamma\Big). \tag{2.26}$$

Note that if we assume that the considered model for the spatial gain is well adapted to the spatial gain, it is possible to ignore the modeling noise $\boldsymbol{\varepsilon}$, and set it equal to zero. Also, considering the noise $\boldsymbol{n}(\mathbf{X}_{\mathcal{M}})$ to have a multivariate normal distribution with a zero mean vector (*i.e.* there is no systematic error), and a deterministic covariance matrix $\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})$, then $p_{\boldsymbol{E}}\Big(\mathbf{e}(\mathbf{X}_{\mathcal{M}}); \Gamma\Big) = p_{\boldsymbol{N}}\Big(\mathbf{n}(\mathbf{X}_{\mathcal{M}})\Big)$ with :

$$p_{\boldsymbol{N}}\Big(\mathbf{n}(\mathbf{X}_{\mathcal{M}})\Big) = \frac{1}{(2\pi)^{M/2} \mid \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \mid} \; e^{-\frac{1}{2}\mathbf{n}(\mathbf{X}_{\mathcal{M}})^T \left(\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\right)^{-1} \mathbf{n}(\mathbf{X}_{\mathcal{M}})}. \tag{2.27}$$

The measurements $\boldsymbol{y}(\mathbf{X}_{\mathcal{M}})$ will then have a multivariate normal distribution with mean $\mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta)$, and the covariance matrix $\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})$:

$$\boldsymbol{y}(\mathbf{X}_{\mathcal{M}}; \Theta) \sim \mathcal{MVN}\Big(\mathbf{h}(\mathbf{X}_{\mathcal{M}}; \Theta), \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\Big). \tag{2.28}$$

In estimation theory, minimizing the variance of an estimator is equivalent to maximizing information [Kay93]. However, deriving the variance of the estimator is not always an easy task. Finding a suitable lower-bound on the variance and minimizing it can be an approximate way to solve the problem [Cra46] ; [Rao45]. A lower-bound on the variance of an unbiased estimator can be obtained by the Cramér–Rao lower bound (CRLB) which is equivalent to the inverse of the Fisher information matrix denoted by $\mathcal{I}(\Theta)$ [Cra46] ; [Rao45] ; [Kay93] ; [Rao94] ; [Dar45] ; [Gar58] ; [Mal99] :

$$\mathrm{Var}(\hat{\Theta}) \geq \mathcal{I}^{-1}(\Theta). \tag{2.29}$$

Therefore, the Fisher information matrix plays an important role in determining the amount of information within the observations. To define the Fisher information matrix, we first start with the log-likelihood function. Looking at the pdf $p_{\boldsymbol{Y}}\Big(\mathbf{y}(\mathbf{X}_{\mathcal{M}}); \Theta\Big)$ (2.26) as a function of Θ, the likelihood function is then defined as :

$$\mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big) \triangleq p_{\boldsymbol{Y}}(\boldsymbol{Y} = \mathbf{y}(\mathbf{X}_{\mathcal{M}}) \mid \Theta). \tag{2.30}$$

The gradient of the log-likelihood function with respect to Θ defines the score function :

$$\mathrm{score}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big) \triangleq \nabla_{\Theta} \log \mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big). \tag{2.31}$$

Now, the Fisher information matrix is the variance of the score function, which itself is a function of the parameter set Θ. With that in mind, the score is zero mean for unbiased estimators, and the Fisher information matrix is defined as [Duc19] :

$$\mathcal{I}(\Theta) = \mathbb{E}_{\mathbf{y}(\mathbf{X}_{\mathcal{M}})}\left[\nabla_{\Theta} \log \mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big) \nabla_{\Theta} \log \mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big)^T\right] \tag{2.32}$$

$$= \int_{-\infty}^{+\infty} \nabla_{\Theta} \log \mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big) \nabla_{\Theta} \log \mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big)^T p_Y(Y = \mathbf{y}(\mathbf{X}_{\mathcal{M}}) \mid \Theta) \; d\mathbf{y}(\mathbf{X}_{\mathcal{M}}), \tag{2.33}$$

where in the above we have a multiple integral over all the elements of the measurement vector $\mathbf{y}(\mathbf{X}_{\mathcal{M}})$. If the log-likelihood function is twice differentiable with respect to Θ, the Fisher information matrix can also be written as :

$$\mathcal{I}(\Theta) = -\mathbb{E}_{\mathbf{y}(\mathbf{X}_{\mathcal{M}})}\left[\nabla^2_{\Theta} \log \mathcal{L}\Big(\Theta \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\Big)\right]. \tag{2.34}$$

2.2.2 Alphabetical criteria

In order to simplify computational operations, different mathematical criteria, known as alphabetical criteria, have been proposed to quantify the information matrix in a scalar [Atk07]. A-optimality, D-optimality, and E-optimality are three common designs which, respectively, maximize the trace, the determinant, and the largest eigenvalue of the Fisher information matrix [CV95]. These optimality criteria are convex functions which have been well studied in [BV04]. Most of the time, choosing one of them depends on several factors and considerations, such as the difficulty of computing the criterion based on the model expression $h(\mathbf{x}, \Theta)$, the assumptions on the noise $\epsilon(\mathbf{x})$, *etc.*

There are two remarks regarding approximating the variance of the ML estimator with the inverse of the Fisher information matrix which are discussed in the following.

The first remark is that, the equality condition of the CRLB in (2.29) is satisfied if 1) a linear model with respect to the parameters Θ is used, *i.e.* $h(\mathbf{x}; \Theta) = \Theta^T \mathbf{x}$, and 2) the measurements noise is additive and normally distributed with a deterministic covariance matrix. Under these two conditions, the covariance of the estimator becomes equal to the inverse of the Fisher information matrix. Therefore, it is possible to characterize the uncertainty by the Fisher information matrix $\mathcal{I}(\Theta)$. In Section 2.2.2, we will go more into the details of the linear model, and show how the exact covariance matrix of the estimator is calculated under the linearity assumption.

The second remark is that the ML estimator is asymptotically unbiased and efficient so that the uncertainty of the parameter Θ (*i.e.* $\text{var}(\hat{\Theta})$) converges to $\mathcal{I}^{-1}(\Theta)$ when the number of measurements M increases to infinity. The consequence is that $\hat{\Theta}$ is asymptotically normally distributed $\underset{M \to +\infty}{\hat{\Theta}} \sim \mathcal{N}(\Theta, \mathcal{I}^{-1}(\Theta))$, where Θ is the true value of the parameters. Therefore, $\mathcal{I}^{-1}(\Theta)$ gives us an idea of the uncertainty of the parameters. However, in practice, the number of measurements cannot be infinite, and as such, the result is only asymptotically true. This approximation will be adequate if : 1) the number of measurements M is large, 2) the model is only slightly non-linear with respect to the parameters, and 3) the measurement errors are additive and normally distributed. Otherwise, if one of these conditions is not met, minimizing the CRLB will not necessarily ensure that the uncertainty of the estimation $\hat{\Theta}$ will be better, and the use of the matrix $\mathcal{I}^{-1}(\Theta)$ to quantify the uncertainty should be used with caution.

□ **Example1 :** linear estimator

Previously, we mentioned that if we use a linear model with respect to the parameters to be estimated, and also considering a normally distributed additive noise, the inverse of the Fisher information matrix will be equal to the covariance matrix of the estimator and the exact uncertainty of the estimation can be computed. In this part, we show how the covariance matrix can be derived.

Consider the following linear model for the spatial gain :

$$\hat{a}(\mathbf{x}) = h(\mathbf{x}; \Theta) = \Theta^T \mathbf{x}. \tag{2.35}$$

Let's assume $\mathbf{X}_{\mathcal{P}} \in \mathbb{R}^{D\times P}$ to be a matrix whose columns are the set of candidates for sensor placement, and the set of positions $\mathbf{X}_{\mathcal{M}} = \{\mathbf{x}_1,\ \mathbf{x}_2, \ldots,\ \mathbf{x}_M\} \subset \mathbf{X}_{\mathcal{P}}$ be selected for sensor placement. Then, by considering additive normally distributed noise, to obtain the maximum-likelihood estimate of Θ, we target the following optimization problem where the loss function is the Mean-Squared-Error (MSE) :

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{M} \left(y(\mathbf{x}_i) - \Theta^T \mathbf{x}_i\right)^2. \tag{2.36}$$

Solving the above problem provides the following estimator [Kay93] :

$$\hat{\Theta} = (\mathbf{X}_{\mathcal{M}}\mathbf{X}_{\mathcal{M}}^T)^{-1}\mathbf{X}_{\mathcal{M}}\boldsymbol{y}_{\mathcal{M}}, \tag{2.37}$$

where $\mathbf{X}_{\mathcal{M}} = [\mathbf{x}_1,\ \mathbf{x}_2, \ldots,\ \mathbf{x}_M] \in \mathbb{R}^{M\times D}$, and $\boldsymbol{y}_{\mathcal{M}} = [y(\mathbf{x}_1),\ y(\mathbf{x}_2), \ldots,\ y(\mathbf{x}_M)]^T \in \mathbb{R}^{M\times 1}$ are, respectively, the sensor coordinates and the corresponding observations.

Now, assume that the noise $\boldsymbol{e}(\mathbf{X}_{\mathcal{M}}) = [e(\mathbf{x}_1),\ e(\mathbf{x}_2), \ldots,\ e(\mathbf{x}_M)]^T$ is zero mean and spatially uncorrelated with variance σ_e^2, and as such, its covariance matrix is $\mathbf{C}^{\boldsymbol{e}}(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) = \mathbb{E}\{\boldsymbol{e}(\mathbf{X}_{\mathcal{M}})\boldsymbol{e}(\mathbf{X}_{\mathcal{M}})^T\} = \sigma_e^2\mathbf{I}_M$, where $\mathbf{I}_M \in \mathbb{R}^{M\times M}$ is the identity matrix. Then, it can be shown that the above estimator is unbiased, *i.e.* $\mathbb{E}\{\hat{\Theta}\} = \Theta$, and the variance of the estimation error, defined by $\Delta = \Theta - \hat{\Theta}$, will be as follows [Kay93] :

$$Cov(\hat{\Theta}) = \mathbb{E}\{(\Theta - \hat{\Theta})(\Theta - \hat{\Theta})^T\} = \sigma_e^2(\mathbf{X}_{\mathcal{M}}\mathbf{X}_{\mathcal{M}}^T)^{-1}. \tag{2.38}$$

As it can be seen, the variance of the estimator for a linear model is proportional to $(\mathbf{X}_{\mathcal{M}}\mathbf{X}_{\mathcal{M}}^T)^{-1}$. Since minimizing the above variance is proportional to maximizing the information, it is desired to collect data at positions such that the minimum variance is obtained, which means that we have a more accurate estimator. However, since minimizing a matrix does not make sense as a criterion, it is required to define a scalar measure of the matrix. As we mentioned before, there exist several scalarization criteria based on the spectrum [1] of (2.38) such as A-optimality, D-optimality, and E-optimality, which can be used here to, respectively, target the trace (the sum of the eigenvalues), the determinant (the product of the eigenvalues), and the largest eigenvalue of $(\mathbf{X}_{\mathcal{M}}\mathbf{X}_{\mathcal{M}}^T)^{-1}$. So, by minimizing any of these measures over the sensor positions $\mathbf{X}_{\mathcal{M}}$, we can obtain maximum information, and consequently, we achieve a more accurate estimator.

□ **Example2 :** best linear unbiased estimator (BLUE)

In this example, we apply best linear unbiased estimator [H71] to estimate the spatial gain at positions of interest based on some observations recorded by a set of sensors. In this regard, [KSG08] has suggested the following experiment. Imagine that we want to estimate the values of the spatial gain in N positions denoted by the matrix $\mathbf{X}_{\mathcal{N}}$ where these positions are unobserved, *i.e.* we can not put any sensors at these positions. Also, suppose that we have K sensors located at positions $\mathbf{X}_{\mathcal{K}}$, and based on the noisy measurements observed by

1. The spectrum of a matrix is the set of its eigenvalues.

these sensors, we want to predict the spatial gain at $\mathbf{X}_{\mathcal{N}}$. To this end, the BLUE suggests the following linear model :

$$\mathbf{a}(\mathbf{X}_{\mathcal{K}}) = \Upsilon \mathbf{a}(\mathbf{X}_{\mathcal{N}}) + \varpi,$$

where $\Upsilon \in \mathbb{R}^{K \times N}$ is the unknown model parameter, and ϖ is an additive model noise with variance σ^2_{ϖ}. To find the parameter Υ, we minimize the mean squared error

$$\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}}) = \underset{\Upsilon}{\operatorname{argmin}}\, \mathbb{E}\Big\{ \big[\mathbf{a}(\mathbf{X}_{\mathcal{K}}) - \Upsilon \mathbf{a}(\mathbf{X}_{\mathcal{N}})\big]^T \big[\mathbf{a}(\mathbf{X}_{\mathcal{K}}) - \Upsilon \mathbf{a}(\mathbf{X}_{\mathcal{N}})\big] \Big\}.$$

We note that here $\mathbf{X}_{\mathcal{N}}$ is known, and $\mathbf{X}_{\mathcal{K}}$ is unknown which we are looking for, and so, we look at the parameter Υ as a function of $\mathbf{X}_{\mathcal{K}}$. By considering the GP assumption on the spatial gain with the covariance matrix C^a, and solving the above problem, we obtain the following solution

$$\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}}) = \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}, \tag{2.39}$$

whose error covariance is

$$Cov\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big) = \mathbb{E}\Big\{ \big[\Upsilon - \hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big] \big[\Upsilon - \hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big]^T \Big\} = \frac{1}{\sigma^2_{\varpi}} \Big[\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}}) \hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})^T \Big]^{-1}. \tag{2.40}$$

So, according to [KSG08], the best sensor locations $\mathbf{X}_{\mathcal{K}}$ to have the most precise estimation of the spatial gain at the positions of interest $\mathbf{X}_{\mathcal{N}}$, are the ones that ***minimize*** the error covariance (2.40). However, since minimizing a matrix is not meaningful, we can use different alphabetical ctiteria such as D-optimality (the log-determinant), A-optimality (the trace), or E-optimality (the magnitude of the largest eigenvalue) to have a scalar interpretation of $Cov\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big)$. So, by applying any of these scalarization functions, the following problem is solved for sensor placement

$$\hat{\mathbf{X}}_{\mathcal{K}} = \underset{\mathbf{X}_{\mathcal{K}}}{\operatorname{argmin}}\, Cov^{(\text{opti})}\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big), \tag{2.41}$$

where $Cov^{(\text{opti})}\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big)$ represents the scalarization of $Cov\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{K}})\big)$ using any of the above-mentioned optimality functions. We again emphasize that in [KSG08], the positions of interest $\mathbf{X}_{\mathcal{N}}$ where we want to estimate the spatial gain are considered to be known and it is not possible to put any sensor at these positions of interest.

However, the problem we are addressing in this thesis is slightly different from the above problem presented in [KSG08]. In this thesis, we are not interested just in the estimation of the spatial gain at *a few* positions, but, we are interested in the estimation of the spatial gain in *the whole underlying spatial field*. Therefore, we assume that any position in the field can be a candidate for sensor placement to estimate the spatial gain at the positions which we do not put a sensor. So, we re-define the experiment as follows.

We assume that we have already put K sensors at positions $\mathbf{X}_{\mathcal{K}}$ (so, $\mathbf{X}_{\mathcal{K}}$ is known), and now we want to select the next N sensor positions $\mathbf{X}_{\mathcal{N}}$ (which are unknown) such that based on the measurements collected by $M = K + N$ sensors, we can have the best estimation of the spatial gain at the rest of the spatial field that we never put any sensor. So, logically, based on the definition of the error covariance of the BLUE, the best choice for the positions of the

next sensors are the ones that have ***larger*** error covariance $Cov\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})\big)$, which is now being considered as a function of the next unknown sensor positions :

$$Cov\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})\big) = \mathbb{E}\Big\{(\Upsilon - \hat{\Upsilon}(\mathbf{X}_{\mathcal{N}}))(\Upsilon - \hat{\Upsilon}(\mathbf{X}_{\mathcal{N}}))^T\Big\} = \frac{1}{\sigma_{\varpi}^2}\Big[\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})^T\Big]^{-1}, \tag{2.42}$$

where

$$\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}}) = \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})\mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}. \tag{2.43}$$

This means that the measurements collected by the primary K sensors do not contain enough information about the spatial gain at the positions with large error covariance, and so, it is crucial to directly collect information about the spatial gain by putting sensors at these positions. Therefore, we solve the following ***maximization*** problem to select the positions of the next N sensors :

$$\hat{\mathbf{X}}_{\mathcal{N}} = \underset{\mathbf{X}_{\mathcal{N}}}{\operatorname{argmax}}\, Cov^{\text{opti}}\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})\big). \tag{2.44}$$

To summarize, as it is seen, unlike in [KSG08], to solve the sensor placement problem in our work, we are looking for the positions with larger error covariance to put new sensors to be able to estimate the spatial gain in *the whole spatial field*.

2.3 Bayesian design for optimal sensor placement

In the previous section, we explained that the classical experimental design considers a deterministic model for the spatial gain $\hat{a}(\mathbf{x}) = h(\mathbf{x}; \Theta)$, and the purpose was the best estimation of the parameters Θ for the estimator to have a low covariance. We also explained that to have a comprehensible meaning of a low variance matrix, different alphabetical criteria can be used to scalarize the covariance matrix. If instead of a deterministic model, a prior stochastic model $\hat{a}(\mathbf{x}) \sim p_{\boldsymbol{A}}(a(\mathbf{x}); \Theta)$ is used for $\boldsymbol{a}(\mathbf{x})$ which is the spatial gain at a single sensor position $\mathbf{x}$, the problem becomes a Bayesian version of the experimental design [Lin56]. In the Bayesian design, if we consider the parameters Θ to be stochastic too, it is called a fully-Bayesian model [Bis06]. In this case, the problem is solved over the best estimation of both the spatial gain and the set of parameters. The fully-Bayesian design is out of the scope of this thesis and here, we consider that the set of parameters Θ is deterministic.

In Bayesian design, since we considered a stochastic model $\boldsymbol{a}(\mathbf{x})$ for the spatial gain, the sensor measurement in (2.24) is rewritten as follows

$$\boldsymbol{y}(\mathbf{X}_{\mathcal{M}}) = \boldsymbol{a}(\mathbf{X}_{\mathcal{M}}) + \boldsymbol{e}(\mathbf{X}_{\mathcal{M}}), \tag{2.45}$$

where $\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})$ is a random vector with distribution $p_{\boldsymbol{A}}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{M}}); \Theta\Big)$ of the M sensors which is our prior knowledge. As mentioned earlier, in this thesis, we consider the parameters Θ to be deterministic, and the purpose of Bayesian design in the optimal sensor placement problem is to find the best M positions to record the noisy measurements $\boldsymbol{y}(\mathbf{X}_{\mathcal{M}})$ such that using interpolation techniques, the best estimation of the spatial gain at the rest of the positions *i.e.* $\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}$ are obtained, where $\mathcal{P}\setminus\mathcal{M}$ represents the indexes of the unobserved positions.

To achieve this goal, by relying on some prior knowledge about the distributions of the spatial gain and the noise, and considering them to be independent, different criteria have been proposed. Among them, entropy and mutual information (MI) are two traditional criteria which have been commonly used for sensor placement. In the rest of this section, we will explain how these two criteria are used for optimal sensor placement.

2.3.1 The entropy criterion

In information theory, entropy is a well-known quantity which is used to determine the amount of uncertainty within the possible values that a random variable can take [Sha48]. In this part, we explain how entropy can be used for optimal sensor placement.

Let $\mathbf{X}_{\mathcal{P}}$ to be the set of possible positions in the space for sensor placement, and $p_{\boldsymbol{A}}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{M}}); \Theta\Big)$ to be the pdf of the spatial gain at the selected positions $\mathbf{X}_{\mathcal{M}} \subset \mathbf{X}_{\mathcal{P}}$. Then, based on the concept of entropy, the most informative subset $\mathcal{M} \subset \mathcal{P}$ with the cardinal $|\mathcal{M}| = M$, is the one that maximizes the entropy, which implies that the spatial gains at the selected positions are mostly uncertain about each other. Therefore, the following problem is

solved [Cur+91] to find the optimal M positions for sensor placement :

$$\mathcal{M} = \underset{\mathcal{M}\subset\mathcal{P};|\mathcal{M}|=M}{\operatorname{argmax}} J_H(\mathbf{X}_{\mathcal{M}}), \tag{2.46}$$

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

$$H\big(\mathbf{a}(\mathbf{X}_{\mathcal{M}})\big) = -\int_{-\infty}^{+\infty} p_{\boldsymbol{A}}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{M}});\Theta\Big)\log p_{\boldsymbol{A}}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{M}});\Theta\Big)d\mathbf{a}(\mathbf{X}_{\mathcal{M}}). \tag{2.47}$$

Using entropy for optimal sensor placement can be seen from another viewpoint. Consider $H\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}) \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\big)$ to be the entropy of the spatial gain at the unobserved positions conditioned on the measurements at the observed positions as follows :

$$\begin{aligned} &J_H\big(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}|\mathbf{X}_{\mathcal{M}}\big) \triangleq H\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}) \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\big) \\ &= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{\boldsymbol{A}}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}),\ \mathbf{a}(\mathbf{X}_{\mathcal{M}});\Theta\Big)\log p_{\boldsymbol{A}}\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}) \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\big)d\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})d\mathbf{a}(\mathbf{X}_{\mathcal{M}}). \end{aligned} \tag{2.48}$$

Then, the sensors are required to be located at positions such that $H\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}) \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\big)$ is minimized which implies that the observations at positions $\mathbf{X}_{\mathcal{M}}$ contain most of the information about the spatial gain, and the unobserved positions do not have much more information than the observed ones have. So, minimizing the conditional entropy of the unobserved positions can be a reasonable solution for sensor placement. This conditional entropy can also be written as follows :

$$H\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}) \mid \mathbf{y}(\mathbf{X}_{\mathcal{M}})\big) = H\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}})\big) - H\big(\mathbf{a}(\mathbf{X}_{\mathcal{M}})\big), \tag{2.49}$$

Minimizing this criterion means that by subtracting the information obtained by the observed positions, not much information remains in the rest of the positions. So, to achieve the best sensor positions, we end up with the following minimization problem which is exactly the same as in (2.46) :

$$\mathcal{M} = \underset{\mathcal{M}\subset\mathcal{P};|\mathcal{M}|=M}{\operatorname{argmax}} H\big(\mathbf{a}(\mathbf{X}_{\mathcal{M}})\big) = \underset{\mathcal{M}\subset\mathcal{P};|\mathcal{M}|=M}{\operatorname{argmax}} J_H(\mathbf{X}_{\mathcal{M}}). \tag{2.50}$$

Finding the optimal solution to (2.50) requires a combinatorial search which has a high computational cost, and according to [KLQ95] this problem is NP-hard. Therefore, a greedy approach can be used to find a near optimal solution to this problem [MBC79] ; [Cre91]. In the greedy approach, we assume that K sensors have already been placed at position $\mathbf{X}_{\mathcal{K}} = \{\mathbf{x}_i\}_{i\in\mathcal{K}}$ with size $|\mathcal{K}| = K \leq |\mathcal{M}| = M$, and we add the rest of the sensors one by one, *i.e.* each time a single sensor is added at position $\mathbf{x}_{\mathcal{N}}$ with the size $|\mathcal{N}| = N = 1$ leading to $K+1$ number of sensors. Taking into account the Gaussian assumption on the model of the spatial gain as in (2.7), the conditional entropy is presented as follows :

$$H\big(a(\mathbf{x}_{\mathcal{N}}) \mid \mathbf{a}(\mathbf{X}_{\mathcal{K}})\big) = \frac{1}{2}\log \mathbf{C}^a(\mathbf{x}_{\mathcal{N}},\mathbf{x}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}}) + \frac{1}{2}(\log(2\pi)+1), \tag{2.51}$$

where, $\mathbf{C}^a(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N} \mid \mathbf{X}_\mathcal{K})$ is defined in (2.17). This criterion measures the uncertainty between the spatial gain at the new position $a(\mathbf{x}_\mathcal{N})$ and the spatial gain at the previously selected positions $\mathbf{a}(\mathbf{X}_\mathcal{K})$, and it is preferred to choose the location that maximizes this quantity. Therefore, by defining

$$J_H\big(a(\mathbf{x}_\mathcal{N}) \mid \mathbf{a}(\mathbf{X}_\mathcal{K})\big) \triangleq H\big(a(\mathbf{x}_\mathcal{N}) \mid \mathbf{a}(\mathbf{X}_\mathcal{K})\big), \tag{2.52}$$

the next sensor position given by this criterion is the following :

$$\hat{\mathbf{x}}_\mathcal{N} = \underset{\mathcal{N}\in\mathcal{P}\setminus\mathcal{K}}{\operatorname{argmax}} J_H\big(a(\mathbf{x}_\mathcal{N}) \mid \mathbf{a}(\mathbf{X}_\mathcal{K})\big)). \tag{2.53}$$

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.* $\mathbf{C}^a$. Also, it is shown [Ram+05] that this criterion attempts to push the sensors far from each other and scatters them in the boundaries of the space. Since the sensors sense their surrounding, pushing them in the boundaries cause some parts of the measurements to be the signals which are not in the region of our interest, and this fact causes waste of information.

2.3.2 The criterion based on the mutual information (MI)

Another way to tackle the problem of optimal sensor placement which resolves the weak point of the entropy mentioned above is to use mutual information (MI). In [KSG08], it is suggested to maximize $\mathrm{MI}\big(\boldsymbol{a}(\mathbf{X}_\mathcal{M}), \boldsymbol{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})\big)$ which is the mutual information between the selected locations for sensor placement $\mathbf{X}_\mathcal{M}$ and the rest of the positions $\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}$:

$$\mathcal{M} = \underset{\mathcal{M}\subset\mathcal{P};|\mathcal{M}|=M}{\operatorname{argmax}} \mathrm{MI}\big(\boldsymbol{a}(\mathbf{X}_\mathcal{M}), \boldsymbol{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})\big), \tag{2.54}$$

where $\mathrm{MI}\big(\boldsymbol{a}(\mathbf{X}_\mathcal{M}), \boldsymbol{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})\big) \triangleq H\big(\mathbf{a}(\mathbf{X}_\mathcal{M})\big) - H\big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}) \mid \mathbf{a}(\mathbf{X}_\mathcal{M})\big)$.

Let's assume that $p_\mathcal{A}\Big(\mathbf{a}(\mathbf{X}_\mathcal{M}); \Theta\Big)$ and $p_\mathcal{A}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}); \Theta\Big)$ are the distributions of the spatial gain over the observed and unobserved locations, respectively. Then, the expected Kullback-Leibler divergence between the pdfs of the spatial gain over the observed and unobserved locations is the mutual information between these two sets of positions

$$\begin{aligned}\mathrm{MI}\big(\boldsymbol{a}(\mathbf{X}_\mathcal{M}), \boldsymbol{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})\big) = \int p_\mathcal{A}\Big(\mathbf{a}(\mathbf{X}_\mathcal{M}); \Theta\Big) \int p_\mathcal{A}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}); \Theta \mid \mathbf{y}(\mathbf{X}_\mathcal{M})\Big) \\ \log \frac{p_\mathcal{A}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}); \Theta \;\mid\; \mathbf{y}(\mathbf{X}_\mathcal{M})\Big)}{p_\mathcal{A}\Big(\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}); \Theta\Big)} d\mathbf{a}(\mathbf{X}_\mathcal{M}) d\mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}}).\end{aligned} \tag{2.55}$$

In [KSG08], it is shown that this criterion outperforms the entropy criterion (2.46). It is also proved that solving (2.54) 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 $\mathbf{X}_{\mathcal{K}}$, and the new single sensor is supposed to be located at $\mathbf{x}_{\mathcal{N}}$ such that the following cost function is maximized :

$$J_{MI}\big(a(\mathbf{x}_{\mathcal{N}}) \mid \mathbf{a}(\mathbf{X}_{\mathcal{K}})\big) \triangleq \mathrm{MI}\big(\mathbf{a}(\mathbf{X}_{\mathcal{M}}), \mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})\big) - \mathrm{MI}\big(\mathbf{a}(\mathbf{X}_{\mathcal{K}}), \mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{K}})\big), \tag{2.56}$$

where $\mathbf{X}_{\mathcal{M}} = [\mathbf{x}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}}]^T$. Therefore, we have the following optimization problem :

$$\hat{\mathbf{x}}_{\mathcal{N}} = \underset{\mathcal{N}\in\mathcal{P}\setminus\mathcal{K}}{\operatorname{argmax}}\, J_{MI}\big(a(\mathbf{x}_{\mathcal{N}}) \mid \mathbf{a}(\mathbf{X}_{\mathcal{K}})\big). \tag{2.57}$$

In [KSG08] it is shown that $J_{MI}\big(a(\mathbf{x}_{\mathcal{N}}) \mid \mathbf{a}(\mathbf{X}_{\mathcal{K}})\big) = H\big(a(\mathbf{x}_{\mathcal{N}}) \mid \mathbf{a}(\mathbf{x}_{\mathcal{K}})\big) - H\big(a(\mathbf{x}_{\mathcal{N}}) \mid \mathbf{a}(\mathbf{X}_{\mathcal{P}\setminus\mathcal{M}})\big)$, and under the Gaussian assumption for the spatial gain presented in (2.15), the above criterion is simplified as follows :

$$\hat{\mathbf{x}}_{\mathcal{N}} = \underset{\mathcal{N}\in\mathcal{P}\setminus\mathcal{K}}{\operatorname{argmax}}\, \frac{\mathbf{C}^a(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}) - \mathbf{C}^a(\mathbf{x}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}})(\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}))^{-1}\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{x}_{\mathcal{N}})}{\mathbf{C}^a(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}) - \mathbf{C}^a(\mathbf{x}_{\mathcal{N}}, \mathbf{X}_{\mathcal{M}})(\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}))^{-1}\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{x}_{\mathcal{N}})}. \tag{2.58}$$

2.4 Conclusion

In this chapter, we reviewed classical approaches for optimal sensor placement which focus on providing the best estimation of a spatial measured field. To this end, we first discussed Gaussian Processes as a useful stochastic tool which is commonly used in this domain. Then, we reviewed the kriging method as an interpolation technique which estimates the spatial gain based on some observations. In the kriging approach, we look for the sensor positions such that the best estimation of the spatial gain at the unobserved positions is achieved. Next, we discussed about optimal design which is a model-based approach. In this method, a model is considered for the spatial gain with a set of parameters Θ, and the purpose is to have the best estimate of the model parameters with low variance. In order to have a comprehensive meaning of a low covariance matrix and to quantify it, we reviewed the alphabetical criteria. We also mentioned that, since calculating the variance of the estimator is not an easy task, usually the inverse of the Fisher information as the lower bound of the variance is targeted. Finally, we reviewed Bayesian design for optimal sensor placement. We explained that this approach is also model-based. However, unlike the optimal design which assumes a deterministic model for the spatial gain, Bayesian design considers a stochastic model for the spatial gain. By relying on some prior knowledge about the distributions of the spatial gain and the noise, two different criteria namely the entropy, and the mutual information were discussed for the sensor placement problem.

Although the different approaches presented in this chapter have shown promising results for the reconstruction of a spatial measured field, *e.g.* the spatial gain, having a good estimation of the measurement field can not guarantee the best extraction of the desired source signal. So, since our main goal is to extract the source signal, in the rest of this thesis, we will present new approaches based on the best estimation in the signal-to-noise ratio (SNR) sense, and we will compare our proposed methods with the classical approaches discussed in this chapter.

Chapitre 3

Average signal to noise ratio as a criterion for optimal sensor placement

Sommaire

As discussed in Chapter 1, unlike the classical kriging approaches which focus on the best estimation of the spatial gain, in this thesis our goal is the best extraction of the source signal. To this aim, this study is to predict the sensor locations so that the extraction of a single source signal of interest from the recordings leads to the best estimation in the signal-to-noise ratio (SNR) sense. By using a stochastic spatial model on the gain and on the noise, and assuming that the sensor measurements follow the model presented in (3.1), we propose a sensor placement criterion based on the maximization of the average SNR. This criterion integrates not only the average gain and noise correlation, but also the uncertainty on the gain.

In the rest of this chapter, we first introduce the linear estimator for source extraction which we use throughout this thesis, and then the details of our first proposed method will be presented. In the end, we illustrate the performance of the proposed method and compare it with the classical kriging methods.

3.1 Linear estimator for source extraction

Let's assume that the observation $y(\mathbf{x}, t)$ corresponds to a single sensor located at position $\mathbf{x}$ and time t is a mixture of a source of interest $s(t)$ and a spatially correlated additive

noise $n(\mathbf{x}, t)$ as

$$y(\mathbf{x}, t) = a(\mathbf{x})s(t) + n(\mathbf{x}, t), \tag{3.1}$$

where $a(\mathbf{x})$ is the spatial gain between the signal of interest $s(t)$ and a sensor at location $\mathbf{x}$. The vector $\mathbf{x} \in \mathbb{R}^D$ represents the coordinates of the sensor in the D-dimensional space. Typically, $D \in \{1, 2, 3\}$, *i.e.* the sensor could be placed on a curve, on a surface or in 3D space, respectively.

Considering that M sensors have been placed at locations $\mathbf{X}_\mathcal{M} = [\mathbf{x}_1,\ \mathbf{x}_2, \ldots,\ \mathbf{x}_M]^T$, we have the corresponding measurement set $\mathbf{y}(\mathbf{X}_\mathcal{M}, t) = [y(\mathbf{x}_1, t),\ y(\mathbf{x}_2, t), \ldots,\ y(\mathbf{x}_M, t)]^T$ at time t being obtained as follows :

$$\mathbf{y}(\mathbf{X}_\mathcal{M}, t) = \mathbf{a}(\mathbf{X}_\mathcal{M})s(t) + \mathbf{n}(\mathbf{X}_\mathcal{M}, t), \tag{3.2}$$

where $\mathbf{a}(\mathbf{X}_\mathcal{M}) = [a(\mathbf{x}_1),\ a(\mathbf{x}_2), \ldots,\ a(\mathbf{x}_M)]^T$ is the set of spatial gains, and $\mathbf{n}(\mathbf{X}_\mathcal{M}, t) = [n(\mathbf{x}_1, t),\ n(\mathbf{x}_2, t), \ldots,\ n(\mathbf{x}_M, t)]^T$ is the noise correlated with the set of sensor positions $\mathbf{X}_\mathcal{M}$. Now, the aim of a linear source extraction is to design an extractor vector $\mathbf{f}(\mathbf{X}_\mathcal{M}) = [f_1(\mathbf{X}_\mathcal{M}),\ f_2(\mathbf{X}_\mathcal{M}), \ldots,\ f_M(\mathbf{X}_\mathcal{M})]^T \in \mathbb{R}^M$ to estimate the source $s(t)$ as follows :

$$\hat{s}(t) = \mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{y}(\mathbf{X}_\mathcal{M}, t), \tag{3.3}$$

which by replacing (3.2), we have :

$$\hat{s}(t) = \mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})s(t) + \mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{n}(\mathbf{X}_\mathcal{M}, t). \tag{3.4}$$

It is seen that the estimation of the source using linear extractor consists of two terms : the first term $\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})s(t)$ is the part related to the source signal, and the second term $\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{n}(\mathbf{X}_\mathcal{M}, t)$ is the remaining noise part. Classically, a criterion to choose the best extractor vector $\mathbf{f}(\mathbf{X}_\mathcal{M})$ is the output signal-to-noise ratio (SNR) which is obtained by dividing the two mentioned parts as follows :

$$\mathrm{SNR}(\mathbf{f}(\mathbf{X}_\mathcal{M})) = \frac{\mathbb{E}_t\Big[\big(\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})s(t)\big)^2\Big]}{\mathbb{E}_t\Big[\big(\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{n}(\mathbf{X}_\mathcal{M}, t)\big)^2\Big]}. \tag{3.5}$$

Assuming that the signal time samples $s(t)$ are temporally zero-mean, independent and identically distributed (iid) and denoting by $\sigma_s^2 = \mathbb{E}_t[s(t)^2]$ the variance of the source, and by $\mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M}) = \mathbb{E}_t[\mathbf{n}(\mathbf{X}_\mathcal{M}, t)\mathbf{n}(\mathbf{X}_\mathcal{M}, t)^T]$ the covariance matrix of the noise, then the SNR (3.5) becomes

$$\mathrm{SNR}(\mathbf{f}(\mathbf{X}_\mathcal{M})) = \sigma_s^2\, \frac{\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})\mathbf{a}(\mathbf{X}_\mathcal{M})^T \mathbf{f}(\mathbf{X}_\mathcal{M})}{\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})\mathbf{f}(\mathbf{X}_\mathcal{M})}. \tag{3.6}$$

Now, by looking at the SNR as a function of the extractor vector, we maximize it over $\mathbf{f}(\mathbf{X}_\mathcal{M})$ as follows to express the best extraction vector $\mathbf{f}^*(\mathbf{X}_\mathcal{M})$ [1] :

$$\mathbf{f}^*(\mathbf{X}_\mathcal{M}) = \underset{\mathbf{f}(\mathbf{X}_\mathcal{M})}{\operatorname{argmax}}\, \frac{\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})\mathbf{a}(\mathbf{X}_\mathcal{M})^T \mathbf{f}(\mathbf{X}_\mathcal{M})}{\mathbf{f}(\mathbf{X}_\mathcal{M})^T \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})\mathbf{f}(\mathbf{X}_\mathcal{M})}. \tag{3.7}$$

1. Note that here the scaling factor σ_s^2 is not tackled since in source extraction, the main goal is to enhance the signal of interest. Additional prior information on the signal amplitude can then be used to properly scale the extracted source.

Taking the derivative of the above cost function and setting it equal to zero leads to the following classical result [Kay93] :

$$\mathbf{f}^*(\mathbf{X}_\mathcal{M}) = \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})^{-1}\mathbf{a}(\mathbf{X}_\mathcal{M}). \tag{3.8}$$

Finally, by replacing $\mathbf{f}^*(\mathbf{X}_\mathcal{M})$ in (3.6) the achieved output SNR is given by

$$\mathrm{SNR}(\mathbf{f}^*(\mathbf{X}_\mathcal{M})) = \sigma_s^2\, \mathbf{a}(\mathbf{X}_\mathcal{M})^T \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})^{-1}\mathbf{a}(\mathbf{X}_\mathcal{M}). \tag{3.9}$$

3.2 Mean of the SNR as a criterion for optimal sensor placement

Let's assume that we want to locate M sensors to extract a source signal. The aim of optimal sensor placement is to find the best locations $\mathbf{X}^*_\mathcal{M}$ so that the output SNR (3.6) is maximized. If we consider a linear source extraction model as discussed in the previous section, according to (3.9), we can see that the output SNR is also a function of the sensor locations $\mathbf{X}_\mathcal{M}$. So, by looking at the SNR as a function of $\mathbf{X}_\mathcal{M}$, we define the following target function

$$J(\mathbf{X}_\mathcal{M}) = \mathbf{a}(\mathbf{X}_\mathcal{M})^T \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})^{-1}\mathbf{a}(\mathbf{X}_\mathcal{M}), \tag{3.10}$$

which needs to be maximized over $\mathbf{X}_\mathcal{M}$ such that :

$$\mathbf{X}^*_\mathcal{M} = \arg\max_{\mathbf{X}_\mathcal{M}} J(\mathbf{X}_\mathcal{M}). \tag{3.11}$$

A difficulty arises from this scheme : to express the criterion (3.10), the optimal extraction vector $\mathbf{f}^*(\mathbf{X}_\mathcal{M})$ (3.8) needs a perfect knowledge of the spatial gain $a(\mathbf{x})$ of the underlying source signal and of the noise's spatial covariance $\mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})$. To overcome this, we first assume that $\mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})$ can be modelled with a covariance kernel function $k^n(\mathbf{x}, \mathbf{x}')$ [RW06] that has been estimated almost perfectly (for instance from previous recordings without the source of interest). On the contrary, the spatial gain $a(\mathbf{x})$ of the source of interest is assumed to be imperfectly known and is modelled as a stochastic Gaussian process :

$$\tilde{a}(\mathbf{x}) \sim \mathcal{GP}\big(m^a(\mathbf{x}), k^a(\mathbf{x}, \mathbf{x}')\big), \tag{3.12}$$

where $m^a(\mathbf{x})$ is its mean function and $k^a(\mathbf{x}, \mathbf{x}')$ is its covariance function. From this modelling, $m^a(\cdot)$ is the main behaviour of the spatial gain and $k^a(\cdot, \cdot)$ represents the uncertainty that we have on it. In practice, these two functions can be expressed either from some prior knowledge of the recorded physical phenomenon, or from previous estimation, *e.g.* the output of independent component analysis applied on previously recorded data. Since $\tilde{a}(\mathbf{x})$ is stochastic, the SNR becomes stochastic, too. Consequently, in practice we suggest the mean output SNR as the criterion to be optimized, which is defined as follows :

$$J_E(\mathbf{X}_\mathcal{M}) = \mathbb{E}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{M})^T \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})^{-1}\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{M})\Big\}. \tag{3.13}$$

Since the expression inside the expectation is scalar, it can be replaced by its trace, and by using the cyclic property of trace, we have the following :

$$J_E(\mathbf{X}_\mathcal{M}) = \mathrm{Tr}\Big[\ \mathbb{E}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{M})\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{M})^T\Big\}\mathbf{C}^n(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})^{-1}\ \Big].$$

Now, setting $\mathbb{E}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{M})\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{M})^T\Big\} = \mathbf{C}^a(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M}) + \mathbf{m}^a(\mathbf{X}_\mathcal{M})\mathbf{m}^a(\mathbf{X}_\mathcal{M})^T$, and using the uncertainty model (3.12), $J_E(\mathbf{X}_\mathcal{M})$ becomes :

$$J_E(\mathbf{X}_\mathcal{M}) = \tag{3.14a}$$

$$\mathbf{m}^a(\mathbf{X}_\mathcal{M})^T\mathbf{C}^n(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})^{-1}\mathbf{m}^a(\mathbf{X}_\mathcal{M}) \tag{3.14b}$$

$$+\ \mathrm{Tr}\Big[\mathbf{C}^n(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})^{-1}\mathbf{C}^a(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})\Big], \tag{3.14c}$$

where $\mathbf{C}^a(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M}) \in \mathbb{R}^{M\times M}$ is the covariance matrix whose (i,j)th element is $k^a(\mathbf{x}_i,\mathbf{x}_j)$. The above equation consists of two terms. The first term (3.14b) is the SNR based on the average knowledge of the spatial gain. The additional term (3.14c) takes into account the uncertainty on the spatial gain. In practice, the optimal sensor locations are thus obtained as

$$\hat{\mathbf{X}}_\mathcal{M} = \underset{\mathbf{X}_\mathcal{M}\in\mathbb{R}^{D\times M}}{\mathrm{argmax}}\ J_E(\mathbf{X}_\mathcal{M}). \tag{3.15}$$

However, this optimization problem is difficult to solve, because the criterion $J_E(\mathbf{X}_\mathcal{M})$ 1) is non-convex, and 2) lies in a high dimensional space ($M \times D$). To overcome these difficulties, we replace (3.15) by

$$\hat{\mathbf{X}}_\mathcal{M} = \underset{\mathbf{X}_\mathcal{M}\subset\mathbf{X}_\mathcal{P}}{\mathrm{argmax}}\ J_E(\mathbf{X}_\mathcal{M}), \tag{3.16}$$

so that the search space has now a finite number of candidates.

After presenting an appropriate objective function and formulating the problem, it is required to provide an efficient way to solve the problem. Nevertheless, directly maximizing (3.16) can lead to a high computational cost because it needs to place M sensors in a D-dimensional space simultaneously, (*i.e.* it is an optimization problem of size $M \times D$). For instance, assume that $D = 3$ and the candidates for sensor placement are in a cube of size $10 \times 10 \times 10$, resulting in $P = 10^3$ total number of candidates. If we aim to use $M = 5$ sensors, to find the best sensor positions out of the candidates that maximize (3.14c), one needs to evaluate a total of $\binom{10^3}{5} = \frac{10^3!}{5!(10^3-5)!} \simeq 8.25\times 10^{12}$ cases that corresponds to a combinatorial search which has a very high computational cost. To avoid this, one can use a greedy approach that selects the M sensors by sequentially selecting $N < M$ sensors at a time. Assuming that K sensors have already been placed at $\mathbf{X}_\mathcal{K}$, and by defining $\mathbf{R}(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M}) \triangleq \mathbf{C}^n(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})^{-1}$, to choose the locations of the N following ones, the criterion (3.14) is recast as

$$J_E(\mathbf{X}_\mathcal{N}|\mathbf{X}_\mathcal{K}) = \mathbb{E}\left\{\begin{bmatrix}\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{K})^T, \tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{N})^T\end{bmatrix}\begin{bmatrix}\mathbf{R}(\mathbf{X}_\mathcal{K},\mathbf{X}_\mathcal{K}) & \mathbf{R}(\mathbf{X}_\mathcal{K},\mathbf{X}_\mathcal{N})\\ \mathbf{R}(\mathbf{X}_\mathcal{N},\mathbf{X}_\mathcal{K}) & \mathbf{R}(\mathbf{X}_\mathcal{N},\mathbf{X}_\mathcal{N})\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{K})\\ \tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{N})\end{bmatrix}\ \middle|\ \mathbf{X}_\mathcal{K}\right\} \tag{3.17}$$

$$= \mathbb{E}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{K})^T\mathbf{R}(\mathbf{X}_\mathcal{K},\mathbf{X}_\mathcal{K})\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{K}) + 2\times\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{K})^T\mathbf{R}(\mathbf{X}_\mathcal{K},\mathbf{X}_\mathcal{N})\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{N})$$

$$+\ \tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{N})^T\mathbf{R}(\mathbf{X}_\mathcal{N},\mathbf{X}_\mathcal{N})\tilde{\boldsymbol{a}}(\mathbf{X}_\mathcal{N})\ \Big|\ \mathbf{X}_\mathcal{K}\Big\}.$$

In (3.17), conditioning on $\mathbf{X}_{\mathcal{K}}$ means that $\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{K}})$ and the upper diagonal block of $\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})$ are independent of the new sensor locations $\mathbf{X}_{\mathcal{N}}$. So, $J_E(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ can be rewritten as

$$J_E(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}) = \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{K}})^T \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{K}}) + 2 \times \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{K}})^T \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \mathbb{E}\Big\{ \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{N}}) \,\Big|\, \mathbf{X}_{\mathcal{K}} \Big\} + \mathrm{Tr}\Big[\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) \, \mathbb{E}\Big\{ \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{N}}) \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{N}})^T \,\Big|\, \mathbf{X}_{\mathcal{K}} \Big\} \Big],$$

where $\mathbb{E}\big\{ \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{N}}) \,\big|\, \mathbf{X}_{\mathcal{K}} \big\} = \mathbf{m}^{\mathbf{a}}\big(\mathbf{X}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}}\big)$ is the conditional mean vector

$$\mathbf{m}^{\mathbf{a}}\big(\mathbf{X}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}}\big) = \mathbf{m}^{\mathbf{a}}(\mathbf{X}_{\mathcal{N}}) + \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}}) \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})^{-1} \big(\mathbf{a}(\mathbf{X}_{\mathcal{K}}) - \mathbf{m}^{\mathbf{a}}(\mathbf{X}_{\mathcal{K}}) \big),$$

and $\mathbb{E}\big\{ \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{N}}) \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{N}})^T \,\big|\, \mathbf{X}_{\mathcal{K}} \big\} = \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}} \big| \mathbf{X}_{\mathcal{K}}) + \mathbf{m}^{\mathbf{a}}\big(\mathbf{X}_{\mathcal{N}} \big| \mathbf{X}_{\mathcal{K}}\big) \mathbf{m}^{\mathbf{a}}\big(\mathbf{X}_{\mathcal{N}} \big| \mathbf{X}_{\mathcal{K}}\big)^T$, with the following conditional covariance matrix

$$\mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}} \big| \mathbf{X}_{\mathcal{K}}) = \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) - \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}}) \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}).$$

Therefore, to find the positions of the N new sensors, the following problem has to be solved :

$$\hat{\mathbf{X}}_{\mathcal{N}} = \underset{\mathbf{X}_{\mathcal{N}} \in \mathbb{R}^{D \times N}}{\mathrm{argmax}} \; J_E(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}). \tag{3.18}$$

Once the sensor locations $\hat{\mathbf{X}}_{\mathcal{M}}$ are obtained sequentially, one can extract the source of interest (3.3) using the estimated extractor vector

$$\hat{\mathbf{f}}(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{m}^a(\mathbf{X}_{\mathcal{M}}). \tag{3.19}$$

It is seen that the proposed approach gives a direct sensor placement criterion for source extraction and it is in contrast to standard sensor placement using kriging which consists of two steps : the standard kriging to select the sensor locations to maximize the information of the spatial gain, and then extracting the source of interest from the mixtures. Fig. 3.1 schematically presents this difference between the kriging approach and our proposed method which had also been mentioned in Chapter 1.

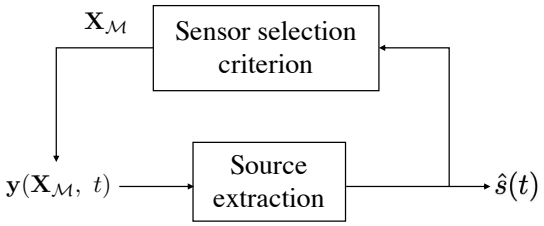
3.3 Performance analysis

Assume that the estimation of $a(\mathbf{x})$ by the mean function $m^a(\mathbf{x})$ can be written as

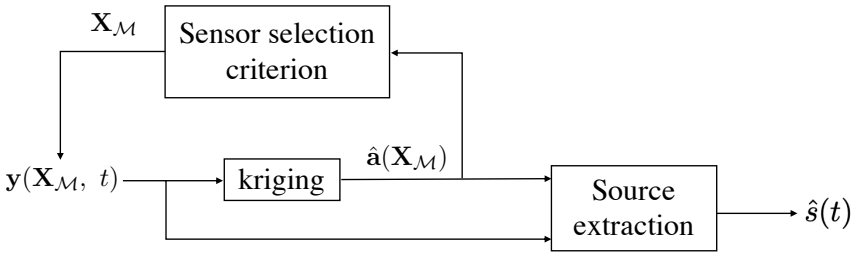
$$m^a(\mathbf{x}) = a(\mathbf{x}) + b^a(\mathbf{x}), \tag{3.20}$$

where $b^a(\mathbf{x})$ would represent the estimation bias. Then, the optimized criterion (3.14) can be expressed as

$$J_E(\mathbf{X}_{\mathcal{M}}) = \tag{3.21a}$$
$$J(\mathbf{X}_{\mathcal{M}}) \tag{3.21b}$$
$$+ 2\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{b}^a(\mathbf{X}_{\mathcal{M}}) + \mathbf{b}^a(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{b}^a(\mathbf{X}_{\mathcal{M}}) \tag{3.21c}$$
$$+ \mathrm{Tr}\Big[\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \Big], \tag{3.21d}$$



(a) Proposed approach



(b) kriging approach

FIGURE 3.1: Comparison between the kriging methods and the proposed sensor selection scheme for source extraction. Top : our proposed approach in this thesis selects sensor positions according to the best extraction of the source signal. Bottom : the classical kriging approaches select sensor positions according to the best extraction of the spatial gain, and then, a source extraction method is used to recover the source signal from the measurements recorded by the sensors.

where $J(\mathbf{X}_{\mathcal{M}})$ is the true value of the SNR with the perfect knowledge on the spatial gain which is represented in (3.10). The criterion J_E consists of three terms : the first term (3.21b) is the true criterion SNR, *i.e.* the SNR assuming a perfect knowledge on the spatial gain $a(\mathbf{x})$, the second term (3.21c) which we denote it by $J_b(\mathbf{X}_{\mathcal{M}})$ is the dependence on the estimation bias $\mathbf{b}^a(\mathbf{X}_{\mathcal{M}})$, *i.e.*, the impact of a deterministic error on the spatial gain, and the third term (3.21d) is the dependence on the uncertainty of the spatial gain $\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})$ and we denote it by $J_u(\mathbf{X}_{\mathcal{M}})$. Consequently, this shows that depending on the particular realization of error term, the optimal selected sensor locations $\hat{\mathbf{X}}_{\mathcal{M}}$ are not necessarily the true optimal sensor locations $\mathbf{X}^*_{\mathcal{M}}$. The larger the bias and uncertainty are, the more different the estimated SNR J_E from the optimal SNR J is.

Next, for the extractor vector (3.19), we reformulate the SNR. Since the spatial gain is not perfectly known, from (3.8) we conclude that the extractor vector becomes stochastic and we represent the random extractor vector as $\boldsymbol{f}(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1}\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})$. Now, to demonstrate the SNR, by recalling it from (3.6), and replacing the extractor vector with $\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1}\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})$, and taking the mean over the random extractor vector, we have the

following

$$\mathrm{SNR}\big(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})\big) = \sigma_s^2 \, \frac{\mathbb{E}_{\boldsymbol{f}}\Big\{\boldsymbol{f}(\mathbf{X}_{\mathcal{M}})^T \mathbf{a}(\mathbf{X}_{\mathcal{M}})\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T \boldsymbol{f}(\mathbf{X}_{\mathcal{M}})\Big\}}{\mathbb{E}_{\boldsymbol{f}}\Big\{\boldsymbol{f}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \boldsymbol{f}(\mathbf{X}_{\mathcal{M}})\Big\}}.$$

In the above equation, we should note that the term $\mathbf{a}(\mathbf{X}_{\mathcal{M}})\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T$ in the numerator is deterministic and obtained by the measurements, and the randomness comes from the randomness of the extractor vector $\boldsymbol{f}(\mathbf{X}_{\mathcal{M}})$. By replacing $\boldsymbol{f}(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1}\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})$, we have

$$\mathrm{SNR}\big(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})\big) = \sigma_s^2 \, \frac{\mathbb{E}_{\tilde{\boldsymbol{a}}}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{a}(\mathbf{X}_{\mathcal{M}})\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})\Big\}}{\mathbb{E}_{\tilde{\boldsymbol{a}}}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})\Big\}}.$$

Note that by $\mathrm{SNR}\big(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})\big)$ we denote the SNR which is obtained by averaging over the random extractor vector (estimated extractor vector). Now, using $\mathbf{f}^*(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1}\mathbf{a}(\mathbf{X}_{\mathcal{M}})$ (3.8), we have

$$\mathrm{SNR}\big(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})\big) = \sigma_s^2 \, \frac{\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \, \mathbb{E}_{\tilde{\boldsymbol{a}}}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})^T\Big\} \, \mathbf{f}^*(\mathbf{X}_{\mathcal{M}})}{\mathbb{E}_{\tilde{\boldsymbol{a}}}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})\Big\}}.$$

If we simplify the denominator as in (3.14), and replace $\mathbb{E}\Big\{\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})\tilde{\boldsymbol{a}}(\mathbf{X}_{\mathcal{M}})^T\Big\} = \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) + \mathbf{m}^a(\mathbf{X}_{\mathcal{M}})\mathbf{m}^a(\mathbf{X}_{\mathcal{M}})^T$ in the numerator, we end up with

$$\mathrm{SNR}\big(\hat{\mathbf{f}}(\mathbf{X}_{\mathcal{M}})\big) = \\ \sigma_S^2 \times \frac{\big[\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{m}^a(\mathbf{X}_{\mathcal{M}})\big]^2 + \mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \mathbf{f}^*(\mathbf{X}_{\mathcal{M}})}{\mathbf{m}^a(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) + \mathrm{Tr}\big[\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\big]}. \quad (3.22)$$

By replacing $\mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) = \mathbf{a}(\mathbf{X}_{\mathcal{M}}) + \mathbf{b}^a(\mathbf{X}_{\mathcal{M}})$, and defining $c^{tr}(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \triangleq \mathrm{Tr}\big[\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\big]$, with a straightforward simplification we obtain

$$\mathrm{SNR}\big(\hat{\mathbf{f}}(\mathbf{X}_{\mathcal{M}})\big) = \sigma_S^2 \times \\ \frac{\big[\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{a}(\mathbf{X}_{\mathcal{M}}) + \mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{b}^a(\mathbf{X}_{\mathcal{M}})\big]^2 + \mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \mathbf{f}^*(\mathbf{X}_{\mathcal{M}})}{\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T \mathbf{f}^*(\mathbf{X}_{\mathcal{M}}) + 2\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{b}^a(\mathbf{X}_{\mathcal{M}}) + \mathbf{b}^a(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{b}^a(\mathbf{X}_{\mathcal{M}}) + c^{tr}(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})}.$$

Finally, by recalling the true SNR with the perfect knowledge on the extractor vector (3.9), *i.e.* $\mathrm{SNR}\big(\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})\big) = \sigma_s^2 \, \mathbf{a}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{a}(\mathbf{X}_{\mathcal{M}})$, the above can be rewritten as

$$\mathrm{SNR}\big(\hat{\mathbf{f}}(\mathbf{X}_{\mathcal{M}})\big) = \mathrm{SNR}\big(\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})\big) \, \frac{\alpha}{\beta}, \quad (3.23)$$

where

$$\beta = 1 + \frac{1}{\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{a}(\mathbf{X}_{\mathcal{M}})} \times \Bigg(2\,\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})^T \mathbf{b}^a(\mathbf{X}_{\mathcal{M}}) + \mathbf{b}^a(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{b}^a(\mathbf{X}_{\mathcal{M}}) \\ + \mathrm{Tr}\big[\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\big] \Bigg),$$

and

$$\alpha = \left[1 + \frac{\mathbf{f}^*(\mathbf{X}_\mathcal{M})^T \mathbf{b}^a(\mathbf{X}_\mathcal{M})}{\mathbf{f}^*(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})}\right]^2 + \frac{\mathbf{f}^*(\mathbf{X}_\mathcal{M})^T \mathbf{C}^a(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M}) \mathbf{f}^*(\mathbf{X}_\mathcal{M})}{\left[\mathbf{f}^*(\mathbf{X}_\mathcal{M})^T \mathbf{a}(\mathbf{X}_\mathcal{M})\right]^2}.$$

This means that the achieved SNR, *i.e.* $\mathrm{SNR}\big(\hat{\mathbf{f}}(\mathbf{X}_\mathcal{M})\big)$, is equal to the oracle one, *i.e.* $\mathrm{SNR}\big(\mathbf{f}^*(\mathbf{X}_\mathcal{M})\big)$, up to a multiplicative term that depends both on the bias $\mathbf{b}^a(\mathbf{X}_\mathcal{M})$ of the estimation of $\mathbf{a}(\mathbf{X}_\mathcal{M})$ and on the uncertainty $\mathbf{C}^a(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})$ of $a(\mathbf{x})$. This multiplicative term is lower than one (by the definition of $\mathbf{f}^*(\mathbf{X}_\mathcal{M})$) indicating a loss of performance due to both the bias of estimation and uncertainty compared to the oracle SNR, *i.e.* $\mathrm{SNR}\big(\mathbf{f}^*(\mathbf{X}_\mathcal{M})\big)$.

3.4 Numerical experiments

In this section, to choose the location of the sensors, the proposed method (3.18) which is based on maximizing the output SNR will be compared to the kriging methods such as entropy (2.53) [SW87a] ; [Cre90], or mutual information (2.57) [KSG08], which are two steps methods : 1) estimating the spatial gain, 2) extraction of the source signal. The methods will be compared according to the achieved output SNR (3.23) relying on the extraction vector $\hat{\mathbf{f}}(\mathbf{X}_\mathcal{M})$ (3.19) that depends on the chosen sensor locations. The optimal SNR (3.9) will be denoted oracle SNR since it assumes a perfect knowledge of the gain $a(\mathbf{x})$.

3.4.1 Numerical setup

For the numerical experiments, the data are generated synthetically, and we consider D, the dimension of the space, to be equal to 1. The range of the sensor position x is normalized, leading to $x \in [0, 1]$ in a regular grid of size P with the index set $\mathcal{P} = \{1,\ 2, \ldots,\ P\}$. So, the underlying grid for sensor locations is denoted by the vector $\mathbf{x}_\mathcal{P} = [x_1,\ x_2, \ldots,\ x_P]^T$. We note that depending on the smoothness of the signals, the size of the grid may change in different experiments which will be specified in each experiment. Generally, the more non-smooth the signal is, the tighter the grid is, which means that we have to increase the size of the grid to be able to detect the important information according to the changes of the signal. In our experiments, we consider three initial sensors ($K = 3$) $\mathbf{x}_\mathcal{K} \subset \mathbf{x}_\mathcal{P}$, one in the middle, and the other two near the right side and the left side of the grid. We assume that the standard deviation of the source signal σ_s is equal to 1. The spatial gain $\tilde{\mathsf{a}}(x)$, and the additive noise $\mathsf{n}(x, t)$ are produced from Gaussian processes $\mathcal{GP}\big(m(x), C(x, x')\big)$. The covariance matrix is generated from a square exponential function $C(x, x') = \sigma^2 \exp(-(x - x')^2/(2\rho^2))$. The smoothness degrees for the spatial gain ρ_a, and for the noise ρ_n, as well as the standard deviations σ_a and σ_n will take different values, which will be mentioned in each experimental part. The standard deviation σ_a (resp. σ_n) is related to $\tilde{\mathsf{a}}(x)$ (resp. to $\mathsf{n}(x, t)$). Similarly, the length scale ρ_a (resp. ρ_n) is related to $\tilde{\mathsf{a}}(x)$ (resp. to $\mathsf{n}(x, t)$). The mean function $m^n(x)$ for the noise is considered to be 0, and $m^a(x)$ for the spatial gain $\tilde{\mathsf{a}}(x)$ is randomly generated by a GP with a zero-mean function, the standard deviation $\sigma_{m_a} = 0.5$, and $\rho_{m_a} = \rho_a$. Also, the mean

estimation bias $b^a(\mathbf{x})$ is produced by a GP with the same smoothness as in the spatial gain, *i.e.* we set $\rho_b = \rho_a$, and also we choose σ_b (the power parameter of the estimation error in (3.20)) such that the SNR between the mean signal $m^a(\mathbf{x})$ and the bias signal $b^a(\mathbf{x})$ becomes 5 dB.

3.4.2 Accuracy of the proposed criterion

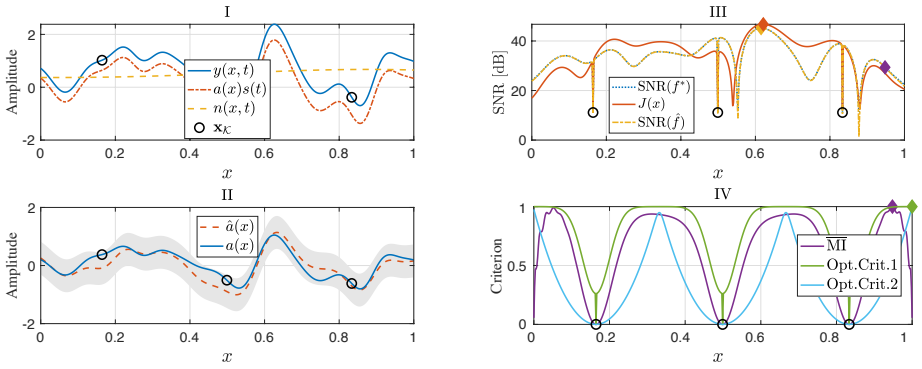
In this experiment, we study the behaviour of the proposed method and evaluate its ability to detect the best positions for sensor placement in terms of the best output SNR. Fig. 3.2 and Fig. 3.3 illustrate the accuracy of the proposed method and compare its performance with that of MI as a classical kriging method for sensor placement.

To optimize (3.14), a grid search method is used : $P = 500$ linearly spaced possible sensor locations are tested between 0 and 1, with three initial sensors arbitrary located at $\mathbf{x}_{\mathcal{K}} = [x_{83},\ x_{250},\ x_{417}]^T = [0.1643,\ 0.4990,\ 0.8337]^T$. To generate $\tilde{\mathsf{a}}(x)$ and its mean $m_a(x)$, the smoothness and power parameters defining the covariance function are $\rho_a = \rho_{m_a} = 0.05$ and $\sigma_a = \sigma_{m_a} = 0.5$, respectively. We note that a small ρ_a corresponds to a non-smooth situation, *e.g.* the spatial field in the ocean, and a large ρ_a corresponds to a smooth spatial gain *e.g.* the maternal abdominal tissues. Also, 500 time samples of $s(t)$ are randomly and independently drawn from a zero mean normal distribution with $\sigma_s = 1$. To set up the parameters of the noise $n(x,t)$, we consider different situations. First, we choose the power parameter σ_n, based on the SNR representing the ratio between the power of the signal and the power the noise according to the following :

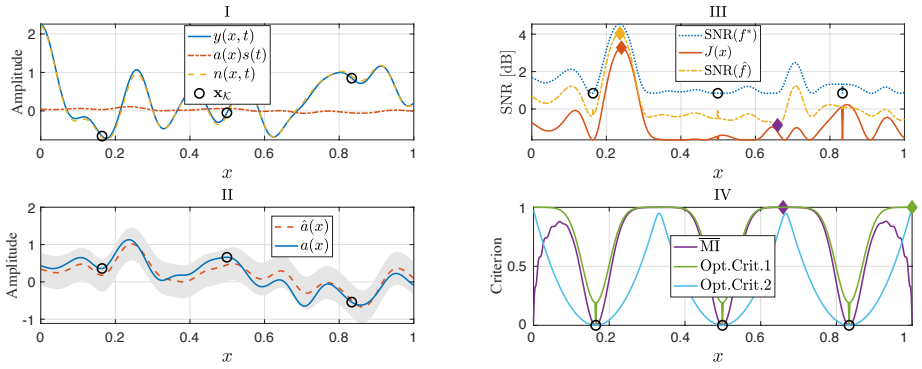
$$\mathrm{SNR}_y[\mathrm{dB}] = 10\ \log\ \left(\frac{\sigma_s^2}{\sigma_n^2}\ \frac{1}{P} \sum_{i=1}^{P} a(x_i)^2 \right). \tag{3.24}$$

We study two cases : in the first case presented in Fig. 3.2, we assume a difficult situation by setting the power parameter of the noise σ_n such that the SNR is -5 dB. In the second case shown in Fig. 3.3, we assume an easier condition and set σ_n to achieve SNR = 0 dB. For each of these cases, we take into account three smoothness degrees for the noise compared to the smoothness of the spatial gain : a) a very smooth noise by setting $\rho_n/\rho_a = 15$, b) a moderate smoothness condition $\rho_n/\rho_a = 1$, and c) almost white noise $\rho_n/\rho_a = 0.01$. For each experiment, we generate 500 time samples of the noise $\mathsf{n}(x,t)$. Each of these three configurations consist of four parts : In part I, a sample signal of the noise (yellow curve), the spatial gain (red curve), and the measurements (blue curve), as well as the positions of three initial sensors (black circles) are depicted. Part-II represents an example of the spatial gain (blue curve), its mean (the red curve), and its uncertainty (the gray shadow). In Part-III we have reported the oracle SNR (blue curve) which is the true SNR with the true extractor vector, and the true SNR with the estimated extractor vector (yellow curve), and the proposed criterion which is the average SNR (red curve). Part-IV depicts the results obtained by the MI criterion (purple curve) and the D-optimality criterion denoted by Opt.Crit.1 (green curve) and Opt.Crit.2 (blue curve) which, respectively, target the posterior covariance (2.18) [2] , and the error covariance obtained from the BLUE (2.42) [3] . In Part-IV, the criteria are normalized between zero and one. Note that in Parts III and IV, the diamonds show the maxima of the corresponding criteria, *e.g.* the red diamond in Part-III is the maxima of the proposed criterion J_E. Also, to compare the results, we have denoted the maximum of J_{MI} with purple color in Part III.

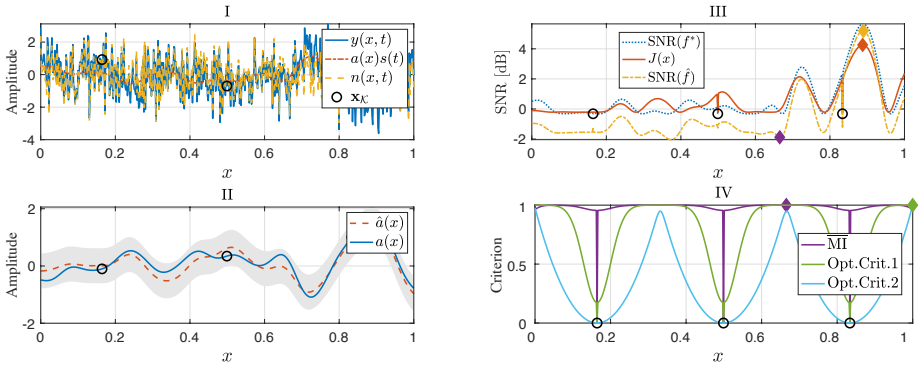
2. $\mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}} \mid \mathbf{X}_{\mathcal{K}}) = \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) - \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}}) \big[\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) + \mathbf{C}^b(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})\big]^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})$.
3. $Cov\big(\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})\big) = \frac{1}{\sigma_{\varpi}^2} \Big[\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}}) \hat{\Upsilon}(\mathbf{X}_{\mathcal{N}})^T\Big]^{-1}$, where $\hat{\Upsilon}(\mathbf{X}_{\mathcal{N}}) = \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}$.



(a) Very smooth noise : $\rho_n/\rho_a = 15$.



(b) Medium noise smoothness : $\rho_n/\rho_a = 1$



(c) White noise : $\rho_n/\rho_a = 0.01$

FIGURE 3.2: Illustration of the proposed method for SNR $= -5$ dB, $\rho_a = 0.05$ and different noise smoothness degrees, ρ_n/ρ_a, equal to a) 15, b) 1, and c) 0.01. For each noise smoothness degree, we have presented : I) an example of the recordings $y(x,t)$. II) Actual gain $\tilde{\mathbf{a}}(x)$ and its estimation $\hat{a}(x)$. Grey shadow is the uncertainty. III) Oracle SNR(f^*) (3.9), estimated SNR($\hat{f}$) (3.14c) and achieved SNR (3.23). IV) MI criterion [KSG08]. In all the plots, the circles denote the three pre-selected sensors. The diamonds indicate the maximum of each function.

Now, we discuss the results obtained by this part of the experiments. Fig. 3.2-(a) corresponds to a difficult situation with SNR $= -5$ dB. Here, we consider a very smooth noise setting $\rho_n/\rho_a = 15$, which can be seen in sub-figure I. In the third sub-figure III, it is seen that the proposed criterion $J_E(\mathbf{x})$ (*i.e.* the estimated SNR from imperfect knowledge of $\tilde{\mathbf{a}}(x)$) has roughly the same behaviour than the oracle SNR (3.9), which requires perfect knowledge of $\tilde{\mathbf{a}}(x)$. The differences are explained by (3.21) and come from the estimation bias $b(x)$ and the uncertainty of the spatial gain. The effect of the bias and the uncertainty on $J_E(\mathbf{x})$ will be more studied in subsection 3.4.3 (next subsection). Interestingly, the maximum value of J_E (3.14) is located in a range of high values for the actual SNR (3.23), thus leading to an output SNR of about 45dB with 4 sensors while the output SNR with only the 3 pre-located sensors is equal to 10dB. Moreover, it is worth noting that the achieved SNR with 4 sensors, $SNR(\hat{\mathbf{f}})$, in orange, can be lower than the output SNR with only the 3 pre-located sensors (*e.g.*, for $x = 0.88$). This means that adding a sensor will not necessarily leads to a better SNR but can deteriorate the quality of the extraction. This counter intuitive result can be explained from the imperfect estimation $\hat{a}(x)$. In this case, based on an inappropriate model, the extraction vector $\hat{\mathbf{f}}(\mathbf{x}_{\mathcal{K}+1})$ which is based on $K+1$ sensors leads to a worse estimation $\hat{s}(t)$ than using only K sensors using $\hat{\mathbf{f}}(\mathbf{x}_{\mathcal{K}})$. Finally, in the 4^{th} sub-figure IV, we can clearly see that the solution provided by the MI criterion is far from the optimal choice of sensor location with respect to the output SNR. Also, as studied in [KSG08], in sub-figure IV, we can see that regardless of the sensor recordings, the optimality criterion pushes the sensor location close to the borders of the space at $\mathbf{x} = 1$, and tries to keep the sensors as far as possible from each other, which obviously does not necessarily provide the best source extraction. This result remains intact for any other settings of the SNR and smoothness ratio between the noise and the spatial gain ($\beta = \dfrac{\rho_n}{\rho_a}$) which will be studied in the following experiments. We note that since in this experiment the new sensors are added one by one, the error covariance becomes a scalar, and as such, the different optimality criteria are all the same, and being equal to the error covariance.

Next, we switch to Fig. 3.2-(b) to examine the performance of the proposed criterion in a more difficult situation where the smoothness of the noise signal is similar to the smoothness of the spatial gain. So, in this figure, we have set $\rho_n/\rho_a = 1$, and as it is seen in sub-figure I, the behaviour of the noise signal $n(x,t)$ is quite similar to the behaviour of the signal $a(x)s(t)$ with regards to the smoothness. Therefore, because the SNR is also quite low (SNR=-5dB), extracting the information of the source signal from the set of measurements becomes more challenging. By looking at the results in the third sub-figure III, and comparing the behaviour of the proposed criterion $J(\mathbf{x})$ with the oracle SNR(f^*), we can see that the proposed criterion interestingly follows the shape of the oracle and wherever SNR(f^*) has a peak, $J(\mathbf{x})$ also contains a peak close to it. Also, comparing $J(\mathbf{x})$ with the true SNR based on estimated extractor vector, *i.e.* SNR($\hat{f}$), we see that the maximum of the proposed criterion is quite close to the maximum of the true SNR based on the estimated extractor vector, as well as the oracle. This shows that even in a more difficult situation of the noise smoothness, the proposed method is still powerful to select a position close to the optimal point. This is unlike the solution obtained by the MI criterion which is far from the optimal answer. Comparing the results of Fig. 3.2-(b)-III with the results of the smooth noise signal in Fig. 3.2-(a)-III,

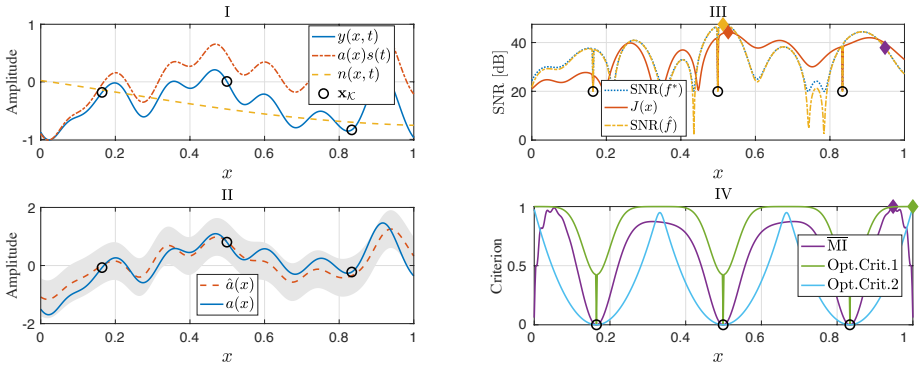
we see that the output SNR is significantly decreased due to the difficulty of the situation concerning the smoothness of the noise. Also, again here we can see that by adding a sensor in a noticeable portion of the spatial field, (*e.g.* $0.3 \leq \mathbf{x} \leq 0.7$) the SNR will be decreased due to the reasons we mentioned before.

Finally, we repeat the experiments for an almost spatially white noise situation in Fig. 3.2-(c). Interestingly, it can be seen that we have the same results as in the previous parts. Moreover, in spite of having a very low SNR equal to -5dB, and a white noise, the proposed method can still find the optimal point for the next sensor position.

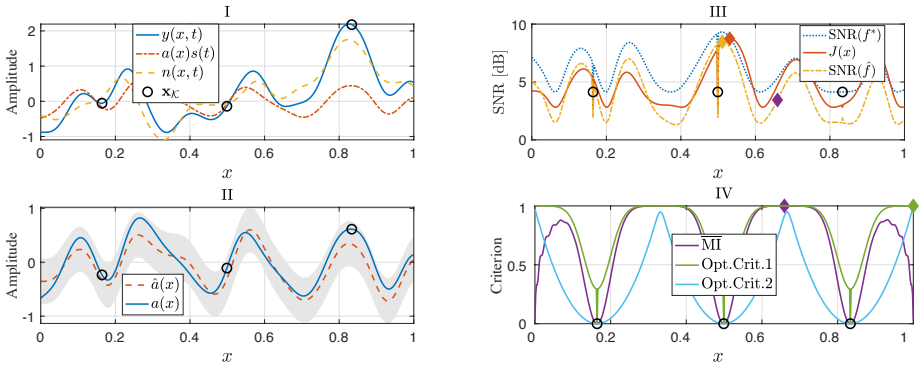
To summarize, from Fig. 3.2, by comparing the results from top to bottom, we can conclude that whenever the noise is significantly smoother than the spatial gain, *i.e.* $\rho_n >> \rho_a$, it is possible to have a large amount of increase in the SNR. In contrast, whenever $\rho_n \approx \rho_a$, *i.e.* when the smoothness of the spatial gain is similar to that of the noise, adding sensor does not provide a lot of increase in the SNR, and the next sensor will be located around $\text{argmax}_x\, a(x)$. In addition, we can see that by considering a white noise, *i.e.* $\rho_n << \rho_a$, the output SNR obtained by the proposed criterion becomes better than the situation where $\rho_n \approx \rho_a$. This means that if the spatial gain and the noise are more different from each other in terms of the smoothness, we will have a better result for source extraction. Finally, it is interesting to pay attention to the behaviour of the J_{MI} when the noise becomes very non-smooth. We can see that in such a situation, J_{MI} has almost a flat shape meaning that it is not useful to find the best sensor location in the sense of increasing the output SNR.

Finally, in Fig. 3.3, we repeat the experiment for a better SNR condition by setting the SNR equal to 0 dB. This figure also studies different smoothness degrees for the noise signal setting ρ_n/ρ_a equal to 15, 1, and 0.01, and the results are reported in parts (a), (b), and (c), respectively. From this figure, we can conclude the same results as in the Fig. 3.2, except that we obtain a better output SNR by adding the new sensor. Specifically, it is interesting to mention the spatially white noise situation in Fig. 3.3-(c), where the effect of the noise in the measurements is decreased due to decreasing the power of noise signal, and we can see in sub-figure I that the measurements better follow the shape of the source signal, *i.e.* $a(x)\, s(t)$. Therefore, the good result is that the output SNR is increased which can be seen in sub-figure III. Furthermore, we note that in sub-figure III the new sensor location is very close to the previous sensor, which means that somewhere close to the previous sensor we still have important information which should be collected by a new sensor.

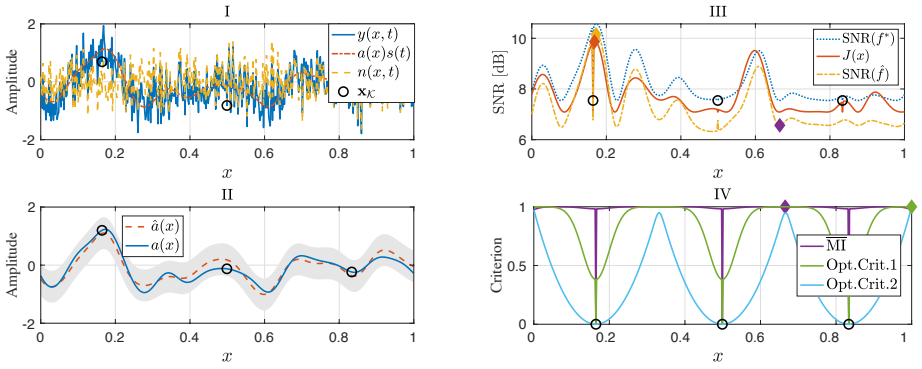
From Fig. 3.3, we can see that the weak point of the proposed method is that the suggested criterion is not powerful to find the positions where placing new sensors may deteriorate the extraction of the source and reduces the output SNR. To give an example, if we look at Fig. 3.3-III, we can see that there are some regions where by adding a new sensor the SNR decreases compared to the initial situation before adding the new sensors (the SNR at the points shown by black circles). This is happening due to the highly uncertain knowledge as well as the bias $b(x)$ we have about the spatial gain at these positions. In the next chapter, we will discuss more about this fact, and present a new method to deal with this issue.



(a) Very smooth noise : $\rho_n/\rho_a = 15$.



(b) Medium noise smoothness : $\rho_n/\rho_a = 1$



(c) White noise : $\rho_n/\rho_a = 0.01$

FIGURE 3.3: The same illustration of the proposed method as in Fig. 3.2 except that the SNR is equal to 0dB.

3.4.3 The effect of bias and uncertainty on the proposed criterion

In Section 3.3, we explained that the uncertainty and the bias of the spatial gain can cause the selected sensor location suggested by the proposed criterion $J_E(\mathbf{X}_\mathcal{M})$ (3.21a) does not necessarily match the true optimal sensor location. In this section, we want to demonstrate the effect of the bias $J_b(\mathbf{X}_\mathcal{M})$ (3.21c) and the uncertainty $J_u(\mathbf{X}_\mathcal{M})$ (3.21d) on the optimal sensor location selected by our proposed criterion $J_E(\mathbf{X}_\mathcal{M})$. To this end, we have prepared Fig.3.4. In this figure, the true value of the SNR with the perfect knowledge of the spatial gain, *i.e.* $J(\mathbf{X}_\mathcal{M})$ (3.21b), the bias $J_b(\mathbf{X}_\mathcal{M})$, and the uncertainty $J_u(\mathbf{X}_\mathcal{M})$ are, respectively, shown in sub-figures (a), (b), and (c). Then, in sub-figure (d) we have demonstrated the proposed criterion which includes the effect of the bias and the uncertainty, *i.e.* $J_E(\mathbf{X}_\mathcal{M}) = J(\mathbf{X}_\mathcal{M})+J_b(\mathbf{X}_\mathcal{M})+J_u(\mathbf{X}_\mathcal{M})$. Moreover, by ignoring the uncertainty of the spatial gain (setting $J_u(\mathbf{X}_\mathcal{M}) = 0$), the effect of the bias is presented in sub-figure (e), where we have plotted $J(\mathbf{X}_\mathcal{M})$ being added with $J_b(\mathbf{X}_\mathcal{M})$ which is the influence of the bias. Finally, sub-figure (f) presents the effect of uncertainty on the proposed criterion where we have plotted $J(\mathbf{X}_\mathcal{M})$ being added with $J_u(\mathbf{X}_\mathcal{M})$, considering $J_b(\mathbf{X}_\mathcal{M}) = 0$. In each sub-figure, the maximum of each function is marked by a diamond. Here, we assume that the SNR of the observations is 0dB, and $\rho_n/\rho_a = 2$ with $\rho_a = \rho_{m^a} = 0.05$. The variances of the source, the mean of the spatial gain, the uncertainty and the bias are, respectively, set as $\sigma_{m^a} = 0.15$, $\sigma_s = 1$, $\sigma_b = 0.2$, $\sigma_u = 0.2$. The size of the spatial grid is 300, with three initial sensors at indexes $\mathcal{K} = [50, 150, 250]^T$.

As it is seen in sub-figure (a), the maximum point suggested by the true value of the SNR is at $\mathbf{x} = 0.42$. Meanwhile, by applying the effect of the bias and the uncertainty in sub-figure (d), this result is at $\mathbf{x} = 0.83$ which is far from the optimal point. Also, by just considering the influence of the bias in sub-figure (e), the suggested criterion is about $\mathbf{x} = 0.83$, and we obtain the maximum at $\mathbf{x} = 0.1$ by looking at sub-figure (f) as the effect of the uncertainty.

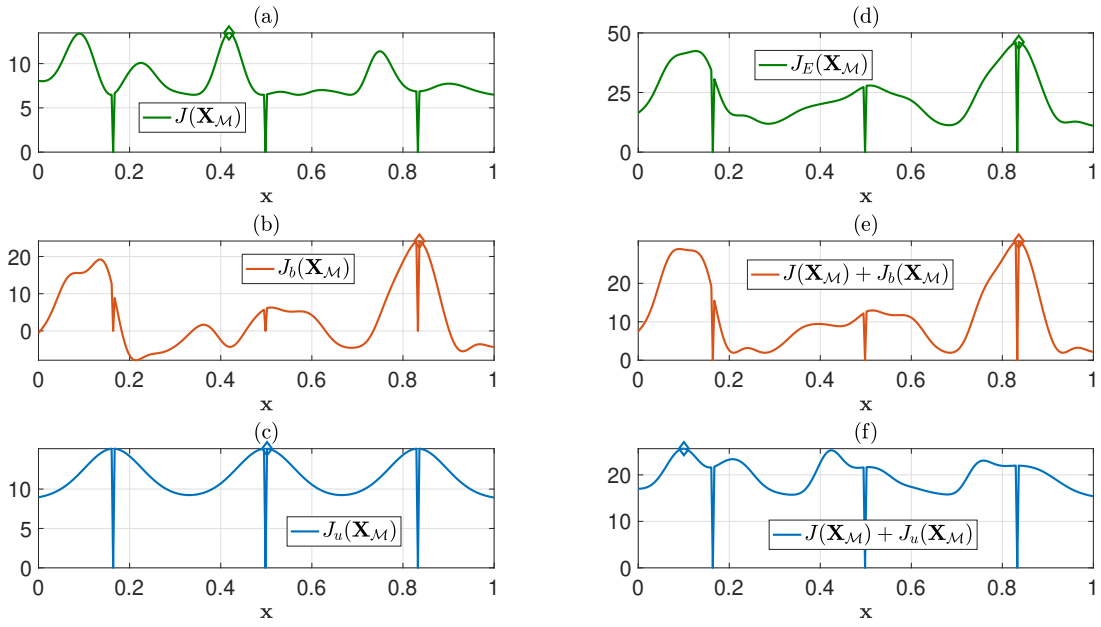


FIGURE 3.4: The effect of the bias and the uncertainty on the proposed criterion. In the left column, from top to bottom, we separately demonstrate the true value of the SNR with the perfect knowledge of the spatial gain, *i.e.* $J(\mathbf{X}_\mathcal{M})$, the bias function $J_b(\mathbf{X}_\mathcal{M})$, and the uncertainty function $J_u(\mathbf{X}_\mathcal{M})$. In the right column, the top sub-figure shows the proposed criterion $J_E(\mathbf{X}_\mathcal{M}) = J(\mathbf{X}_\mathcal{M}) + J_b(\mathbf{X}_\mathcal{M}) + J_u(\mathbf{X}_\mathcal{M})$, the middle one presents only the effect of the bias on the proposed criterion, *i.e.* $J(\mathbf{X}_\mathcal{M}) + J_b(\mathbf{X}_\mathcal{M})$, and the last sub-figure shows the effect of uncertainty by ignoring the effect of the bias, *i.e.* $J(\mathbf{X}_\mathcal{M}) + J_u(\mathbf{X}_\mathcal{M})$.

3.4.4 Comparison with state of the art

In this part, firstly, the sensitivity of the methods to the smoothness of the noise relative to the smoothness of the spatial gain, *i.e.* ρ_n/ρ_a are compared. In Fig. 3.5 the actual SNRs are plotted according to different ratios ρ_n/ρ_a. In this experiment, a grid search method is used : $P = 200$ linearly spaced possible sensor locations were tested between 0 and 1, with three initial sensors arbitrary located at the set of indices $\mathcal{K} = \{33, 100, 167\}$, corresponding to the positions $\mathbf{x}_\mathcal{K} = [x_{33},\ x_{100},\ x_{167}]^T = [0.1608,\ 0.4975,\ 0.8342]^T$. To generate $\tilde{\mathsf{a}}(x)$ and its mean $m_a(x)$, we set $\rho_a = \rho_{m_a} = 0.05$ and $\sigma_a = \sigma_{m_a} = 0.5$. Also, we put $\sigma_s = 1$.

To set up the parameters of the noise $n(x,t)$, we consider two different situations. In the first one shown in Fig. 3.5-(a), we set σ_n such that the SNR (3.24) is equal to 3dB. In the second case presented in Fig. 3.5-(b), we assume a more difficult situation by setting the power parameter of the noise σ_n such that the SNR is equal to 0 dB. For each of these two cases, we take into account different smoothness degrees for the noise compared to the smoothness of the spatial gain, and we choose 35 different values for ρ_n/ρ_a, starting from a white noise $\rho_n/\rho_a = 0.01$ to a very smooth noise $\rho_n/\rho_a = 25$. For each experiment, we generate 50 Monte Carlo (MC) of the noise $\mathsf{n}(x,t)$ and the spatial gain $\tilde{\mathsf{a}}(x)$.

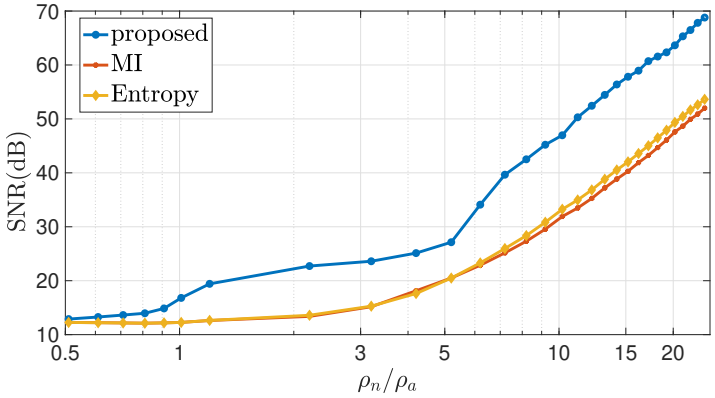
From Fig. 3.5-(a), it can be concluded that the proposed method performs better in average than the other methods in different smoothness situations. For instance, consider the point where $\rho_n/\rho_a = 3$; the resulting SNR after adding one new sensor using the proposed method is about 23dB, while the value of SNRs using MI and entropy are around 13dB. Note that depending on the ratio ρ_n/ρ_a, the optimal location of the new sensor varies. We also mention that for very small amounts of ρ_n/ρ_a, none of the compared methods provide a significant improvement. Next, by looking at Fig. 3.5-(b), we can see that even in a more difficult situation in terms of the SNR, the proposed method still outperforms the classical kriging methods. However, this superiority is less than the previous situation due to the higher level of the noise. Note that in sub-figure-a, the average standard error of the mean (SEM) for the average SNR is 0.18, 0.38, and 0.31 for the proposed method, MI and entropy, respectively. Also, these values for second case in sub-figure-b are equal to 0.23, 0.16, and 0.16, respectively. We recall that the SEM is calculated as follows :

$$\text{SEM} = \frac{\text{SD}}{\sqrt{N_{MC}}} \tag{3.25}$$

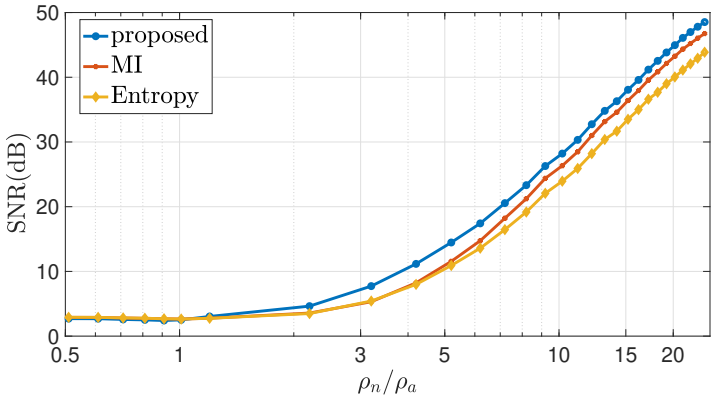
where $N_{MC} = 50$ is the number of MCMC, and SD is the standard deviation defined as

$$\text{SD} = \sqrt{\frac{1}{N_{MC}-1}\sum_{i=1}^{N_{MC}}(\text{SNR} - \overline{\text{SNR}})^2} \tag{3.26}$$

with $\overline{\text{SNR}}$ being the mean of the SNR.



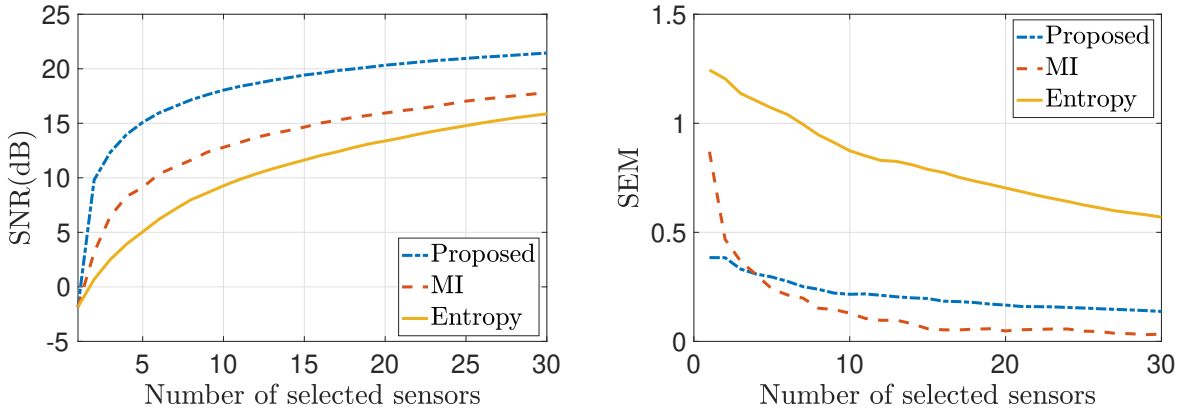
(a) SNR = 3dB



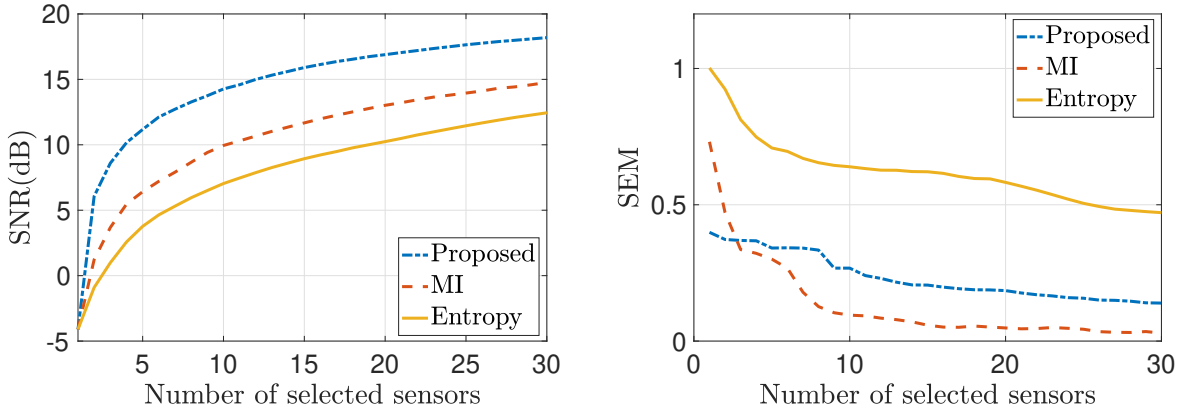
(b) SNR = 0dB

FIGURE 3.5: The dependency of the performance on the smoothness ratio between noise and spatial gain $\frac{\rho_n}{\rho_a}$. In this figure, the average SEM for the proposed method, MI and entropy is, respectively, equal to 0.18, 0.38, and 0.31 in part (a), and 0.23, 0.16, and 0.16 in part (b).

In the second part of the simulations (Fig. 3.6), the efficiency of the proposed method to choose the sensor locations is tested. To set up the simulations, we set the parameters as in Fig. 3.5, except that we start with one initial sensor in the middle of the grid at $x = 0.5$, and we set the ratio between the smoothness parameter of the noise and of the spatial gain to be $\rho_n/\rho_a = 0.01$. We do our experiment for two different values of the SNR equal to 3dB and 0dB, respectively, in sub-figures (a) and (b). A number of 50 data sets are also randomly generated for each number of sensors. The sensor locations were chosen with a greedy approach : at each iteration, a single new sensor is added to the previous set of sensors. Three different criteria are compared to select the new sensor location : the proposed method, MI and entropy. Fig. 3.6 shows the actual output SNR (3.23) for each size of the set of sensors on the left column, and the SEM (3.26) for the average SNR on the right column. From the results, we can see that the performance of our proposed method is significantly better than the other two methods, specially when the number of sensors is very limited. For instance, if a maximum number of 10 sensors can be used, by looking at Fig. 3.5-(a) we see that the results achieved by the proposed method is about 11dB better than entropy and also 6dB better than MI. Also, to reach an SNR of 15dB, the proposed criterion needs about 5 sensors, while 15 sensors are required using the MI method and we need 30 sensors with the entropy criterion. This shows the efficiency of the proposed approach to choose relevant sensor locations for source extraction. We can have similar interpretation from Fig. 3.5-(b), and we can see that even in a harder situation regarding the SNR, the proposed method keeps its significant superiority for sensor placement. In both cases, we can see that the entropy criterion does not perform as well as MI and the proposed criterion, and the results obtained by this criterion have a high SEM. We also note that, for small number of sensors the proposed criterion has a better performance in terms of SEM. However, MI outperforms our method in this regard by increasing the number of sensors.



(a) SNR = 3dB



(b) SNR = 0dB

FIGURE 3.6: Comparing the performance of the proposed method with entropy and MI. Left : improvement of the average SNR for 50 MC versus the number of selected sensors using greedy method. Right : The standard error of the mean (SEM) for the average output SNR versus the number of selected sensor.

3.5 Conclusion

In this chapter we studied the problem of optimal sensor placement for signal extraction from noisy underdetermined measurements collected by a limited number of sensors. Based on a linear estimator to extract the source of interest, the average output SNR of the linearly extracted signal has been proposed as a quality criterion to select sensor locations. Such a criterion takes into account the uncertainty on the spatial gain of the source to be extracted, and it provides a suitable solution for the optimal sensor placement problem. Numerical simulations have shown the superior efficiency and accuracy of the proposed method in the source extraction problem when compared to the classical sensor placement criteria such as entropy and mutual information. This superiority is proven from different aspects, such as the smoothness conditions of the noise, and different SNR situations.

As we mentioned in this chapter, our proposed method is not able to retrieve the positions where placing new sensors may deteriorate the extraction of the source and reduces the output SNR due to the highly uncertain knowledge and the bias $b(x)$. In the next chapter, we will go into a more in-depth analysis of the influence of modeling parameters, and we will introduce a new robust algorithm based on the pdf of the SNR.

Chapitre 4

Robust sensor placement for signal extraction

Sommaire

In the previous chapter, by considering the noisy sensor measurements model

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

and assuming a GP model for the spatial gain $a(\mathbf{x})$ and for the additive noise $n(\mathbf{x}, t)$, the output SNR of the estimated source signal was derived as a function of $a(\mathbf{x})$ and of the sensor locations $\mathbf{X}_{\mathcal{M}}$, by using a linear source extraction approach. To model the imperfect knowledge on the spatial gain, a stochastic model on the spatial gain $a(\mathbf{x})$ is considered. As a consequence, the output SNR is also stochastic. So, we suggested the average SNR as a

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, the selected sensor positions may not improve the SNR, and they may even decrease the SNR.

In this chapter, we propose to extend the idea presented in the previous chapter 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 pdf of the SNR, we indicate how this pdf can be evaluated using the same GP assumption on the spatial gain. Moreover, when we choose the sensor positions in a sequential fashion, we show that the criterion can 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. Furthermore, the results confirm the superior robustness of this approach compared to the average SNR criterion.

The rest of this chapter is organized as follows. In Section 4.1, we present the problem statement. In Section 4.2, the proposed criterion for sensor placement based on the pdf of the SNR is introduced. Section 4.3 presents how the criterion becomes simpler with the sequential approach. In Section 4.4, the simulation results are presented. Then, in Section 4.5, some more numerical tools for a better evaluation of different methods will be presented. Finally, Section 4.6 concludes this chapter.

4.1 Problem statement

In this section, we will present the problem that we want to address. Assume that M sensors are going to be used at locations $\mathbf{X}_{\mathcal{M}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M]^T$ to extract a source signal $s(t)$. The set of noisy measurements related to each sensor is denoted by the vector $\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t) = [y(\mathbf{x}_1, t), y(\mathbf{x}_2, t), \ldots, y(\mathbf{x}_M, t)]^T$ described through the following model

$$\mathbf{y}(\mathbf{X}_{\mathcal{M}}, t) = \mathbf{a}(\mathbf{X}_{\mathcal{M}})s(t) + \mathbf{n}(\mathbf{X}_{\mathcal{M}}, t), \tag{4.1}$$

where $\mathbf{n}(\mathbf{X}_{\mathcal{M}}, t) = [n(\mathbf{x}_1, t), n(\mathbf{x}_2, t), \ldots, n(\mathbf{x}_M, t)]^T$ and $\mathbf{a}(\mathbf{X}_{\mathcal{M}}) = [a(\mathbf{x}_1), a(\mathbf{x}_2), \ldots, a(\mathbf{x}_M)]^T$ are, respectively, the corresponding noise and spatial gain vectors. In the previous chapter we showed that by considering the observation model in (4.1), and using a linear estimator described by the extractor vector $\mathbf{f}(\mathbf{X}_{\mathcal{M}}) \in \mathbb{R}^M$, the extracted source signal $\hat{s}(t)$ is given by

$$\hat{s}(t) = \mathbf{f}(\mathbf{X}_{\mathcal{M}})^T \mathbf{y}(\mathbf{X}_{\mathcal{M}}, t) = \mathbf{f}(\mathbf{X}_{\mathcal{M}})^T \mathbf{a}(\mathbf{X}_{\mathcal{M}})s(t) + \mathbf{f}(\mathbf{X}_{\mathcal{M}})^T \mathbf{n}(\mathbf{X}_{\mathcal{M}}, t). \tag{4.2}$$

Also, we showed from above, the SNR of the estimation of $s(t)$ is derived as

$$\mathrm{SNR}(\mathbf{f}(\mathbf{X}_{\mathcal{M}})|\mathbf{X}_{\mathcal{M}}) = \frac{\mathbb{E}_t\big[\big(\mathbf{f}(\mathbf{X}_{\mathcal{M}})^T\mathbf{a}(\mathbf{X}_{\mathcal{M}})s(t)\big)^2\big]}{\mathbb{E}_t\big[\big(\mathbf{f}(\mathbf{X}_{\mathcal{M}})^T\mathbf{n}(\mathbf{X}_{\mathcal{M}},t)\big)^2\big]} = \frac{\sigma_s^2\mathbf{f}(\mathbf{X}_{\mathcal{M}})^T\mathbf{a}(\mathbf{X}_{\mathcal{M}})\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T\mathbf{f}(\mathbf{X}_{\mathcal{M}})}{\mathbf{f}(\mathbf{X}_{\mathcal{M}})^T\mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})\mathbf{f}(\mathbf{X}_{\mathcal{M}})}, \tag{4.3}$$

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 $\sigma_s^2 = \mathbb{E}_t[s(t)^2]$ and $\mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}}) = \mathbb{E}^t[\mathbf{n}(\mathbf{X}_{\mathcal{M}},t)\mathbf{n}^T(\mathbf{X}_{\mathcal{M}},t)]$. Accordingly, the linear estimator that maximizes the SNR is

$$\mathbf{f}^*(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}\mathbf{a}(\mathbf{X}_{\mathcal{M}}) \tag{4.4}$$

and the corresponding SNR is given by

$$\mathrm{SNR}(\mathbf{f}^*(\mathbf{X}_{\mathcal{M}})|\mathbf{X}_{\mathcal{M}}) = \sigma_s^2\,\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T\mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}\mathbf{a}(\mathbf{X}_{\mathcal{M}}). \tag{4.5}$$

Now, by treating the resulting output SNR in (4.5) as a function of sensor locations $\mathbf{X}_{\mathcal{M}}$ and maximizing it over $\mathbf{X}_{\mathcal{M}}$, 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 $a(\mathbf{x}_i)$, $i \in \{1,\ 2,\ \ldots,\ M\}$. 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 the spatial gain as $\boldsymbol{a}(\mathbf{x})$:

$$\boldsymbol{a}(\mathbf{x}) \sim \mathcal{GP}\Big(m^a(\mathbf{x}), C^a(\mathbf{x},\mathbf{x}')\Big). \tag{4.6}$$

Consequently, the randomness of $\boldsymbol{a}(\mathbf{x})$ leads to a stochastic SNR :

$$\mathrm{SNR}(\boldsymbol{f}^*(\mathbf{X}_{\mathcal{M}})|\mathbf{X}_{\mathcal{M}}) = \sigma_s^2\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})^T\mathbf{R}(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})\boldsymbol{a}(\mathbf{X}_{\mathcal{M}}), \tag{4.7}$$

where $\boldsymbol{f}^*(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})$ is the corresponding random extractor vector, and assuming the noise covariance matrix to be invertible, we have defined

$$\mathbf{R}(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}}) \triangleq \mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}. \tag{4.8}$$

To deal with this stochastic representation of the SNR, in Chapter 3 we suggested to estimate the SNR with its mean

$$\begin{aligned} J_E(\mathbf{X}_{\mathcal{M}}) =& \mathbb{E}\Big\{\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})^T\mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})\Big\} \\ =& \mathbf{m}^a(\mathbf{X}_{\mathcal{M}})^T\mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}\mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) + \mathrm{Tr}\Big[\mathbf{C}^n(\mathbf{X}_{\mathcal{M}},\mathbf{X}_{\mathcal{M}})^{-1}\mathbf{C}^a(\mathbf{X}_{\mathcal{M}},\ \mathbf{X}_{\mathcal{M}})\Big]. \end{aligned}$$

However, in a very uncertain condition where the covariance of the spatial gain $C^a(\mathbf{x},\mathbf{x}')$ takes large values compared to its mean $m^a(\mathbf{x})$, it is possible to choose sensor positions where the mean of the SNR is quite large, but at the same time we have a high dispersion causing the actual SNR to be much smaller. Therefore, to target the SNR in (4.5), we need to use a quantitative statistical measure of the SNR distribution which is robust against the uncertainty in the spatial gain model (4.6), *i.e.* large values of $C^a(\mathbf{x},\mathbf{x}')$. This will be presented in the next part.

Remark 1
If we have a prior information on $\boldsymbol{a}(\boldsymbol{X}_\mathcal{K}) = [a(\boldsymbol{x}_1), a(\boldsymbol{x}_2), \ldots, a(\boldsymbol{x}_K)]^T$*, for example if we have previously placed the sensors at these positions and measured the gains, then the prior (4.6) can be updated to include this information, simply by conditioning the GP on the measurements of the gains. As explained in Chapter 2.1, the conditioned GP is still a GP but with conditioned mean and covariance functions presented in* (2.16) *and* (2.18)*, respectively [RW06].*

4.2 Proposed criterion

Since the pdf of a random variable contains all the information about it, here, by computing 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 (4.7). Then, based on the obtained pdf, we present a robust criterion for sensor placement.

To simplify notation, we define the random variable

$$\mathsf{w}(\mathbf{X}_\mathcal{M}) \triangleq \frac{1}{\sigma_s^2}\,\mathrm{SNR}(\boldsymbol{f}^*(\mathbf{X}_\mathcal{M})|\mathbf{X}_\mathcal{M}) = \boldsymbol{a}(\mathbf{X}_\mathcal{M})^T\mathbf{C}^n(\mathbf{X}_\mathcal{M},\mathbf{X}_\mathcal{M})^{-1}\boldsymbol{a}(\mathbf{X}_\mathcal{M}). \tag{4.9}$$

Now, in order to find the distribution of $\mathsf{w}(\mathbf{X}_\mathcal{M})$, we propose the following :

Proposition 1 (Distribution of $\mathsf{w}(\mathbf{X}_\mathcal{M})$)
If the spatial gain $\mathsf{a}(\boldsymbol{x})$ *follows the GP model* $\mathsf{a}(\boldsymbol{x}) \sim \mathcal{GP}\Big(m^a(\boldsymbol{x}), C^a(\boldsymbol{x},\boldsymbol{x}')\Big)$*, then* $\mathsf{w}(\boldsymbol{X}_\mathcal{M})$ *is the weighted sum of* M *independent random variables as follows :*

$$\mathsf{w}(\boldsymbol{X}_\mathcal{M}) = \sum_{i=1}^{M} d_i \mathsf{y}_i^2, \tag{4.10}$$

where y_i*'s are independent normally distributed random variables* $\mathsf{y}_i \sim \mathcal{N}\left(m_{\mathsf{y}_i}, 1\right)$*, with the mean*

$$m_{\mathsf{y}_i} = \boldsymbol{u}_i^T\boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M},\boldsymbol{X}_\mathcal{M})^{-\frac{1}{2}}\boldsymbol{m}^a(\boldsymbol{X}_\mathcal{M}) \tag{4.11}$$

and, d_i*'s and* $\boldsymbol{u}_i$*'s are, respectively, the eigenvalues and eigenvectors of the matrix* $\boldsymbol{A}$ *defined as follows :*

$$\boldsymbol{A} \triangleq \boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M},\boldsymbol{X}_\mathcal{M})^{\frac{1}{2}}\boldsymbol{R}(\boldsymbol{X}_\mathcal{M},\boldsymbol{X}_\mathcal{M})\boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M},\boldsymbol{X}_\mathcal{M})^{\frac{1}{2}}, \tag{4.12}$$

and $\boldsymbol{m}^a(\boldsymbol{X}_\mathcal{M}) = [m^a(\boldsymbol{x}_1),\ m^a(\boldsymbol{x}_2),\ \ldots,\ m^a(\boldsymbol{x}_\mathcal{M})]^T$.

To prove the above proposition, we first need to present the following lemma.

Lemma 1
Let $\boldsymbol{q}$ *be a normally distributed random vector with zero mean and an identity covariance matrix* $\boldsymbol{I}_M$ *:* $\boldsymbol{q} \sim \mathcal{N}(0, \boldsymbol{I}_M)$*. Now, we define the random variable* y_i *as follows :*

$$\mathsf{y}_i \triangleq \boldsymbol{u}_i^T(\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_\mathcal{M})), \tag{4.13}$$

where

$$\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}) = \boldsymbol{C}^a(\boldsymbol{X}_{\mathcal{M}}, \boldsymbol{X}_{\mathcal{M}})^{-\frac{1}{2}} \boldsymbol{m}^a(\boldsymbol{X}_{\mathcal{M}}). \tag{4.14}$$

Then, it is straightforward to show that for any $i \neq j \in \{1, 2, \ldots, M\}$, y_i *and* y_j *are uncorrelated :*

$$Corr(\mathsf{y}_i, \mathsf{y}_j) = \frac{\mathbb{E}\Big\{(\mathsf{y}_i - \mathbb{E}\{\mathsf{y}_i\})(\mathsf{y}_j - \mathbb{E}\{\mathsf{y}_j\})\Big\}}{\sqrt{\mathbb{E}\{\mathsf{y}_i^2\} - \mathbb{E}\{\mathsf{y}_i\}^2}\sqrt{\mathbb{E}\{\mathsf{y}_j^2\} - \mathbb{E}\{\mathsf{y}_j\}^2}} = 0. \tag{4.15}$$

Taking into account that $\mathbb{E}\{\boldsymbol{q}\} = 0$, *from* (4.13) *we have*

$$\mathbb{E}\{\mathsf{y}_i\} = \mathbb{E}\{\boldsymbol{u}_i^T(\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}))\} = \boldsymbol{u}_i^T\mathbb{E}\{\boldsymbol{q}\} + \boldsymbol{u}_i^T\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}) = \boldsymbol{u}_i^T\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}), \tag{4.16}$$

and similarly $\mathbb{E}\{\mathsf{y}_j\} = \boldsymbol{u}_j^T\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})$. *So, now we have*

$$\begin{aligned}
&\mathbb{E}\Big\{(\mathsf{y}_i - \mathbb{E}\{\mathsf{y}_i\})(\mathsf{y}_j - \mathbb{E}\{\mathsf{y}_j\})\Big\} = \\
&\mathbb{E}\Big\{\big[\boldsymbol{u}_i^T(\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})) - \boldsymbol{u}_i^T\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})\big]\big[\boldsymbol{u}_j^T(\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})) - \boldsymbol{u}_j^T\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})\big]\Big\} = \\
&\mathbb{E}\Big\{\boldsymbol{u}_i^T\boldsymbol{q}\boldsymbol{u}_j^T\boldsymbol{q}\Big\} = \mathbb{E}\Big\{\boldsymbol{u}_i^T\boldsymbol{q}\boldsymbol{q}^T\boldsymbol{u}_j\Big\} = \boldsymbol{u}_i^T\mathbb{E}\Big\{\boldsymbol{q}\boldsymbol{q}^T\Big\}\boldsymbol{u}_j = \boldsymbol{u}_i^T\boldsymbol{I}\boldsymbol{u}_j = \boldsymbol{u}_i^T\boldsymbol{u}_j = 0.
\end{aligned} \tag{4.17}$$

Note that, since $\boldsymbol{u}_i$ *and* $\boldsymbol{u}_j$ *are the eigenvectors of* $\boldsymbol{A}$ (4.12), *and* $\boldsymbol{A}$ *is symmetric, they are orthogonal and* $\boldsymbol{u}_i^T\boldsymbol{u}_j = 0$. *Now, by replacing* (4.17) *in the numerator of* (4.15), *we can conclude that* $Corr(\mathsf{y}_i, \mathsf{y}_j) = 0$, *and thus,* 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* $h : \mathbb{R}^M \longrightarrow \mathbb{R}^{M'}$ *is diagonal, the non-linear functions will not affect the independence of the random variables. So, by defining* $h(\mathsf{y}_i) \triangleq d_i\mathsf{y}_i^2$, *since the Jacobian matrix of the function* h *over the independent random variables* $\mathsf{y}_i, i \in \{1, 2, \ldots, M\}$ *is diagonal, it can be concluded that for any* $i \neq j \in \{1, 2, \ldots, M\}$, $\Gamma_i \triangleq d_i\mathsf{y}_i^2$ *and* $\Gamma_j \triangleq d_j\mathsf{y}_j^2$ *are also independent.* ■

Proof 1 (proof of Proposition 1)
From (4.13), *the squared form of* y_i *is as follows :*

$$\begin{aligned}
\mathsf{y}_i^2 =& [\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})]^T\boldsymbol{u}_i\boldsymbol{u}_i^T[\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})] \\
=& \boldsymbol{q}^T\boldsymbol{u}_i\boldsymbol{u}_i^T\boldsymbol{q} + 2(\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}))^T\boldsymbol{u}_i\boldsymbol{u}_i^T\boldsymbol{q} + (\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}))^T\boldsymbol{u}_i\boldsymbol{u}_i^T(\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})).
\end{aligned}$$

According to the above equation, we have the following :

$$\begin{aligned}
\sum_{i=1}^{M} d_i\mathsf{y}_i^2 =& \boldsymbol{q}^T\Bigg[\sum_{i=1}^{M}(d_i\boldsymbol{u}_i\boldsymbol{u}_i^T)\Bigg]\boldsymbol{q} + 2(\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}))^T\Bigg[\sum_{i=1}^{M}(d_i\boldsymbol{u}_i\boldsymbol{u}_i^T)\Bigg]\boldsymbol{q} \\
&+ (\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}))^T\Bigg[\sum_{i=1}^{M}(d_i\boldsymbol{u}_i\boldsymbol{u}_i^T)\Bigg]\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}}).
\end{aligned}$$

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

$$\sum_{i=1}^{M} d_i\mathsf{y}_i^2 = \boldsymbol{q}^T\boldsymbol{A}\boldsymbol{q} + 2\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})^T\boldsymbol{A}\boldsymbol{q} + \boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})^T\boldsymbol{A}(\boldsymbol{m}'(\boldsymbol{X}_{\mathcal{M}})). \tag{4.18}$$

By replacing (4.14) *and* (4.12) *in* (4.18)*, we obtain*

$$\sum_{i=1}^{M} d_i \mathsf{y}_i^2 = \Big[\boldsymbol{m}^a(\boldsymbol{X}_\mathcal{M}) + \boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M}, \boldsymbol{X}_\mathcal{M})^{\frac{1}{2}}\boldsymbol{q}\Big]^T \boldsymbol{R}(\boldsymbol{X}_\mathcal{M}, \boldsymbol{X}_\mathcal{M})\Big[\boldsymbol{m}^a(\boldsymbol{X}_\mathcal{M}) + \boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M}, \boldsymbol{X}_\mathcal{M})^{\frac{1}{2}}\boldsymbol{q}\Big].$$

Since $\boldsymbol{q}$ *is a normally distributed random vector with zero mean and an identity covariance matrix, and also because* $\boldsymbol{m}^a(\boldsymbol{X}_\mathcal{M})$ *and* $\boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M}, \boldsymbol{X}_\mathcal{M})$ *are the mean vector and covariance matrix of the model of the spatial gain, respectively, we conclude that* $\boldsymbol{m}^a(\boldsymbol{X}_\mathcal{M}) + \boldsymbol{C}^a(\boldsymbol{X}_\mathcal{M}, \boldsymbol{X}_\mathcal{M})^{\frac{1}{2}}\boldsymbol{q}$ *follows the same distribution as* $\boldsymbol{a}(\boldsymbol{X}_\mathcal{M})$*. Therefore, we have*

$$\sum_{i=1}^{M} d_i \mathsf{y}_i^2 = \boldsymbol{a}(\boldsymbol{X}_\mathcal{M})^T \boldsymbol{R}(\boldsymbol{X}_\mathcal{M}, \boldsymbol{X}_\mathcal{M})\boldsymbol{a}(\boldsymbol{X}_\mathcal{M})^T = \frac{1}{\sigma_s^2} SNR(\boldsymbol{f}^*(\boldsymbol{X}_\mathcal{M})|\boldsymbol{X}_\mathcal{M}) = \mathsf{w}(\boldsymbol{X}_\mathcal{M}). \quad \blacksquare$$

Lemma 2
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 non-central chi-squared distribution with the number of degrees of freedom* $k_i = 1$*, and non-centrality parameter* $\lambda_i = m_{\mathsf{y}_i}^2$ *[AS72]. So, the pdf of the random variable* $\mathsf{v}_i \triangleq \mathsf{y}_i^2$ *becomes*

$$g_{\mathsf{v}_i}(v_i; k_i, \lambda_i) = \frac{1}{2}\exp^{-\frac{(v_i+\lambda_i)}{2}} \left(\frac{v_i}{\lambda_i}\right)^{(\frac{k_i}{4}-\frac{1}{2})} I_{\frac{k_i}{2}-1}(\sqrt{\lambda_i v_i}), \tag{4.19}$$

where $I.(\cdot)$ *is the modified Bessel function of the first kind. Now, by defining* $\Gamma_i = d_i \mathsf{v}_i$ *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 :*

$$g_{\Gamma_i}(\gamma_i) = \frac{1}{d_i} g_{\mathsf{v}_i}(\frac{\gamma_i}{d_i}; k_i, \lambda_i), \tag{4.20}$$

with $g_{\mathsf{v}_i}(.)$ *being the pdf of the random variable* $\mathsf{v}_i = \mathsf{y}_i^2$ *with non-central chi-squared distribution defined in* (4.19)*.* $\blacksquare$

Proposition 2 (Distribution of $\mathsf{w}(\mathbf{X}_\mathcal{M}) = \mathrm{SNR}(\boldsymbol{f}^*(\mathbf{X}_\mathcal{M})|\mathbf{X}_\mathcal{M})/\sigma_s^2$)
From Lemma 1 and Lemma 2, it can be concluded that $\mathsf{w}(\boldsymbol{X}_\mathcal{M}) = SNR(\boldsymbol{f}^*(\boldsymbol{X}_\mathcal{M})|\boldsymbol{X}_\mathcal{M})/\sigma_s^2$ *is the sum of* M *independent random variables* Γ_i *each having a pdf defined in* (4.20)*. 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 *:*

$$\begin{aligned} g_\mathsf{w}(w) =& g_{\Gamma_1}(w) * g_{\Gamma_2}(w) * \cdots * g_{\Gamma_M}(w) \\ =& \frac{1}{\prod_{i=1}^{M} d_i} g_{\mathsf{v}_1}(\frac{w}{d_1}; k_1, \lambda_1) * g_{\mathsf{v}_2}(\frac{w}{d_2}; k_2, \lambda_2) * \cdots * g_{\mathsf{v}_M}(\frac{w}{d_M}; k_M, \lambda_M). \quad \blacksquare \end{aligned} \tag{4.21}$$

Now, by knowing the pdf of $\mathsf{w}(\mathbf{X}_\mathcal{M}) = \mathrm{SNR}(\boldsymbol{f}^*(\mathbf{X}_\mathcal{M})|\mathbf{X}_\mathcal{M})/\sigma_s^2$, 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. 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, and we introduce the criterion $J_P(\mathbf{X}_{\mathcal{M}}, \theta)$ which is based on the cumulative distribution function (cdf) of the SNR :

$$J_P(\mathbf{X}_{\mathcal{M}}, \theta) = Pr(w(\mathbf{X}_{\mathcal{M}}) > \theta) = 1 - G_w(\theta), \tag{4.22}$$

where $G_w(\cdot)$ denotes the cdf of the SNR$/\sigma_s^2$ (*i.e.* $w(\mathbf{X}_{\mathcal{M}})$) conditioned on the sensors at positions $\mathbf{X}_{\mathcal{M}}$. By knowing the pdf of $w(\mathbf{X}_{\mathcal{M}})$, from (4.21) the cdf becomes as follows :

$$G_w(\theta; \mathbf{k}, \Lambda) = \frac{1}{\prod_{i=2}^{M} d_i} \times G_{v_1}\Big(\frac{\theta}{d_1}; k_1, \lambda_1\Big) * g_{v_2}\Big(\frac{\theta}{d_2}; k_2, \lambda_2\Big) * \cdots * g_{v_M}\Big(\frac{\theta}{d_M}; k_M, \lambda_M\Big), \tag{4.23}$$

where $\mathbf{k} = \{k_1, k_2, \ldots, k_M\}$, $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$, and $G_{v_1}(\cdot)$ is the corresponding cdf of $g_{v_1}(\cdot)$. The proof of (4.23) is presented in Appendix A. Note that from Lemma 2 we also have :

$$G_{v_1}(v_1; k_1, \lambda_1) = 1 - Q_{\frac{k_1}{2}}(\sqrt{\lambda_1}, \sqrt{v_1}), \tag{4.24}$$

where λ_i and k_i are defined in Lemma 2, and Q is the Marcum-Q-function [NA75]. So, based on the criterion (4.22), we solve the following problem for sensor placement :

$$\hat{\mathbf{X}}_{\mathcal{M}} = \underset{\mathbf{X}_{\mathcal{M}}}{\operatorname{argmax}}\, J_P(\mathbf{X}_{\mathcal{M}}, \theta), \tag{4.25}$$

Remark 2 (Links with maximum likelihood)
It is worth mentioning that if we consider a target value α of SNR$/\sigma_s^2$, then the maximum likelihood can also be used as a sensor placement criterion. In this case, we look for the sensor positions matrix $\boldsymbol{X}_{\mathcal{M}}$ maximizing

$$J_{ML}(\boldsymbol{X}_{\mathcal{M}}, \alpha) = g_w(\alpha). \tag{4.26}$$

However, this criterion does not have an intuitive interpretation regarding robustness and maximization of SNR as (4.22).

The advantage of the criterion (4.22) is that the free parameter θ can be used to control the risk we want to take when placing new sensors. The effect of θ is depicted in Fig. 4.1. In this figure, we have plotted the pdf of the SNR, *i.e.* $g_w(w)$ (4.21), in two example positions : $\mathbf{x}_1$ which has a low variance, and $\mathbf{x}_2$ which has a large variance. Then, we considered two different cases. In the first case, the parameter θ is set to a value equal to 10. In the second case, we reduced the parameter θ to a smaller value equal to 5. The green shadow shows the cdf of the SNR (our proposed criterion) which depends on the value we choose for the parameter θ. From this figure, we can see that by increasing θ to sufficiently high values (sub-figures 4.1a and 4.1b), the upper tail of the output SNR will be compared for two different positions, forcing the criterion to be high for positions leading to a high SNR but which are conservative (having smaller variance). 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

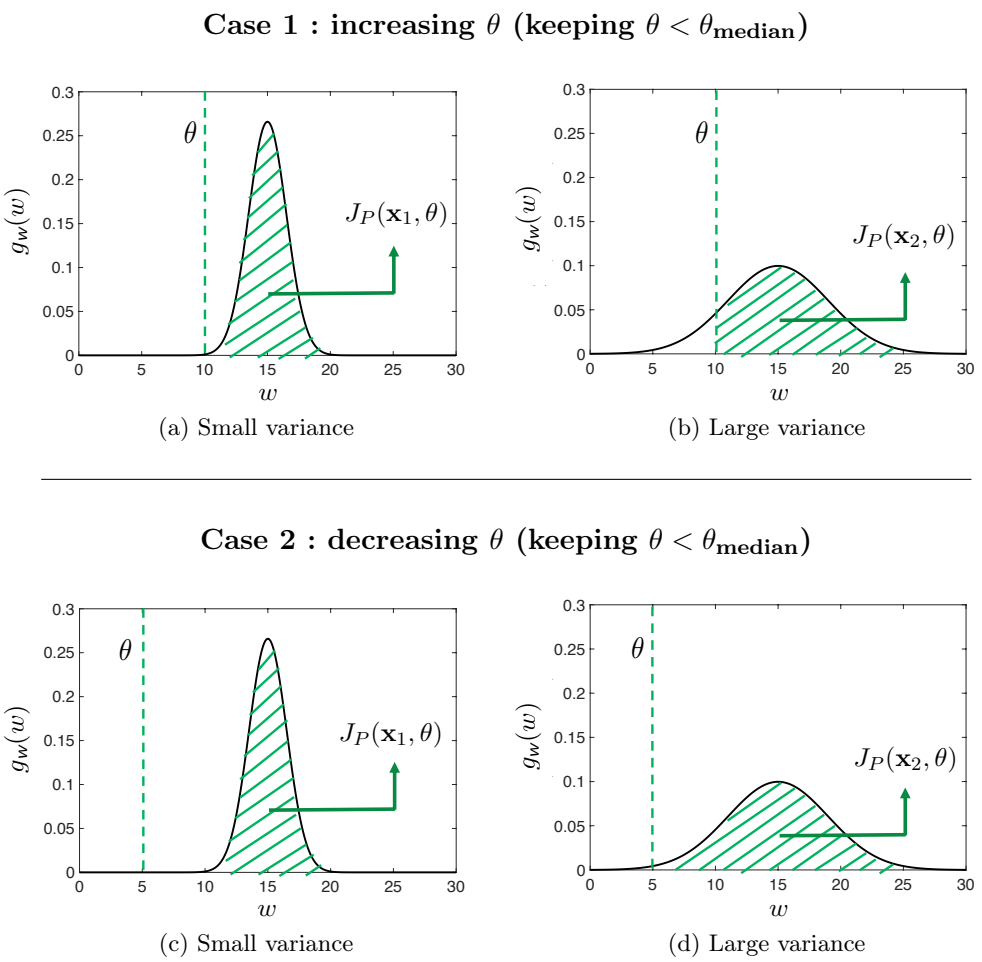


FIGURE 4.1: The effect of θ on the cdf of the SNR (the proposed criterion $J_P(\mathbf{x}, \theta)$). a) In the case 1 we see that increasing θ forces the criterion to be high for positions leading to a high SNR but which are conservative (having low variance). b) In the case 2 we see that by reducing the parameter θ most of the positions will have similar criterion values closed to 1, and consequently, risky sensor positions with a very large average SNR, but at the same time, a large uncertainty, will not be discarded.

value (sub-figures 4.1c and 4.1d), most of the positions will have similar criterion values close to 1 (the green shadows), and, as a consequence, risky sensor positions (the positions with large variance) leading to a very high average SNR values will not be discarded. We note that this result is valid if we consider that θ is always smaller than the median θ_{median} [1], and if we set $\theta > \theta_{\text{median}}$ the results are in opposite. 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. We note that, in Fig.4.1, we have considered that the pdf has the same mean equal to 15 at all the positions. However, this does not hold in practice, and the pdf can take different mean values at different positions. On the other hand, selecting the parameter θ depends on the mean of

1. The median of a population is the value where at least half of the population is less than this value.

the pdf. This means that for different positions, different values for θ should be considered based on the mean of the SNR at the corresponding positions. To solve this difficulty, one can shift the pdf of the SNR to its mean at each position, and then, set a fixed threshold for all the shifted pdfs.

4.3 Sequential sensor placement

In practice, to optimize the criterion (4.22), since there is no evidence that this criterion is neither convex nor monomodal, a grid search is used. We are thus looking for the best subset $\mathbf{X}_{\mathcal{M}}$ from a spatial grid of total P positions $\mathbf{X}_{\mathcal{P}}$ to place M sensors such that our proposed criterion in (4.22) is maximized. The optimal solution to this problem requires a combinatorial search over $\frac{P!}{M!(P-M)!}$ possibilities. This search has a high computational cost, especially if the grid is very tight, *i.e.* when P is large. Therefore, it is essential to provide a less costly solution.

In this section, we present two solutions that reduce the complexity. 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.

4.3.1 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 prediction, the underlying criterion becomes maximum for the totally selected positions up to each iteration. Assume that K sensors have already been located at $\mathbf{X}_{\mathcal{K}}$ and we want to add the next N sensors at $\mathbf{X}_{\mathcal{N}}$ such that $K+N$ is smaller than the total number of sensors M. Among all the possibilities for the N new sensors, we search for the subset $\mathbf{X}_{\mathcal{N}}$ such that $\mathbf{X}_{\mathcal{M}} = \{\mathbf{X}_{\mathcal{K}} \cup \mathbf{X}_N\}$ maximizes (4.22). This approach is presented in Algorithm (1). Since at each iteration $N < M$, the computational complexity will be reduced compared with a direct search for $\mathbf{X}_{\mathcal{M}}$. However, such complexity reduction comes with the cost of a possible sub-optimality of the chosen sensor positions.

4.3.2 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. For instance, when we start with some initial sensors at positions $\mathbf{X}_{\mathcal{K}}$, we consider an approximation for the spatial gain based on the observations obtained by the sensors at $\mathbf{X}_{\mathcal{K}}$. Then, by adding the new N sensors and using the information obtained by the new measurements, we can apply some methods such as

Algorithm 1 Greedy method

Inputs : $K = 0$, $\mathbf{X}_{\mathcal{K}} = \emptyset$
At each iteration : prediction of N new sensors positions
for $j = 1, 2, \cdots, M$ **do**
    $\hat{\mathbf{X}}_{\mathcal{N}} = \text{argmax}_{\mathbf{X}_{\mathcal{N}}} J_P(\mathbf{X}_{\mathcal{N}} \cup \mathbf{X}_{\mathcal{K}}; \theta)$
    $X_{\mathcal{K}} \longleftarrow \mathbf{X}_{\mathcal{N}} \cup \mathbf{X}_{\mathcal{K}}$
end for
$\mathbf{X}_{\mathcal{M}} \longleftarrow \mathbf{X}_{\mathcal{K}}$
Output : $\mathbf{X}_{\mathcal{M}}$

independent component analysis (ICA) [Com94], to update the value of the spatial gains, and have a better estimation of it. 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 $\boldsymbol{a}(\mathbf{x})$ 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. This approach is presented in Algorithm (2).

In practice, information on the spatial gains may be obtained either if a ground truth signal is available [SJ+18], or with blind signal processing techniques such as ICA [Com94] or sparse component analysis (SCA) [ZP01]. In the following, we suppose that one of these approaches is used and that the new information available on the gain is equivalent to the noisy measurements of it, denoted $\mathbf{z}(\mathbf{X}_{\mathcal{K}}) = [z(\mathbf{x}_1),\ z(\mathbf{x}_2), \ldots, z(\mathbf{x}_K)]^T$, which is obtained at the previously placed sensor positions $\mathbf{X}_{\mathcal{K}}$:

$$\mathbf{z}(\mathbf{X}_{\mathcal{K}}) = \mathbf{a}(\mathbf{X}_{\mathcal{K}}) + \boldsymbol{v}(\mathbf{X}_{\mathcal{K}}), \tag{4.27}$$

where $\mathbf{a}(\mathbf{X}_{\mathcal{K}})$ is considered to be the true value of the spatial gain at the preselected positions $\mathbf{X}_{\mathcal{K}}$, and $\boldsymbol{v}(\mathbf{x})$ is a zero mean Gaussian process independent of the measurement noise $\boldsymbol{n}(\mathbf{x}, t)$, with the covariance function $C^v(\mathbf{x}, \mathbf{x}')$. So, $z(\mathbf{x}) \sim \mathcal{GP}\big(a(\mathbf{x}), C^v(\mathbf{x}, \mathbf{x}')\big)$ is an unbiased estimation of the spatial gain $a(\mathbf{x})$.

Once we placed the $\mathbf{X}_{\mathcal{K}}$ sensors and measured $\mathbf{z}(\mathbf{X}_{\mathcal{K}})$, we can update the criterion for placing the new N sensors by conditioning the gain on $\mathbf{z}(\mathbf{X}_{\mathcal{K}})$. The gain vector can then be written as $\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})^T = \big[\mathbf{z}(\mathbf{X}_{\mathcal{K}})^T\ \boldsymbol{a}(\mathbf{X}_{\mathcal{N}})^T\big]$, where $\mathbf{z}(\mathbf{X}_{\mathcal{K}})$ are the estimated gains for which we have measurements, $\boldsymbol{a}(\mathbf{X}_{\mathcal{N}})$ is the random vector corresponding to the spatial gains at the candidate positions $\mathbf{X}_{\mathcal{N}}$, and both are conditioned on $\mathbf{z}(\mathbf{X}_{\mathcal{K}})$. As mentioned in Remark 1, since both the gains and $\mathbf{z}(\mathbf{X}_{\mathcal{K}})$ are Gaussian, the conditioned gains are Gaussian, too, with mean and covariance given by [RW06]

$$\begin{aligned}\mathbf{m}^a\big(\mathbf{X}_{\mathcal{M}}|\mathbf{z}(\mathbf{X}_{\mathcal{K}})\big) &= \mathbb{E}\Big\{\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})|\mathbf{z}(\mathbf{X}_{\mathcal{K}})\Big\} \\ &= \mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) + \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{K}})\Big[\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) + \mathbf{C}^v(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})\Big]^{-1}\Big[\mathbf{z}(\mathbf{X}_{\mathcal{K}}) - \mathbf{m}^a(\mathbf{X}_{\mathcal{K}})\Big],\end{aligned} \tag{4.28}$$

Algorithm 2 Sequential approach

1: **Inputs :** $K = 0$, $\mathbf{X}_{\mathcal{K}} = \emptyset$
2: **At each iteration :**
3: **for** $j = 1, 2, \cdots, M$ **do**
4: **Estimation of the spatial gain :** $\mathbf{a}(\mathbf{X}_{\mathcal{K}}) \longrightarrow \mathbf{z}(\mathbf{X}_{\mathcal{K}})$
5: **prediction of** N **new sensors positions :** $\hat{\mathbf{X}}_{\mathcal{N}} = \text{argmax}_{\mathbf{X}_{\mathcal{N}}} J_P\left(\mathbf{X}_{\mathcal{N}} \cup X_{\mathcal{K}}; \theta | \mathbf{z}(\mathbf{X}_{\mathcal{K}})\right)$
6: $\mathbf{X}_{\mathcal{K}} \longleftarrow \mathbf{X}_{\mathcal{N}} \cup \mathbf{X}_{\mathcal{K}}$
7: **end for**
8: $\mathbf{X}_{\mathcal{M}} \longleftarrow \mathbf{X}_{\mathcal{K}}$
9: **Output :** $\mathbf{X}_{\mathcal{M}}$

and

$$\begin{aligned}\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}} | \mathbf{z}(\mathbf{X}_{\mathcal{K}})) &= \mathbb{E}\left\{ \Big[\boldsymbol{a}(\mathbf{X}_{\mathcal{M}}) - \mathbf{m}^a\big(\mathbf{X}_{\mathcal{M}} | \mathbf{z}(\mathbf{X}_{\mathcal{K}})\big)\Big] \Big[\boldsymbol{a}(\mathbf{X}_{\mathcal{M}}) - \mathbf{m}^a\big(\mathbf{X}_{\mathcal{M}} | \mathbf{z}(\mathbf{X}_{\mathcal{K}})\big)\Big]^T \right\} \\ &= \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) - \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{K}}) \Big[\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) + \mathbf{C}^v(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})\Big]^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{K}})^T, \end{aligned} \tag{4.29}$$

where $\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{K}})$ and $\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})$ are blocks of the partitioned covariance matrix

$$\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) = \left[\begin{array}{cc} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{K}}) & \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{N}}) \end{array} \right] = \left[\begin{array}{cc} \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) & \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \\ \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})^T & \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) \end{array} \right]. \tag{4.30}$$

The estimated SNR, *i.e.* $\widehat{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{N}}) | \mathbf{X}_{\mathcal{K}}, \mathbf{z}(\mathbf{X}_{\mathcal{K}}))$, is similar to (4.7), but, since $\mathbf{X}_{\mathcal{K}}$ is fixed, it is now conditioned on $\mathbf{z}(\mathbf{X}_{\mathcal{K}})$ and also it is a function of $\mathbf{X}_{\mathcal{N}}$ only. As $\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})$ is Gaussian, the distribution of $\boldsymbol{w}(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}}) \triangleq (1/\sigma_s^2) \widehat{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{N}}) | \mathbf{X}_{\mathcal{K}}, \mathbf{z}(\mathbf{X}_{\mathcal{K}}))$ can be obtained in a similar way as presented in Subsection 4.2, equation (2). With the distribution of $\boldsymbol{w}(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}})$, we can modify the criterion (4.22) to have the following robust sequential sensor placement criterion :

$$J_P(\mathbf{X}_{\mathcal{N}}, \theta | \mathbf{z}(\mathbf{X}_{\mathcal{K}})) = \Pr(\boldsymbol{w}(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}}) > \theta) = 1 - G_{\boldsymbol{w}(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}})}(\theta). \tag{4.31}$$

In the next section, by assuming a perfect knowledge on the spatial gain, we will show how the distribution of $\boldsymbol{w}(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}})$ can be derived.

4.3.3 Perfect gain information

In this section by assuming a perfect knowledge on the spatial gain, we derive the distribution of $\boldsymbol{w}(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}})$. If the information on the gain is assumed to be perfectly known at sensor locations $\mathbf{X}_{\mathcal{K}}$, that is

$$\mathbf{z}(\mathbf{X}_{\mathcal{K}}) = \mathbf{a}(\mathbf{X}_{\mathcal{K}}),$$

then using (4.9) we have

$$\begin{aligned}
&\frac{1}{\sigma_s^2}\widehat{\mathrm{SNR}}\big(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{N}})|\mathbf{X}_{\mathcal{K}}, \mathbf{z}(\mathbf{X}_{\mathcal{K}}) = \mathbf{a}(\mathbf{X}_{\mathcal{K}})\big) \qquad (4.32)\\
&= \begin{bmatrix} \mathbf{a}(\mathbf{X}_{\mathcal{K}}) \\ \boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) \end{bmatrix}^T \begin{bmatrix} \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) & \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \\ \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})^T & \mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) \end{bmatrix} \begin{bmatrix} \mathbf{a}(\mathbf{X}_{\mathcal{K}}) \\ \boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) \end{bmatrix}\\
&= \mathbf{a}(\mathbf{X}_{\mathcal{K}})^T\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})\mathbf{a}(\mathbf{X}_{\mathcal{K}}) + 2\mathbf{a}(\mathbf{X}_{\mathcal{K}})^T\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})\boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) + \boldsymbol{a}(\mathbf{X}_{\mathcal{N}})^T\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}), \boldsymbol{a}(\mathbf{X}_{\mathcal{N}}).
\end{aligned}$$

where

$$\mathbf{R}(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) = \begin{bmatrix} \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) & \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}}) \\ \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})^T & \mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) \end{bmatrix}.$$

The conditioned random vector $\boldsymbol{a}(\mathbf{X}_{\mathcal{N}})|\mathbf{z}(\mathbf{X}_{\mathcal{N}})$ is also Gaussian whose mean $\mathbf{m}^a(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ and covariance $\mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ can be obtained, respectively, from (4.28) and (4.29) by setting the noise covariance equal to zero, *i.e.* $\mathbf{C}^v(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) = \mathbf{0}$. They are given by

$$\begin{aligned}
\mathbf{m}^a(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}) &= \mathbb{E}\Big\{\boldsymbol{a}(\mathbf{X}_{\mathcal{N}})|\mathbf{a}(\mathbf{X}_{\mathcal{K}}) = \mathbf{z}(\mathbf{X}_{\mathcal{K}})\Big\} \qquad (4.33)\\
&= \mathbf{m}^a(\mathbf{X}_{\mathcal{N}}) + \mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})^T\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})^{-1}\big[\mathbf{a}(\mathbf{X}_{\mathcal{K}}) - \mathbf{m}(\mathbf{X}_{\mathcal{K}})\big]. \qquad (4.34)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}) =& \mathbb{E}\Big\{\big[\boldsymbol{a}(\mathbf{X}_{\mathcal{N}})\mathbf{m}^a(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})\big]\big[\boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) - \mathbf{m}^a(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})\big]^T|\mathbf{a}(\mathbf{X}_{\mathcal{K}}) = \mathbf{z}(\mathbf{X}_{\mathcal{K}})\Big\}\\
=& \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}}) - \mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}})\mathbf{C}^a(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}})^{-1}\mathbf{C}^a(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}})^T. \qquad (4.35)
\end{aligned}$$

Since by definition (4.8), $\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})$ is invertible, then, by factorizing (4.33), we have

$$\begin{aligned}
&\widehat{\mathrm{SNR}}\Big(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{N}})|\mathbf{X}_{\mathcal{K}}, \mathbf{a}(\mathbf{X}_{\mathcal{K}}) = \mathbf{z}(\mathbf{X}_{\mathcal{K}})\Big) = \sigma_s^2\times\\
&\quad\Bigg\{\Big[\mathbf{a}(\mathbf{X}_{\mathcal{K}})^T\Big(\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) - \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})^T\Big)\mathbf{a}(\mathbf{X}_{\mathcal{K}})\Big]\\
&\qquad + \Big[\Big(\boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) + \mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}})\mathbf{a}(\mathbf{X}_{\mathcal{K}})\Big)^T\times\\
&\qquad\quad \mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})\Big(\boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) + \mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{K}})\mathbf{a}(\mathbf{X}_{\mathcal{K}})\Big)\Big]\Bigg\}. \qquad (4.36)
\end{aligned}$$

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

$$\begin{aligned}
w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}) \triangleq& \frac{1}{\sigma_s^2}\widehat{\mathrm{SNR}}\left(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{N}})|\mathbf{X}_{\mathcal{K}}, \mathbf{a}(\mathbf{X}_{\mathcal{K}}) = \mathbf{z}(\mathbf{X}_{\mathcal{K}})\right) \qquad (4.37)\\
&- \mathbf{a}(\mathbf{X}_{\mathcal{K}})^T\left(\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) - \mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})\mathbf{R}(\mathbf{X}_{\mathcal{N}}, \mathbf{X}_{\mathcal{N}})^{-1}\mathbf{R}(\mathbf{X}_{\mathcal{K}}, \mathbf{X}_{\mathcal{N}})^T\right)\mathbf{a}(\mathbf{X}_{\mathcal{K}})
\end{aligned}$$

has a distribution that can be obtained as described in Subsection 4.2. So, we have the following proposition.

Proposition 3 (Distribution of $w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$)
If the spatial gain $a(\boldsymbol{x})$ follows the GP model $a(\boldsymbol{x}) \sim \mathcal{GP}\Big(m^a(\boldsymbol{x}), C^a(\boldsymbol{x}, \boldsymbol{x}')\Big)$, then, we can show that $w(\boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}})$ is a weighted sum of N independent random variables as follows[2] *:*

$$w(\boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}}) = \sum_{i=1}^{N} d_i y_i^2, \tag{4.38}$$

where y_i's are independent normally distributed random variables $y_i \sim \mathcal{N}\left(m_{y_i}, 1\right)$, with the mean

$$m_{y_i} = \boldsymbol{u}_i^T \boldsymbol{C}^a(\boldsymbol{X}_{\mathcal{N}}, \boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}})^{-\frac{1}{2}} \boldsymbol{m}^a(\boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}}) \tag{4.39}$$

and, d_i's and $\boldsymbol{u}_i$'s are, respectively, the eigenvalues and eigenvectors of the matrix $\boldsymbol{A}$ defined as follows :

$$\boldsymbol{A} \triangleq \boldsymbol{C}^a(\boldsymbol{X}_{\mathcal{N}}, \boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}})^{\frac{1}{2}} \boldsymbol{R}(\boldsymbol{X}_{\mathcal{N}}, \boldsymbol{X}_{\mathcal{N}}) \boldsymbol{C}^a(\boldsymbol{X}_{\mathcal{N}}, \boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}})^{\frac{1}{2}}. \tag{4.40}$$

The conditional mean of the spatial gain $\boldsymbol{m}^a(\boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}})$, and its conditional covariance $\boldsymbol{C}^a(\boldsymbol{X}_{\mathcal{N}}, \boldsymbol{X}_{\mathcal{N}}|\boldsymbol{X}_{\mathcal{K}})$ are, respectively, presented in (4.34) *and* (4.35)*.*

Since y_i is a normally distributed scalar with non-zero mean m_{y_i} (4.39) *and variance one, its squared form y_i^2 follows a non-central 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 $v_i \triangleq y_i^2$ will be similar to* (4.19) *with the difference that here u_i and d_i are, respectively, the eigenvectors and eigenvalues of* (4.40)*, and $\lambda_i = m_{y_i}^2$ where m_{y_i} is the conditional mean presented in* (4.39)*. So, by defining $\Gamma_i = d_i v_i$, the distribution of Γ_i will be as in* (4.20)*. On the other hand, similar to Lemma 1 we can show that for any $i \neq j \in \{1, 2, \ldots, N\}$, $\Gamma_i \triangleq d_i y_i^2$ and $\Gamma_j \triangleq d_j y_j^2$ are independent. Therefore, due to independence, the pdf of the SNR is given by the convolution product, denoted by $*$, between the pdf of N random variables Γ_i :*

$$\begin{aligned} g_w(w) &= g_{\Gamma_1}(w) * g_{\Gamma_2}(w) * \cdots * g_{\Gamma_N}(w) \\ &= \frac{1}{\prod_{i=1}^{N} d_i} g_{v_1}(\frac{w}{d_1}; k_1, \lambda_1) * g_{v_2}(\frac{w}{d_2}; k_2, \lambda_2) * \cdots * g_{v_N}(\frac{w}{d_N}; k_N, \lambda_N). \end{aligned} \tag{4.41}$$

Now, by knowing the pdf of $w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}) = \mathrm{SNR}(\boldsymbol{f}^*(\mathbf{X}_{\mathcal{N}})|\mathbf{X}_{\mathcal{K}})/\sigma_s^2$, it is possible to redefine our proposed criteria :

$$\hat{\mathbf{X}}_{\mathcal{N}} = \underset{\mathbf{X}_{\mathcal{N}}}{\operatorname{argmax}}\, J_P(\mathbf{X}_{\mathcal{N}}, \theta|\mathbf{X}_{\mathcal{K}}), \tag{4.42}$$

where the criterion $J_P(\mathbf{X}_{\mathcal{N}}, \theta|\mathbf{X}_{\mathcal{K}})$ is based on the cumulative distribution function (cdf) of $w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ (and equivalently the cdf of $\mathrm{SNR}(\boldsymbol{f}^*(\mathbf{X}_{\mathcal{N}})|\mathbf{X}_{\mathcal{K}})/\sigma_s^2$) :

$$J_P(\mathbf{X}_{\mathcal{N}}, \theta|\mathbf{X}_{\mathcal{K}}) = Pr(w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}}) > \theta) = 1 - G_{w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})}(\theta), \tag{4.43}$$

in which $G_w(\cdot)$ represents the cdf of $w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$. By knowing the pdf of $w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$, from (4.41) the cdf becomes as follows :

$$G_{w(\mathbf{X}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})}(\theta; \mathbf{k}, \Lambda) = \frac{1}{\prod_{i=1}^{N} d_i} \times G_{v_1}\Big(\frac{\theta}{d_1}; k_1, \lambda_1\Big) * g_{v_2}\Big(\frac{\theta}{d_2}; k_2, \lambda_2\Big) * \cdots * g_{v_N}\Big(\frac{\theta}{d_N}; k_N, \lambda_N\Big), \tag{4.44}$$

2. The proof is similar to the proof of proposition 1.

where $\mathbf{k} = \{k_1, k_2, \ldots, k_N\}$, $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$, and $G_{\nu_1}(\cdot)$ is the corresponding cdf of $g_{\nu_1}(\cdot)$ (4.24).

So, by having the cdf of $w(\mathbf{X}_\mathcal{N}|\mathbf{X}_\mathcal{K})$, the criterion (4.43) can be used in Algorithm 2 to sequentially add new sensors.

4.3.4 Adding one sensor at a time under the perfect gain information setting

A special case for which $J_P\big(\mathbf{X}_\mathcal{N}, \theta|\mathbf{z}(\mathbf{X}_\mathcal{K}) = \mathbf{a}(\mathbf{X}_\mathcal{K})\big)$ has a known analytical form is when we add new sensors one by one, *i.e.* $N = 1$. In this case, $a(\mathbf{x}_\mathcal{N})$ is a scalar random variable and both $R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})$ and $C^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})$ are also scalars. Therefore, we can rewrite (4.36) as follows

$$\widehat{\mathrm{SNR}}\left(\hat{\boldsymbol{f}}(\mathbf{x}_\mathcal{N})|\mathbf{X}_\mathcal{K}, \mathbf{a}(\mathbf{X}_\mathcal{K})\right) = \sigma_s^2 \times \Bigg[R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N}) \left(a(\mathbf{x}_\mathcal{N}) + \frac{\mathbf{R}(\mathbf{x}_\mathcal{N}, \mathbf{X}_\mathcal{K})\mathbf{a}(\mathbf{X}_\mathcal{K})}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})}\right)^2 + \mathbf{a}(\mathbf{X}_\mathcal{K})^T \left(\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K}) - \frac{\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{x}_\mathcal{N})\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{x}_\mathcal{N})^T}{R_{NN}}\right)\mathbf{a}(\mathbf{X}_\mathcal{K})\Bigg].$$

If we define

$$q(\mathbf{x}_\mathcal{N}) \triangleq \frac{1}{\sqrt{C^a}(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})}\left(a(\mathbf{x}_\mathcal{N}) + \frac{\mathbf{R}(\mathbf{x}_\mathcal{N}, \mathbf{X}_\mathcal{K})\mathbf{a}(\mathbf{X}_\mathcal{K})}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})}\right),$$

then

$$\widehat{\mathrm{SNR}}(\hat{\boldsymbol{f}}(\mathbf{x}_\mathcal{N})|\mathbf{X}_\mathcal{K}, \mathbf{a}(\mathbf{X}_\mathcal{K})) = \sigma_s^2 \times \Bigg[R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})C^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})q(\mathbf{x}_\mathcal{N})^2 + \mathbf{a}(\mathbf{X}_\mathcal{K})^T \left(\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K}) - \frac{\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{x}_\mathcal{N})\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{x}_\mathcal{N})^T}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})}\right)\mathbf{a}(\mathbf{X}_\mathcal{K})\Bigg]. \tag{4.45}$$

From the Gaussian assumption, we have $q(\mathbf{x}_\mathcal{N}) \sim \mathcal{N}\left(m^q(\mathbf{x}_\mathcal{N}), \sigma_q(\mathbf{x}_\mathcal{N})^2\right)$, where $\sigma_q(\mathbf{x}_\mathcal{N})^2 = 1$, and

$$m^q(\mathbf{x}_\mathcal{N}) = \frac{1}{\sqrt{C^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})}}\left(m^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K}) + \frac{\mathbf{R}(\mathbf{x}_\mathcal{N}, \mathbf{X}_\mathcal{K})\mathbf{a}(\mathbf{X}_\mathcal{K})}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})}\right).$$

Therefore, it can be concluded that $q(\mathbf{x}_\mathcal{N})^2$ has a non-central chi-squared distribution as $q(\mathbf{x}_\mathcal{N})^2 \sim \chi^2(1, \lambda)$, with degree of freedom $k = 1$ and non-centrality parameter

$$\lambda = m^q(\mathbf{x}_\mathcal{N}) = \frac{1}{\sqrt{C^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})}}\left(m^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K}) + \frac{\mathbf{R}(\mathbf{x}_\mathcal{N}, \mathbf{X}_\mathcal{K})\mathbf{a}(\mathbf{X}_\mathcal{K})}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})}\right).$$

Using expression (4.45) and (4.24), we can give an expression for the robust placement criterion

$$J_P(\mathbf{X}_\mathcal{N}, \theta|\mathbf{a}(\mathbf{X}_\mathcal{K})) = \Pr(w(\mathbf{X}_\mathcal{N}|\mathbf{X}_\mathcal{K}) > \theta) = 1 - Q_{\frac{1}{2}}(\sqrt{\lambda}, \sqrt{\theta'}), \tag{4.46}$$

where

$$\theta' = \frac{\theta - \mathbf{a}(\mathbf{X}_\mathcal{K})^T \left(\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{X}_\mathcal{K}) - \frac{\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{x}_\mathcal{N})\mathbf{R}(\mathbf{X}_\mathcal{K}, \mathbf{x}_\mathcal{N})^T}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})}\right)\mathbf{a}(\mathbf{X}_\mathcal{K})}{R(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})C^a(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})}. \tag{4.47}$$

So, under the assumption of having a perfect knowledge on the spatial gain, the criterion (4.46) is used in Algorithm 2 to sequentially add new sensors.

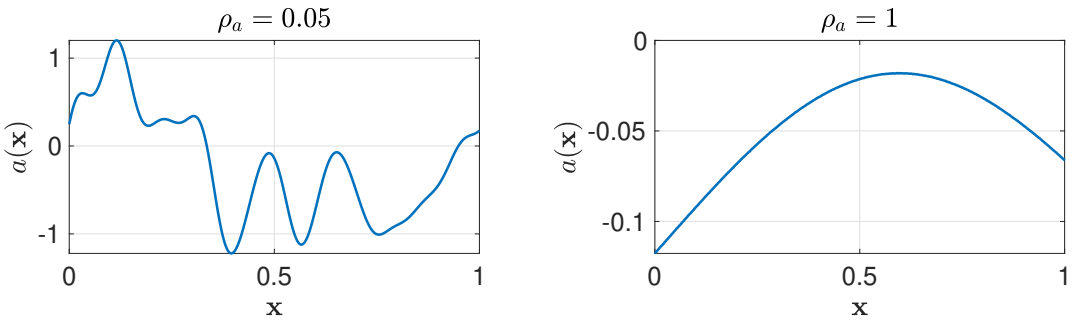


FIGURE 4.2: The effect of the parameter ρ_a on the smoothness of the spatial gain. By increasing ρ_a, the spatial gain becomes smoother.

4.4 Numerical experiments

Experiments concerning the robustness of the proposed criterion J_P (4.22) and of the criteria from Chapter 3 based the mean of the SNR J_E (3.14c), the entropy J_H (2.52) and the mutual information J_{MI} (2.56) are presented in this section. All the simulations 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. We first discuss about the concept of robustness and how to measure it. Next, the effect of θ in J_P is shown. Afterwards, we compare the different criteria in terms of robustness and output SNR. Finally, we will introduce more numerical tools to evaluate the performance of the criteria from different aspects.

4.4.1 Simulation setup

In the numerical experiments, we consider a one-dimension grid $\mathbf{X}_{\mathcal{P}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_P]^T$ in the normalized range $\mathbf{x}_i \in [0, 1]$ where $i = 1, 2, \ldots, P$ for possible sensor locations. Note that to keep the consistency throughout this chapter, in the experimental part we represent the scalars with bold lowercase letters. Depending on the smoothness of the signals, the size P of the grid and the number of initial sensors K, as well as their positions $\mathbf{X}_{\mathcal{K}}$ will be changed in different simulation parts. In all the cases, we assume that

$$\mathbf{m}^a(\mathbf{X}_{\mathcal{K}}) = \mathbf{a}(\mathbf{X}_{\mathcal{K}}) + \mathbf{b}(\mathbf{X}_{\mathcal{K}}) + \bar{\boldsymbol{u}}(\mathbf{X}_{\mathcal{K}}), \tag{4.48}$$

where $\mathbf{m}^a(\mathbf{X}_{\mathcal{K}})$ is the mean of the spatial gain which will be used as the estimation of the spatial gain, $\mathbf{a}(\mathbf{X}_{\mathcal{K}})$ is the true value of the spatial gain, $\mathbf{b}(\mathbf{X}_{\mathcal{K}})$ is a bias in the spatial gain, and $\bar{\boldsymbol{u}}(\mathbf{X}_{\mathcal{K}})$ is the uncertainty of the spatial gain. Note that, the sum of the uncertainty and the bias denotes the error $\boldsymbol{v}(\mathbf{X}_{\mathcal{K}})$ in equation (4.27), *i.e.* $\boldsymbol{v}(\mathbf{X}_{\mathcal{K}}) = \mathbf{b}(\mathbf{X}_{\mathcal{K}}) + \bar{\boldsymbol{u}}(\mathbf{X}_{\mathcal{K}})$. We use GP models $\mathcal{GP}\big(m(\mathbf{x}), C(\mathbf{x}, \mathbf{x}')\big)$, 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 $\boldsymbol{a}(\mathbf{x})$, the noise $\boldsymbol{n}(\mathbf{x})$, and the uncertainty $\bar{u}(\mathbf{x})$. Here, ρ is a smoothness parameter where a small ρ compared to 1 means fast spatial changes, while a large $\rho \simeq 1$ means smooth changes. The effect of ρ_a on the smoothness of the spatial gain is depicted in Fig. 4.2 for two different values : $\rho_a = 0.05$ on the left, and $\rho_a = 1$ on the right. 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. This is due to the fact that the scaling factor of the spatial gain is different at each position, and so the aspect ratio between the spatial gain at each pair of the positions will be changed. The subscripts $(.)_a$, $(.)_n$, $(.)_b$, and $(.)_u$ 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 prior mean functions to be equal to 0 except the mean function of the spatial gain which is generated according to (4.48). 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 $J_P\left(\mathbf{x}_{\mathcal{N}}, \theta | \hat{\mathbf{a}}(\mathbf{X}_{\mathcal{K}})\right)$, it is required to choose a relevant value for the parameter θ. So, we suggest the following model :

$$\theta = \widehat{\mathrm{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}})) + \delta \left[\widehat{\mathrm{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})) - \widehat{\mathrm{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}}))\right]. \tag{4.49}$$

In the above equation, by using (4.33), $\widehat{\mathrm{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}}))$ and $\widehat{\mathrm{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ are the means 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 factor of the increase in the SNR after adding new sensors. Specifically, by setting $\delta = 0$, we set the parameter θ equal to the initial value of the SNR before adding the new sensor. On the other hand, if we set $\delta = 1$, the parameter θ becomes equal to the average SNR after adding the new sensors. In the rest of the simulation part, we parameterize J_P as a function of δ instead of θ.

4.4.2 Failure region and Failure[%] as measures for the uncertainty effect

Although J_E and J_P are criteria targeting SNR, uncertainty on $\boldsymbol{a}(\mathbf{x})$ may lead to large values of these criteria at positions where in practice the value of the true output $\mathrm{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})|\mathbf{X}_{\mathcal{M}})$ may be decreased. The true SNR is given by (4.3), where $\mathbf{f}(\mathbf{X}_{\mathcal{M}}) = \hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})$ is the estimated extraction vector. This vector is given by $\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1}\mathbf{m}^a(\mathbf{X}_{\mathcal{M}})$ in the non-sequential approach and by $\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}) = \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1}\mathbf{m}^a(\mathbf{X}_{\mathcal{M}}|\mathbf{X}_{\mathcal{K}})$ in the sequential approach. As a consequence, the true output SNR can be rewritten as

$$\mathrm{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})|\mathbf{X}_{\mathcal{M}}) = \frac{\sigma_s^2 \hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})^T \mathbf{a}(\mathbf{X}_{\mathcal{M}})\mathbf{a}(\mathbf{X}_{\mathcal{M}})^T \hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})}{\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})}. \tag{4.50}$$

Note that (4.50) differs from the estimated SNR, *i.e.* $\widehat{\mathrm{SNR}}$ (4.7), since this true SNR (4.50) depends on the true spatial gain $\mathbf{a}(\mathbf{X}_{\mathcal{M}})$, contrary to $\widehat{\mathrm{SNR}}$ which only depends on the estimation of the spatial gain $\boldsymbol{a}(\mathbf{X}_{\mathcal{M}})$. In the rest, the set of positions that deteriorates the true SNR is called *failure region* (FR).

Fig. 4.3 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 $\mathbf{X}_{\mathcal{K}} = [0.25, 0.5, 0.75]^T$,

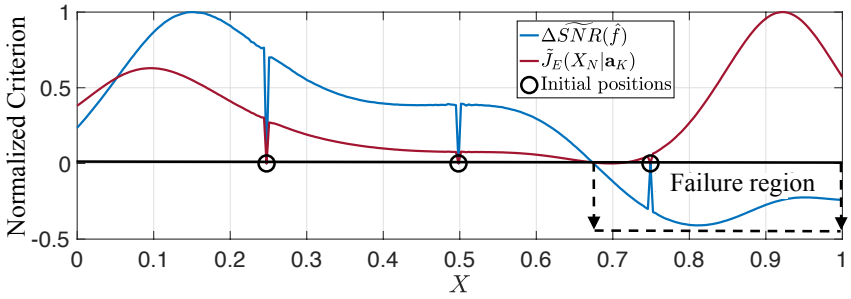


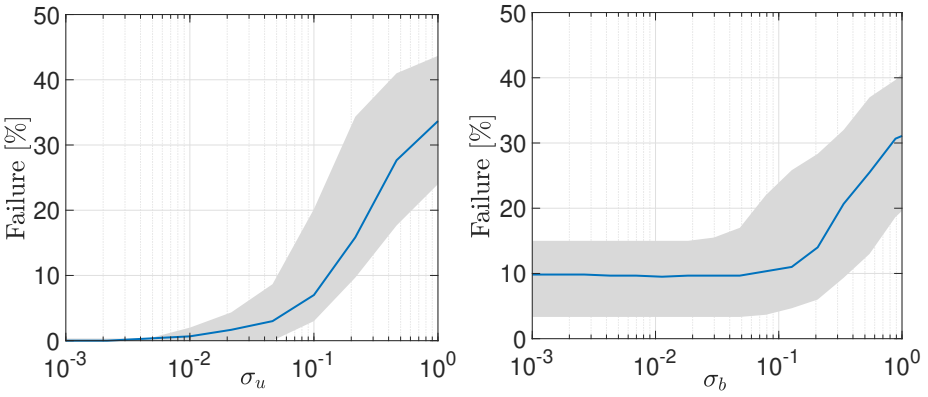
FIGURE 4.3: Failure region is where the true value of the SNR with estimated extraction vector $\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})$ is smaller than its initial value before adding the new sensors, *i.e.* when $\Delta\widetilde{\text{SNR}}$ (the blue curve) defined in (4.51) is negative. Here, 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 $\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})$. The tilde superscript represents the normalization of the criteria.

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 $\sigma_s = 2$. To illustrate the failure phenomenon, we use J_E (3.14) as the placement criterion. Now, the aim is to find the best location for the 4^{th} sensor. In this figure, the tilde superscript indicates that the function is normalized such that its maximum value is equal to one, and the initial value is equal to zero. The normalized variation of SNR, denoted $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$, is as follows :

$$\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})) = \frac{\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})) - \text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}}))}{\text{SNR}_{max}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})) - \text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}}))}, \tag{4.51}$$

where $\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}}))$ represents the initial value of the SNR before adding the new sensor, and $\text{SNR}_{max}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ is the maximum value of the SNR after adding the new sensor. Note that in this chapter, 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. 4.3 it is seen that, due to the uncertainty, the normalized variation of the true SNR, *i.e.* $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$, can take negative values at some regions, which means that by placing sensors at these positions, the SNR will take smaller values compared to the initial SNR. Therefore, since $\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})) < \text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{K}}))$ in FR, the numerator of (4.51) becomes negative, and as such, $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ 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. Therefore, this is an example of the lack of robustness of the criterion J_E.

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 sensors positions within the FR, here denoted N_{FR}. For a total size



(a) The effect of uncertainty (b) The effect of bias

FIGURE 4.4: 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 $\sigma_u = 0.01$.

of spatial grid N_P, we define the failure rate as :

$$\text{Failure}\,[\%] \triangleq \frac{\text{Size of the FR}}{\text{Total size of the spatial grid}} = \frac{N_{FR}}{N_P}. \tag{4.52}$$

Fig. 4.4 illustrates the effect of uncertainty and bias on Failure[%]. In this figure, three initial sensors are placed at $\mathbf{X}_{\mathcal{K}} = \{0.25, 0.5, 0.75\}$, and we place the 4^{th} sensor using J_E. The smoothness parameters are set as $\rho_a = \rho_u = 0.2$, $\rho_n = 0.01$, and the variances are set $\sigma_a = 0.5$ and $\sigma_n = 0.01$. We generate $N^a_{MC} = 10$ realizations for the spatial gain, and for each realization, we consider 50 runs for the bias, uncertainty and noise ($N^b_{MC} = N^u_{MC} = N^n_{MC} = 50$), leading to a total number of $N_{MC} = 500$ Monte-Carlo (MC) realizations. Then, Failure[%] is evaluated according to (4.52). In this figure, the effects of the bias and the uncertainty on Failure[%] are studied separately. Firstly, in Fig. 4.4a, 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[%] increase. This experiment is repeated in Fig. 4.4b 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 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 implies Failure[%] of about 10 % which remains almost constant at this amount for $\sigma_b \leq 0.1$. If the level of bias goes beyond this value, the effect of the bias on Failure[%] becomes more significant, which means that the suggested model for the spatial gain is not appropriate, and better methods need to be used in order to provide the best 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 $b(x) = 0$.

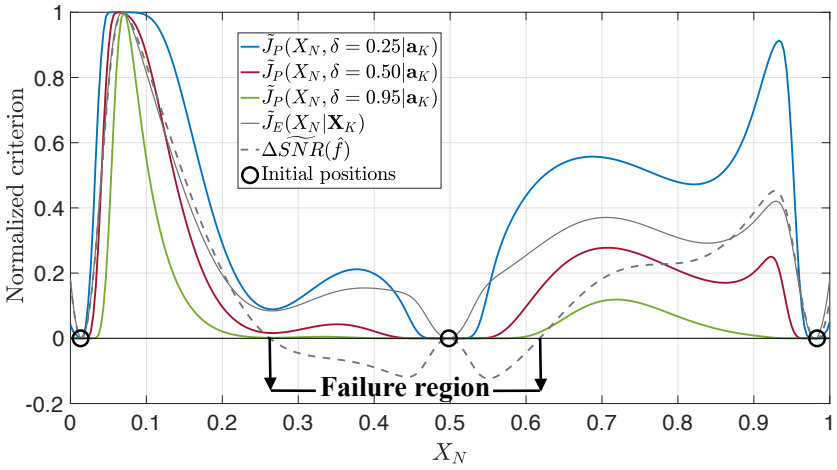


FIGURE 4.5: Effect of parameter δ. The solid curves are the proposed criterion for different δ. The superscript tilde represents the normalization of the function. By increasing δ, the algorithm is more robust against FR (see the text below for more details). However, other candidates for sensor placement 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. 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 being equal to 300. Starting with three initial sensors at $\mathbf{X}_\mathcal{K} = \{0.05, 0.5, 0.95\}$, we look for the best position for the 4^{th} sensor.

4.4.3 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. 4.5 is provided where the proposed criterion J_P (4.43) is used as the target function with three different values of δ (0.25, 0.5, and 0.95), which are depicted with blue, red, and green color, respectively. Moreover, $\widetilde{\Delta\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))$ which is the true output SNR normalized according to (4.51) is depicted with a dashed curve. The plots for $J_E(\mathbf{x}_\mathcal{N}|\mathbf{X}_\mathcal{K})$ and $J_P(\mathbf{x}_\mathcal{N}, \delta|\mathbf{X}_\mathcal{K})$ are normalized such that their maximum and initial values are equal to 1 and 0, respectively. The normalized forms are denoted by a superscript tilde. Note that the locations of maxima for different criteria will not be changed by the normalization.

As depicted in Fig. 4.5, 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 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 FR and keeping the regions with a high increase in the SNR.

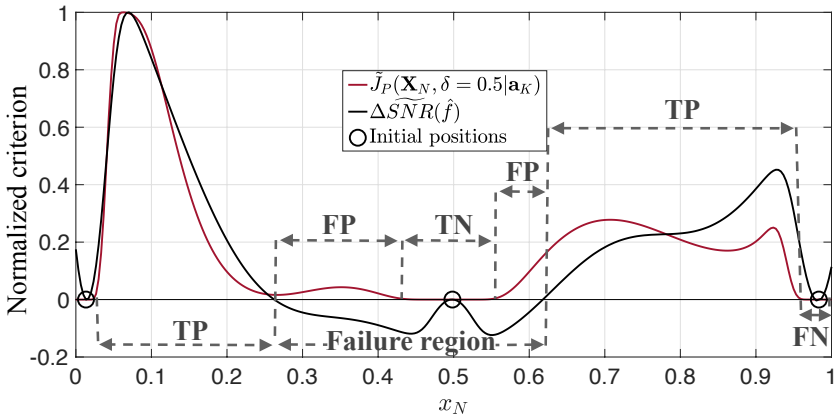
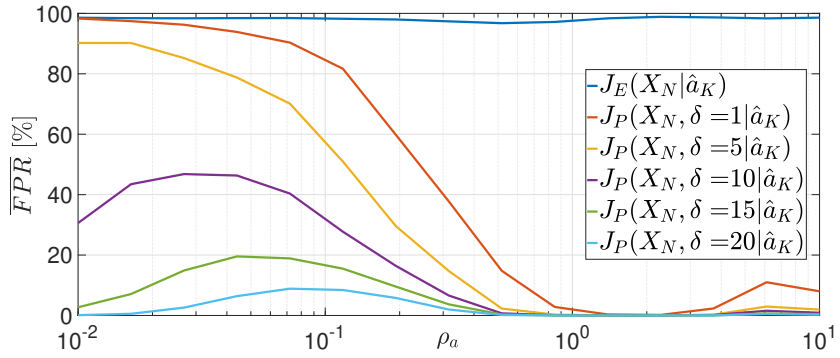


FIGURE 4.6: Region classification : Positive (P) vs. Negative (N), and True (T) vs. False (F). P and N represent the regions where $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ is positive and negative, respectively. T is a situation when the criterion and $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ are both positive, or the criterion is zero and $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ is negative. F is when the criterion is positive while $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ is negative, or the criterion is zero whenever $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ is positive. Here, the parameter configuration is 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 $\mathbf{X}_{\mathcal{K}} = \{0.05, 0.5, 0.95\}$.

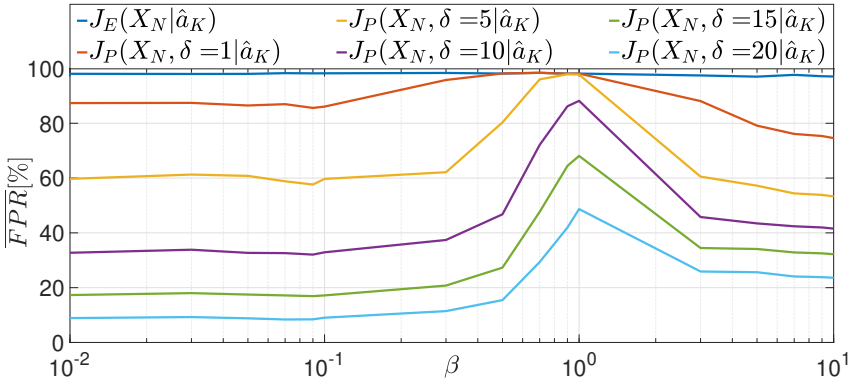
Finally, we note that by increasing δ, the proposed criterion J_P becomes flatter with a single narrow peak around its maximum. Moreover, as δ increases, the decrease of the criterion is not uniform with respect to $\mathbf{x}$. For instance, considering the green curve when the parameter δ is increased to 0.95. We see that by increasing δ the peaks at positions $\mathbf{x} = 0.93$ and $\mathbf{x} = 0.35$ become zero, meanwhile at position $\mathbf{x} = 0.7$ we still have a peak, which means that this position can be a possible candidate for sensor placement. We note that sometimes due to some limitations we cannot put the sensor at the maxima of the criteria and we look for the next best candidate to put the sensor. For instance, if the suggested position by the criterion is very close to the location of the previously located sensors, we need to look for the next best candidate to put the new sensor.

4.4.4 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. Based on Fig. 4.6, we call the regions with positive $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ as positive (P), and the regions with negative $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ as negative regions (N). Here, $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ denotes the normalized output SNR defined in (4.51). Accordingly, the true positive (TP) is the size of the region where both $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ and the criterion are positive. We recall that, by the size of a region we mean the number of grid points which are in the corresponding region. The number of positions that $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ 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



(a) The influence of **a** smoothness, with $\beta = 0.1$.



(b) The influence of smoothness ratio, with $\rho_a = 0.1$.

FIGURE 4.7: The effect of the smoothness of the spatial gain and the noise on the robustness. (a) Here, different smoothness values of the spatial gain are 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.

the region where $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))$ is negative, and the criterion is zero. Finally, false negative (FN) is related to the position with positive $\Delta\widetilde{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))$, 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 criterion as follows :

$$FPR = \frac{\textit{false positive}}{\textit{total number of negatives}} = \frac{FP}{TN + FP}, \tag{4.53}$$

where, the denominator is actually the size of the FR. It is noticeable that if the size of the FR is small, although the 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. 4.4, we set $\sigma_u \in [0.1, 0.8]$ which provides a moderate Failure[%] between 8% and 30%.

Fig. 4.7a is provided to show the robustness of the proposed criterion $J_P(\mathbf{x}_\mathcal{N}, \delta|\mathbf{a}(\mathbf{X}_\mathcal{K}))$

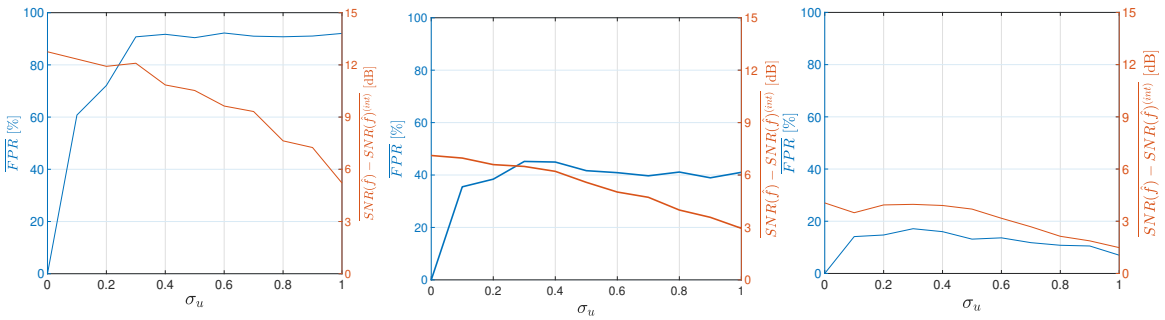
in terms of the average $FPR[\%]$ and compare it with the criterion $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ (3.17) for different smoothness conditions ρ_a. The larger the smoothness parameter ρ_a is, the smoother the spatial gain is. In this figure, three initial sensors are used at $\mathbf{X}_{\mathcal{K}} = \{0.05, 0.5, 0.95\}$ in a spatial grid of size 300, and we look for the 4^{th} sensor position. We note that in this figure, we intentionally selected the parameter δ between 0 and 1 to better visualise the results. In the later simulations, based on the set up of the other parameters, we also chose larger values of δ. 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 $\left[10^{-2},\ 10\right]$ 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 larger β corresponds to the smoother noise signal. A 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. 4.7. $\overline{FPR}[\%]$ represents the average value of $FPR[\%]$ over the whole Monte-Carlo realizations. As depicted in this figure, we can see that by increasing the parameter δ, the proposed criterion becomes more robust against the uncertainty and $\overline{\text{FPR}}[\%]$ becomes smaller. In contrast, $\overline{\text{FPR}}[\%]$ is almost constant and equal to 98% for different smoothness degrees using $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ (3.17) which shows its inability to eliminate the positions in FR.

Now, the effect of the noise smoothness is shown in Fig. 4.7b. 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 $\left[10^{-2},\ 10\right]$, thus leading to different values of ρ_n. It is seen that as the smoothness of the noise is varied, the proposed method shows a better performance compared to J_E. For instance, considering $\beta = 0.1$ and $\delta = 15$, the value of $\overline{\text{FPR}}[\%]$ is less than 20% for J_P, meanwhile, it is almost constant and equal to 98% for J_E. Generally, by increasing the parameter δ, the FR is better eliminated by the proposed method, and $\overline{\text{FPR}}[\%]$ becomes smaller. We also 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.

4.4.5 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 unable 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. 4.8



(a) $\delta = 1$ (b) $\delta = 10$ (c) $\delta = 20$

FIGURE 4.8: 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, $\overline{FPR}[\%]$ 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.

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 : $\sigma_u \in \{0, 0.1, 0.2, \ldots, 1\}$. We used 10 MC realizations for $\mathsf{a}(\mathbf{x})$ 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.* $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))} = \overline{\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M})) - \text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))^{(init)}}$ is presented in orange with the y-axis on the right. Here, $\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))^{(init)}$ denotes the true value of the SNR at the initial situation with a unique sensor located at $\mathbf{x} = 0.5$. Note that $\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M})) - \text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_\mathcal{M}))^{(init)}$ is the variation of the SNR, which is equivalent to the denominator of (4.51). This experiment is repeated for three different values of δ, namely 1, 10, and 20.

As Fig. 4.8 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. 4.8-a with Fig. 4.8-b, we see that by increasing δ from 1 to 10, the average FPR becomes smaller for different values of σ_u. However, the average improvement in the SNR also decreases. Going through a more strict selection for δ, in Fig. 4.8-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 have less increase in the SNR to reduce $\overline{FPR}[\%]$, and so, we can choose a larger δ. In contrast, when σ_u is large, since the output SNR is affected by this uncertainty, it is not desirable to lose a large amount of the SNR to have a smaller value of $\overline{FPR}[\%]$. 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 $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))}$ and $\overline{FPR}[\%]$ 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 the 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.

4.4.6 Sequential approach

In this part, we want to compare the proposed method $J_P(\mathbf{x}_{\mathcal{N}}, \delta|\mathbf{X}_{\mathcal{K}})$ with $J_H(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$, $J_{MI}(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ and $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ by using a sequential approach (Section 4.3.2). This comparison is made based on two performance indexes : the average output SNR, *i.e.* $\overline{\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))}$, and the average $\overline{FPR}[\%]$. 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 $\mathbf{x}_1 = 0.5$. 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, and to recover the source signal with a desired accuracy, we need to collect more information by using a 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 large 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 $\overline{FPR}[\%]$. In this part a number of $N^a_{MC} = 10$ realizations are used for the spatial gain $a(\mathbf{x})$, and for each realization we consider 10 runs for the bias, uncertainty and noise ($N^b_{MC} = N^u_{MC} = N^n_{MC} = 10$), 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 $n(\mathbf{x})$.

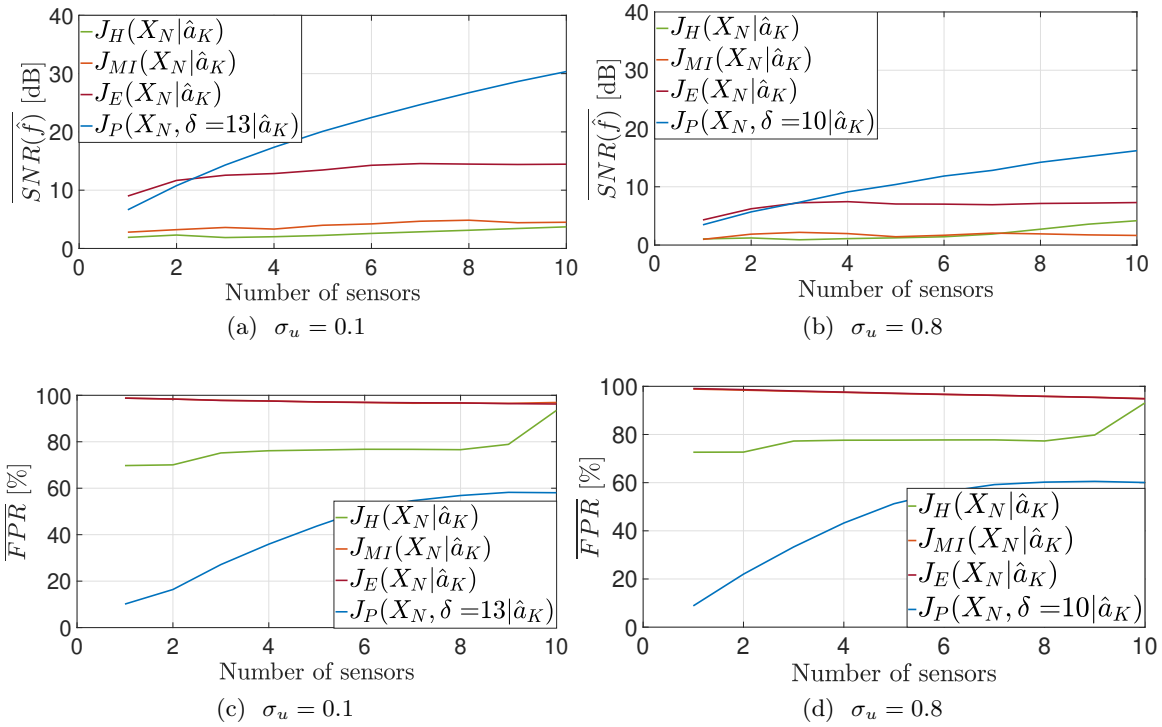
Fig. 4.9 presents the simulation results for different uncertainty levels. The first column is related to $\sigma_u = 0.1$, and the second column is for $\sigma_u = 0.8$. The first row shows $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))}$ by adding new sensors, and the second row corresponds to $\overline{FPR}[\%]$. In all the sub-figures, the blue, red, orange, and green curves show the performances of the proposed criterion $J_P(\mathbf{x}_{\mathcal{N}}, \delta|\hat{\mathbf{a}}_K)$, and other criteria based on the mean of the SNR $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$, the mutual

information $J_{MI}(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$, and the entropy $J_H(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$, respectively.

We begin with a low level of uncertainty $\sigma_u = 0.1$ in the first column. By looking at Fig. 4.9-a, it is seen that the proposed criterion significantly outperforms the other methods and provides a larger value for $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))}$ when the number of the sensors increases. By sequentially adding 10 sensors, $J_P(\mathbf{x}_{\mathcal{N}}, \delta|\hat{\mathbf{a}}_K)$ provides $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))} = 30\ dB$, which still can be increased by adding more sensors. The next best criterion is $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$, which gives $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))} = 15\ dB$. 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. 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 $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$. 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, the proposed criterion starts outperforming $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$. Therefore, in this case, to avoid increasing the computational complexity, one can simply start adding primary sensors using $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$. 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 Chapter 3, $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ has notably a better performance compared to the classical kriging approaches $J_H(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ and $J_{MI}(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$.

To complete the discussion, sub-figure (c) shows the results of $\overline{FPR}[\%]$. 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 $\overline{FPR}[\%]$. Note that the curves related to $J_{MI}(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ and $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ are quite close to each other. In sub-figure (c), $J_H(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ seems to perform better than $J_{MI}(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$ and $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{X}_{\mathcal{K}})$, and it provides approximately $\overline{FPR}[\%] = 70\%$ up to adding 9 sensors, with step-wise 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 $\overline{FPR}[\%]$ starts from 10%, and slightly increases by adding the number of sensors up to about 8 sensors. Then, $\overline{FPR}[\%]$ 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 $\overline{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ (concluded from sub-figure (a)), we can be hopeful that $\overline{FPR}[\%]$ 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 see that the proposed criterion has a superior performance compared to the other methods both in $\overline{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ and $\overline{FPR}[\%]$. Sub-figure (b) shows that by adding 10 sensors, the proposed criterion provides $\overline{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})) = 16\ dB$, which is 9 dB higher than the result achieved by $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$. In this part, by choosing $\delta = 10$, we try to keep the results for $\overline{FPR}[\%]$ 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,



(a) $\sigma_u = 0.1$ (b) $\sigma_u = 0.8$

(c) $\sigma_u = 0.1$ (d) $\sigma_u = 0.8$

FIGURE 4.9: 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. Note that in (c) and (d), J_{MI} and J_E largely overlap.

considering that 10 sensors are added, for the proposed criterion, $\overline{\text{SNR}}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ 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 $J_E(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ is from 15 dB to 8 dB, and for $J_{MI}(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ it is from 4 dB to 2 dB. The results for $J_H(\mathbf{x}_{\mathcal{N}}|\mathbf{a}(\mathbf{X}_{\mathcal{K}}))$ remain almost unchanged under different level of uncertainties.

4.5 Other numerical tools for performance evaluation

Previously, in Subsection 4.4.2, we discussed about FR as a notion to determine the regions which have negative SNR caused by the uncertainty, and then, we denoted the portion of the spatial grid within the FR by Failure[%]. Following that, in Subsection 4.4.4, we introduced FPR as a measure to represent the robustness of a criterion against uncertainty by using the concept of failure. We explained that one can report the robustness of a criterion against uncertainty by counting the number of the times that the criterion suggests to put a sensor in the FR (*i.e.*, the probability that the criterion takes its maximum value at FR).

Although FPR is a useful tool to measure the robustness of a criterion against uncertainty, our demand is not just to avoid failure region, but also we have to find the position which mostly increases the SNR, as described in subsection 4.4.5. Therefore, it is necessary to present more comprehensive measures to take into account different aspects of performance evaluations. In other words, we need a measure which reports the ability of a criterion to propose a position that provides an as large as possible increase in the SNR as well as less probability of being in the FR. In this part, such numerical tools will be provided to achieve this aim.

4.5.1 Thresholding instead of maximizing

To move forward to our goal in this part, we first explain about an important fact for sensor placement. Up to now, different criteria have been suggested to be maximized for sensor placement, and we used the measures $\overline{FPR}[\%]$ and $\overline{\Delta\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))}$ for performance evaluation which are based on the sensor positioning obtained according to the maxima of the criteria. However, choosing the maximum of a criterion may not be an efficient strategy to properly evaluate the performance of the criterion from the two aspects of our goal, *i.e.* having a large increase in the average SNR and low uncertainty. To explain more, we give two examples. In the first example, let's consider J_P as the criterion for sensor placement. The position that maximizes the criterion J_P may have the lowest uncertainty against FR, however, the amount of increase in the average SNR may not be significant. On the other hand, there may be other non-maximizer points with a little bit higher uncertainty compared to the maximum of the criterion, but these non-maximizer points can have a significant increase in the average SNR. It means that it is worth to accept this small increase in the uncertainty to have a large increase in the SNR. As the second example, consider J_E as the sensor placement criterion. Although the maximum of J_P has the largest increase in the average SNR, this point can contain too much uncertainty of being in the FR. So, there can be non-maximizer points with a little bit less increase in the average SNR, but a remarkably smaller uncertainty. According to this explanation, and the examples of J_P and J_E, we can conclude that the performances of different criteria cannot be properly evaluated just based on the maxima of the criteria. Therefore, we suggest that instead of considering the maxima of the criteria, we assume a threshold ξ for the minimum increase of the SNR that we expect to have, and then, among all the positions above this threshold, we select the ones with smaller uncertainties for sensor placement. In the following, we introduce a new region and numerical tools based on thresholding the criteria (instead of maximizing them) to better evaluate and compare the performances of different criteria.

4.5.2 Region of interest (ROI)

In this part, besides FR, we define a new region based on a threshold ξ which is a fraction of the maximum value of the SNR. In Fig. 4.10 an example of this threshold is demonstrated with a green dashed line for $\xi = 0.5$, which is 50% of the maximum SNR. According to this threshold, the positions where the true value of the SNR, *i.e.* $\text{SNR}(\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}}))$ takes its value

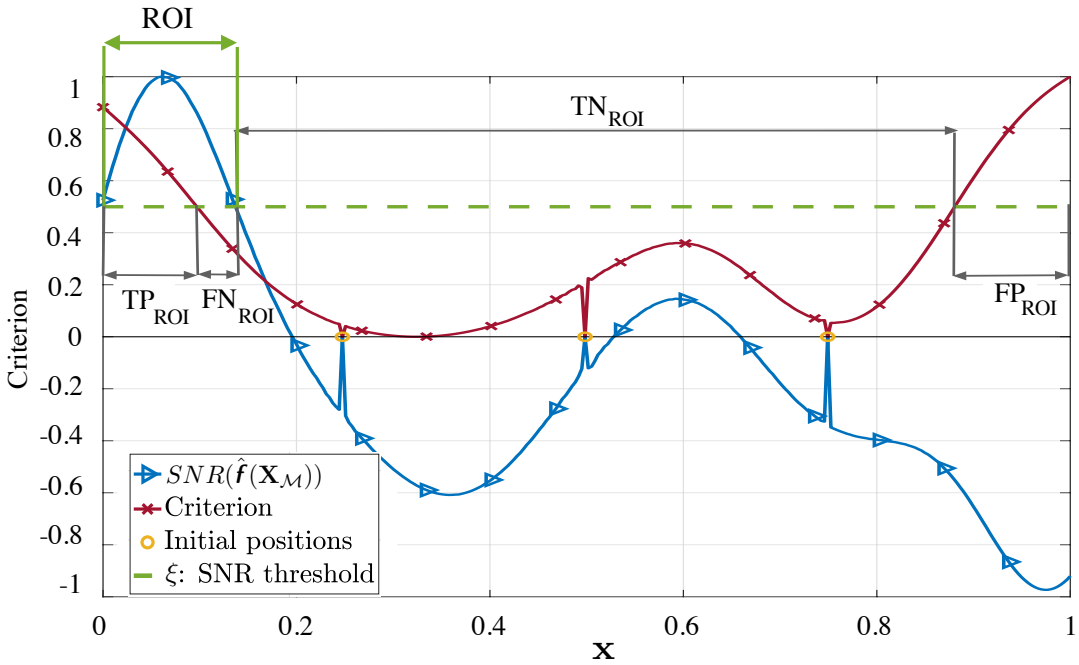


FIGURE 4.10: Region of interest (ROI) is the set of positions where the true value of the SNR, *i.e.* SNR($\hat{\boldsymbol{f}}(\mathbf{X}_{\mathcal{M}})$) takes its value above a threshold ξ. Based on ROI, four different regions are defined : I) TP_{ROI} : the regions where both the sensor placement criterion and the true SNR are in the ROI. II) TN_{ROI} : the regions where both the sensor placement criterion and the true SNR are out of the ROI. III) FP_{ROI} : the regions where the sensor placement criterion is in the ROI, but the true SNR is out of the ROI. IV) FN_{ROI} : the regions where the sensor placement criterion is out of the ROI, but the true SNR is in the ROI.

above ξ are considered as our desired positions, and we call the set of these positions as the *region of interest* (ROI). So, ξ can declare the minimum increase in the SNR that we expect to have by adding the new sensor. Based on the definition of ROI, we also define the following four segments

— True positive (TP_{ROI}) : the regions where both the sensor placement criterion and the true SNR are in the ROI.

— True negative (TN_{ROI}) : the regions where both the sensor placement criterion and the true SNR are out of the ROI.

— False positive (FP_{ROI}) : the regions where the sensor placement criterion is in the ROI, but the true SNR is out of the ROI.

— False negative (FN_{ROI}) : the regions where the sensor placement criterion is out of the ROI, but the true SNR is in the ROI.

These segments are demonstrated in Fig. 4.10, and from now on, by abbreviations TP_{ROI}, TN_{ROI}, FP_{ROI} and FN_{ROI}, we mention to the size of the corresponding segments. Next, we will explain how this new region classification can provide a better comprehension about the performance of a criterion from different aspects.

As mentioned before, the performance of a criterion should be evaluated in two aspects : being able to have a significant increase in the SNR, and the robustness against uncertainty (avoiding failure region). The advantage of defining ROI is that, instead of maximizing a criterion for sensor placement, we can define a region as the candidate to add the new sensor, and it helps us to achieve the two goals. For instance, if the maximum of the criterion J_E which maximizes the average SNR contains a high uncertainty, we can choose another position in ROI with a little bit less increase in the average SNR (depending on the value of ξ), but instead having a much smaller uncertainty.

With the threshold ξ, we can have a trade off between the amount of increase in the SNR, and reducing the possibility of choosing locations in the FR. To be clear, let us explain what happens if ξ is increased or decreased : a) by decreasing ξ, the size of the ROI will be larger, and among all possible positions in ROI, we select the position where the SNR has less uncertainty on the model. Therefore, although the selected position may have a smaller increase in the SNR compared to the maximum value of the SNR, it is more probable that the SNR takes this value, and so, less probable to be in the FR. b) by increasing ξ, the size of the ROI will be smaller. In this case, it is insisted that we want the SNR to increase significantly, and the positions with a small or moderate values of the SNR are ignored. Although in this situation we may have a large increase in the SNR, it is more probable that the proposed position is in the FR. It is due to the fact that, there may exist some positions out of the ROI with less uncertainty on the SNR, while the positions in the ROI may have a high degree of uncertainty.

To summarize, ξ is a tool to make a trade-off between increasing the SNR and decreasing the uncertainty of the SNR. The larger ξ is, the more risk we take to have a larger increase in the SNR. The smaller ξ is, the less probable it is to select the positions in FR. Depending on the application and the level of the risk we want to take regarding the uncertainty, as well as the amount of the increase in SNR that we expect, one can choose the threshold ξ.

In the following, we introduce three different performance indices relying on the threshold ξ and ROI to better evaluate the performance of a criterion.

4.5.3 Recall, Specificity, and their Harmonic mean

According to Fig. 4.10, it is important to well detect both the positive segments (regions where the true SNR is above the threshold ξ) and the negative segments (regions where the true SNR is bellow the threshold ξ). We can say that a criterion has a good performance if a large portion of the ROI can be recognized as positive, and also a large portion out of ROI is recognized as negative by the criterion. These two aspects can be separately measured by two performance indexes *Recall* and *Specificity* which are defined in the following.

I) **Recall :** If we consider a specific threshold for the SNR ξ, then, we say that a criterion has a good performance to detect the positive region when $\mathrm{TP}_{\mathrm{ROI}}$ is large relative to the total size of the positive region. Accordingly, we define *Recall* (Rcl) as the ratio between

$\mathrm{TP}_{\mathrm{ROI}}$ and the size of ROI which measures the percentage of the ROI detected by a criterion :

$$\mathrm{Rcl} = \frac{\text{True positives}}{\text{Number of actual positive}} = \frac{\mathrm{TP}_{\mathrm{ROI}}}{\mathrm{TP}_{\mathrm{ROI}} + \mathrm{FN}_{\mathrm{ROI}}}. \tag{4.54}$$

Note that if we denote the size of the ROI as |ROI| in the denominator, $\mathrm{TP}_{\mathrm{ROI}}+\mathrm{FN}_{\mathrm{ROI}} = |\mathrm{ROI}|$, which is the total size of the region where the true SNR is above ξ.

II) **Specificity :** If we consider a specific threshold for the SNR ξ, then, we say that a criterion has a good performance to detect the negative region when $\mathrm{TN}_{\mathrm{ROI}}$ is large relative to the total size of the negative region. Accordingly, we define *Specificity* (Spc) as the ratio between $\mathrm{TN}_{\mathrm{ROI}}$ and the size of the region out of ROI, which is the percentage of positions out of ROI detected by a criterion :

$$\mathrm{Spc} = \frac{\text{true negative}}{\text{no. of actual negative}} = \frac{\mathrm{TN}_{\mathrm{ROI}}}{\mathrm{TN}_{\mathrm{ROI}} + \mathrm{FP}_{\mathrm{ROI}}}. \tag{4.55}$$

Here, the denominator, $\mathrm{TN}_{\mathrm{ROI}}+\mathrm{FP}_{\mathrm{ROI}} = |\overline{ROI}|$ is the total size of the region out of the ROI which is denoted by $|\overline{ROI}|$.

III) **Harmonic mean, an average between Recall and Specificity :** Rcl and Spc do not necessarily behave in the same manner and it is possible to have a large value for Rcl but the Spc takes a quite smaller value. For instance, although an algorithm may be able to sufficiently detect positive positions with a large Rcl, it may fail to shrinkage the negative region and detect some positions in $\overline{ROI}$ as positive which causes Spc to be small. The opposite may also be the case, that is, a criterion has a good performance in ignoring $\overline{ROI}$ and has a large Spc, but it fails to properly detect ROI and has a small Rcl. Therefore, it is necessary to make an average between Rcl and Spc to take into account both aspects of our goal. For instance, *Harmonic mean* (Hrm) can be used which is defined as follows

$$\mathrm{Hrm} = 2 \times \frac{\mathrm{Rcl} \times \mathrm{Spc}}{\mathrm{Rcl} + \mathrm{Spc}}. \tag{4.56}$$

In the next part, we will present some numerical experiments to compare the performance of different criteria based on Rcl, Prc, and their Harmonic mean Hrm.

4.5.4 Performance evaluation

In this part, we study and compare the ability of our proposed criteria J_P and J_E to detect the positive and negative regions. To this end, Fig. 4.11 is provided where we have reported Rcl and Spc as well as their mean Hrm versus different threshold levels ξ. In this figure, the parameters are set the same as 4.4.3, and 100 Monte-Carlo realizations of $\mathsf{a}(\mathbf{x})$, $m^a(\mathbf{x})$, and $\mathsf{n}(\mathbf{x})$ are used to report the measures in average. Depending on the threshold level ξ, we set the parameter δ for J_P such that this criterion has its best performance in terms of the underlying measure. This means that, we try different values of δ, and among them, we select the one which has the largest Rcl, Spc, and Hrm to be reported.

In this figure, we see that J_P performs better than J_E both in terms of Rcl and Spc. However, this criterion shows a dual behaviour in terms of Hrm as an average between Rcl

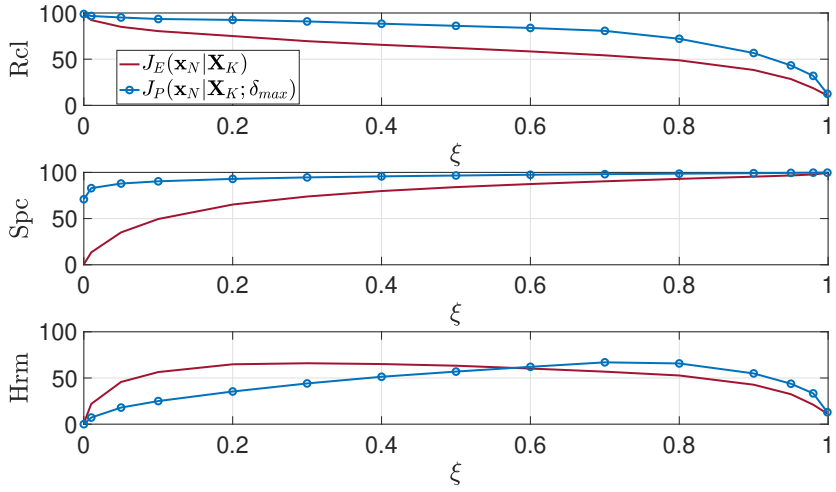


FIGURE 4.11: Performance in terms of Rcl, Spc, and Hrm.

and SPC. Using J_P, the measure Hrm takes larger values for higher thresholds ξ, but smaller values for lower thresholds. It is important to note that, if a specific value for δ maximizes Rcl, it does not necessarily maximizes Spc, and vice versa. This is why the Hrm, in average, does not show the same behavior as in Rcl and Spc in the first and second rows of Fig. 4.11.

From Fig. 4.11, we can conclude that since for the smaller values of the threshold ξ Hrm is larger in case of using J_P, it is possible to find a value for δ such that J_P has a better performance than J_E both in terms of Rcl and Spc. However, by increasing ξ and bringing the threshold level close to the maximum of the normalized true SNR, it is not possible to have a large Rcl and Spc at the same time, and depending on the application, we have to choose either having a large Rcl or a large Spc to set the parameter δ.

4.6 Conclusion

In this chapter, by considering a linear signal extraction, we targeted the predicted output SNR for the 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, *i.e.* $Pr(\text{SNR}(\mathbf{X}_\mathcal{M}) \geq \theta)$.

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 J_E 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 gains of the already placed batches of sensors is available.

Numerical results showed a consistent superiority of the proposed criterion compared with the classical kriging and the average SNR criteria in terms of the output SNR and robustness against the uncertainty on the model of the spatial gain.

In the next chapter, we will present a gradient-based technique to develop a fast optimization algorithm. This technique is applied on our first proposed criterion J_E in a one dimensional case, which can also be extended for the new criterion in this chapter J_P, as well as higher dimensional cases.

Chapitre 5

Gradient-based optimization approach

Sommaire

So far in this thesis, by assuming a stochastic model, we were focused on presenting some robust sensor placement criteria based on the maximization of the SNR of the desired source. To this end, we first suggested to target the average SNR J_E in Chapter 3, and then, a more general and robust criterion based on the pdf of the SNR J_P was proposed in Chapter 4. Both methods have shown a significant improvement compared to the classical kriging approaches using criteria based on the entropy J_H, and the mutual information J_{MI} in terms of the output SNR. In all these cases, to solve the maximization problem, we used a greedy approach in which the sensors are added one by one from a fixed grid of candidate sensor positions. However, the greedy method is restricted to be on a predefined grid and by increasing the size of the grid, the computational cost will be increased, too. Consequently, to have better results, the grid should be fine enough, leading to a high computational cost. To solve these issues, in this chapter, we propose a first order optimization-based approach that in contrast to the one-by-one strategy adopted by the greedy method on a grid, optimizes all the sensor positions at once and does not discretize the search space. In addition, since placing 2 sensors very close to each other may not be feasible, *e.g.* due to the physical size of the sensors, a regularizing term is added to avoid choosing too close sensor positions. As the cost function is non-convex, to avoid bad local minima, the proposed algorithm is initialized by the solution of the greedy approach. In this chapter, we will implement our proposed optimization method on our first suggested criterion J_E. However, it can be extended to the other criteria as well.

In the rest of this chapter, we first present the details of the proposed method in Section 5.1. Then, Section 5.2 presents the performance of the proposed method and compares it with greedy approach. Finally, the conclusion is presented in Section 5.3.

5.1 Proposed method

In Chapter 3 we showed that if we assume that M sensors are located at positions $\mathbf{X}_{\mathcal{M}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M]^T$, by using a linear source extractor, and having a GP assumption on the spatial gain and the noise, then, the the mean of the SNR as a criterion to be optimized is obtained as follows :

$$J_E(\mathbf{X}_{\mathcal{M}}) = \mathbf{m}^a(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) + \mathrm{Tr}\left[\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})\right], \tag{5.1}$$

where $\mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \in \mathbb{R}^{M \times M}$ and $\mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}}) \in \mathbb{R}^{M \times M}$ are the covariance matrices of the spatial gain and the noise, respectively. Moreover, $\mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) = [m^a(\mathbf{x}_1), m^a(\mathbf{x}_2), \ldots, m^a(\mathbf{x}_M)]^T$ is the set of means at locations $\{\mathbf{x}_i\}_{i \in \{1,\cdots,M\}}$. Then, since directly maximizing (5.1) in a grid requires a combinatorial search, leading to a high computational cost, a greedy approach has been introduced in Chapter 3 that selects the M sensors by sequentially selecting $N < M$ sensors at a time. Assuming that K sensors have already been placed, to choose the locations of the next N sensors, the following criterion is optimized :

$$J_E(\mathbf{X}_{\mathcal{N}} | \mathbf{X}_{\mathcal{K}}) = \mathbb{E}\left\{ \left[\boldsymbol{a}(\mathbf{X}_{\mathcal{K}})^T,\ \boldsymbol{a}(\mathbf{X}_{\mathcal{N}})^T\right] \begin{bmatrix} \mathbf{C}^n(\mathbf{X}_{\mathcal{K}},\ \mathbf{X}_{\mathcal{K}}) & \mathbf{C}^n(\mathbf{X}_{\mathcal{K}},\ \mathbf{X}_{\mathcal{N}}) \\ \mathbf{C}^n(\mathbf{X}_{\mathcal{N}},\ \mathbf{X}_{\mathcal{K}}) & \mathbf{C}^n(\mathbf{X}_{\mathcal{N}},\ \mathbf{X}_{\mathcal{N}}) \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{a}(\mathbf{X}_{\mathcal{K}}) \\ \boldsymbol{a}(\mathbf{X}_{\mathcal{N}}) \end{bmatrix} \middle| \mathbf{X}_{\mathcal{K}} \right\}. \tag{5.2}$$

However, optimizing the above function by a combinatorial search and by increasing the grid size significantly increases the computational cost.

In this section, we present our proposed framework to solve the optimization problem for sensor placement. Unlike the greedy approach, our proposed method directly provides the positions of all the required number of sensors. By considering a one dimensional situation, we want to minimize $f(\mathbf{X}_{\mathcal{M}}) \triangleq -J_E(\mathbf{X}_{\mathcal{M}})$ where J_E is presented in (5.1). In order to control the average distances between each pair of the sensors, we constrain the sum of the squared distances to be greater than a threshold. Furthermore, due to the spatial constraints of the boundaries, we consider a normalized case where $0 \leq \mathbf{x}_i \leq 1$ [1]. Therefore, we study the following minimization problem

$$\min_{\mathbf{X}_{\mathcal{M}}} f(\mathbf{X}_{\mathcal{M}}) \quad \text{s.t.} \quad \begin{cases} \|\mathbf{D}\mathbf{X}_{\mathcal{M}}\|_2^2 \geq \epsilon \\ 0 \leq \mathbf{x}_i \leq 1, \quad i \in \{1, 2, \ldots, M\} \end{cases} \tag{5.3}$$

where $\mathbf{D} \in \mathbb{R}^{\frac{M(M-1)}{2} \times M}$ is a matrix that enumerates all the possible combinations of positions in pairs of size two *i.e.* $\|\mathbf{D}\mathbf{X}_{\mathcal{M}}\|_2^2 = \sum_{i=1}^{M} \sum_{j>i}^{M} |\mathbf{x}_i - \mathbf{x}_j|^2$. For instance, if the number of sensors is $M = 3$, then

$$\mathbf{D}\mathbf{X}_{\mathcal{M}} = \underbrace{\begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{pmatrix}}_{\mathbf{X}_{\mathcal{M}}} = \begin{pmatrix} \mathbf{x}_1 - \mathbf{x}_2 \\ \mathbf{x}_1 - \mathbf{x}_3 \\ \mathbf{x}_2 - \mathbf{x}_3 \end{pmatrix}.$$

1. Note that to keep the consistency of this chapter, in the experimental part we represent the scalars with bold lowercase letters.

So, $\|\mathbf{D}\mathbf{X}_{\mathcal{M}}\|_2^2$ shows the averaged distances between each pair of sensor positions, and by constraining it to be greater than ϵ, we want the minimum distances between the sensors to be greater than ϵ in average. The MATLAB code to generate the matrix $\mathbf{D}$ is presented in Appendix B. To solve (5.3), we define an auxiliary variable $\mathbf{z}(\mathbf{X}_{\mathcal{M}}) \triangleq \mathbf{D}\mathbf{X}_{\mathcal{M}}$, and reformulate (5.3) as the following problem :

$$\min_{\mathbf{X}_{\mathcal{M}},\mathbf{z}(\mathbf{X}_{\mathcal{M}})} f(\mathbf{X}_{\mathcal{M}}) \quad \text{s.t.} \quad \begin{cases} \mathbf{z}(\mathbf{X}_{\mathcal{M}}) \in \mathcal{A}_\epsilon, \\ \mathbf{z}(\mathbf{X}_{\mathcal{M}}) = \mathbf{D}\mathbf{X}_{\mathcal{M}}, \\ 0 \leq \mathbf{x}_i \leq 1, \quad i \in \{1, \ldots, M\}, \end{cases} \tag{5.4}$$

where $\mathcal{A}_\epsilon = \left\{ \mathbf{z}(\mathbf{X}_{\mathcal{M}}) \in \mathbb{R}^M \;\middle|\; \|\mathbf{z}(\mathbf{X}_{\mathcal{M}})\|_2^2 \geq \epsilon \right\}$. To solve (5.4), we use the penalty method [Ber99], by adding the constraint $\mathbf{z}(\mathbf{X}_{\mathcal{M}}) = \mathbf{D}\mathbf{X}_{\mathcal{M}}$ as a penalty to the target function with the penalty parameter α :

$$\begin{aligned} \min_{\mathbf{X}_{\mathcal{M}},\mathbf{z}(\mathbf{X}_{\mathcal{M}})\in\mathcal{A}_\epsilon} & \left\{ f(\mathbf{X}_{\mathcal{M}}) + \frac{1}{2\alpha}\|\mathbf{z}(\mathbf{X}_{\mathcal{M}}) - \mathbf{D}\mathbf{X}_{\mathcal{M}}\|_2^2 \right\} \\ & \text{s.t.} \quad 0 \leq \mathbf{x}_i \leq 1 \quad i \in \{1, \ldots, M\}. \end{aligned} \tag{5.5}$$

Then, an alternating minimization method is used to optimize it. At iteration l, the cost is first optimized over $\mathbf{z}(\mathbf{X}_{\mathcal{M}})$, fixing $\mathbf{X}_{\mathcal{M}}$ to its current estimate $\mathbf{X}_M^{(l)}$. That is :

$$\mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(l)} = \underset{\mathbf{z}(\mathbf{X}_{\mathcal{M}})\in\mathcal{A}_\epsilon}{\operatorname{argmin}} \frac{1}{2\alpha}\|\mathbf{z}(\mathbf{X}_{\mathcal{M}}) - \mathbf{D}\mathbf{X}_{\mathcal{M}}^{(l)}\|_2^2. \tag{5.6}$$

The solution to the above minimization is a projection onto the set $\mathcal{A}_\epsilon$ as follows :

$$\mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(l)} = \begin{cases} \mathbf{X}_{\mathcal{M}}^{(l)}, & \text{if} \quad \|\mathbf{X}_{\mathcal{M}}^{(l)}\|_2^2 \geq \epsilon \\ \dfrac{\mathbf{X}_{\mathcal{M}}^{(l)}}{\|\mathbf{X}_{\mathcal{M}}^{(l)}\|_2^2}\epsilon, & \text{otherwise.} \end{cases} \tag{5.7}$$

For the second step, the variable $\mathbf{z}(\mathbf{X}_{\mathcal{M}})$ is fixed as in (5.7), and we do the minimization over $\mathbf{X}_{\mathcal{M}}$ as follows :

$$\mathbf{X}_{\mathcal{M}}^{(l+1)} = \underset{\mathbf{X}_{\mathcal{M}}}{\operatorname{argmin}} \left\{ f(\mathbf{X}_{\mathcal{M}}) + \frac{1}{2\alpha}\|\mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(l)} - \mathbf{D}\mathbf{X}_{\mathcal{M}}\|_2^2 \right\} \quad \text{s.t.} \quad 0 \leq \mathbf{x}_i \leq 1, \quad i \in \{1, \ldots, M\}. \tag{5.8}$$

Since the constraint is a quite simple convex set, to solve (5.8), a projected gradient descent is used : after a gradient descent (GD) update, the result is projected onto $[0, 1]$. That is, by defining the cost function as follows :

$$g(\mathbf{X}_{\mathcal{M}}) = f(\mathbf{X}_{\mathcal{M}}) + \frac{1}{2\alpha}\|\mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(l)} - \mathbf{D}\mathbf{X}_{\mathcal{M}}\|_2^2, \tag{5.9}$$

the gradient step to optimize (5.8) is

$$\mathbf{X}_{\mathcal{M}}^{(l+1)} = \mathbf{X}_{\mathcal{M}}^{(l)} - \mu\nabla_{\mathbf{X}_{\mathcal{M}}} g(\mathbf{X}_{\mathcal{M}}^{(l)}), \tag{5.10}$$

where, $\nabla g(\mathbf{X}_{\mathcal{M}}^{(l)})$ is the gradient of the smooth function $g(.)$ at the previously updated point $\mathbf{X}_{\mathcal{M}}^{(l)}$, and $\mu > 0$ is a step size. To derive the gradient of $g(.)$, we can write :

$$\nabla_{\mathbf{X}_{\mathcal{M}}} g(\mathbf{X}_{\mathcal{M}}^{(l)}) = \nabla_{\mathbf{X}_{\mathcal{M}}} f(\mathbf{X}_{\mathcal{M}}^{(l)}) - \alpha^{-1}\mathbf{D}^T(\mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(l)} - \mathbf{D}\mathbf{X}_{\mathcal{M}}^{(l)}). \tag{5.11}$$

We use the chain rule to calculate $\nabla f(\mathbf{X}_{\mathcal{M}}^{(l)})$ as follows :

$$\begin{aligned}\frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{X}_{\mathcal{M}}^{(l)}) =& \mathrm{Tr}\left[\left(\frac{\partial f}{\partial \mathbf{m}^a}(\mathbf{X}_{\mathcal{M}})\right)^T \frac{\partial \mathbf{m}^a}{\partial \mathbf{x}_i}(\mathbf{X}_{\mathcal{M}})\right] \\ &+\mathrm{Tr}\left[\left(\frac{\partial f}{\partial \mathbf{C}^a}(\mathbf{X}_{\mathcal{M}})\right)^T \frac{\partial \mathbf{C}^a}{\partial \mathbf{x}_i}(\mathbf{X}_{\mathcal{M}})\right] \\ &+\mathrm{Tr}\left[\left(\frac{\partial f}{\partial \mathbf{C}^n}(\mathbf{X}_{\mathcal{M}})\right)^T \frac{\partial \mathbf{C}^n}{\partial \mathbf{x}_i}(\mathbf{X}_{\mathcal{M}})\right].\end{aligned}$$

where $\mathbf{x}_i$ is the i^{th} element of $\mathbf{X}_{\mathcal{M}}$. The above expression is simplified to the following :

$$\begin{aligned}\frac{\partial f(\mathbf{X}_{\mathcal{M}}^{(l)})}{\partial \mathbf{x}_i} = \mathrm{Tr}\Big[&- 2\mathbf{m}^a(\mathbf{X}_{\mathcal{M}})^T \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \frac{\partial \mathbf{m}^a(\mathbf{X}_{\mathcal{M}})}{\partial \mathbf{x}_i} \\ &- \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \frac{\partial \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})}{\partial \mathbf{x}_i} \\ &+ \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} [\mathbf{m}^a(\mathbf{X}_{\mathcal{M}}) \mathbf{m}^a(\mathbf{X}_{\mathcal{M}})^T \\ &+ \mathbf{C}^a(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})] \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})^{-1} \frac{\partial \mathbf{C}^n(\mathbf{X}_{\mathcal{M}}, \mathbf{X}_{\mathcal{M}})}{\partial \mathbf{x}_i}\Big],\end{aligned}$$

where $\dfrac{\partial \mathbf{m}^a}{\partial \mathbf{x}_i}(\mathbf{X}_{\mathcal{M}}) = \left[\dfrac{\partial m_j^a}{\partial \mathbf{x}_i}\right]_j$, m_j^a denotes the jth element of the mean vector, and $\dfrac{\partial \mathbf{C}}{\partial \mathbf{x}_i}(\mathbf{X}_{\mathcal{M}}) = \left[\dfrac{\partial \mathbf{C}_{ij}}{\partial \mathbf{x}_i}\right]_{(i,j)}$, in which $\mathbf{C}$ represents any covariance matrix, with $\mathbf{C}_{ij}$ corresponding to its (i,j)th entry. In this way, we have computed the gradient of $f(\mathbf{X}_{\mathcal{M}}^{(l)})$ over the i^{th} sensor position, providing thus the expression of the gradient vector $\nabla_{\mathbf{X}_{\mathcal{M}}} f(\mathbf{X}_{\mathcal{M}}^{(l)})$.

To determine the gradient step size μ in (5.10), we use a backtracking line search strategy [Ber99]. Starting with an initial value for μ, this strategy sequentially reduces μ by a decreasing factor $0 < \beta < 1$ until the following inequality is satisfied

$$g(\mathbf{X}_{\mathcal{M}}^{(l)} - \mu \nabla g(\mathbf{X}_{\mathcal{M}}^{(l)})) > g(\mathbf{X}_{\mathcal{M}}^{(l)}) - \frac{\mu}{2}\varrho^{(l)},$$

where $\varrho^{(l)} = \|\nabla g(\mathbf{X}_{\mathcal{M}}^{(l)})\|_2^2$. By satisfying the above condition we aim to have a large enough decrease of $g(.)$. After updating $\mathbf{X}_{\mathcal{M}}$ using (5.10), all the element of $\mathbf{X}_{\mathcal{M}}^{(l+1)}$ are projected into $[0, 1]$.

Finally, to solve (5.5) we start with an initial point and alternate between the projection step (5.7) and the GD step (5.10). As done in standard penalty methods [Ber99], the problem (5.5) should be solved for a decreasing sequence of α, *i.e.* $\{\alpha_0, \alpha_1, \ldots\}$, where $\alpha_{j+1} = \eta\alpha_j$, with $0 < \eta < 1$. For each fixed value of α, we perform a few iterations between (5.7), (5.10),

Algorithm 3 Alternating minimization (AM) for solving (5.4)

Inputs : $\left\{\mathbf{X}_{\mathcal{M}}^{(0)}, \mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(0)}\right\}$, μ_0, α_0, Q (number of iterations)
Initialization : Set $\mu = \mu_0$, $l = 0$
for $j = 1, 2, \cdots, Q$ **do**
  while stopping criterion not met **do**
    $\varrho^{(l)} = \|\nabla g(\mathbf{X}_{\mathcal{M}}^{(l)})\|_2^2$
    while $g(\mathbf{X}_{\mathcal{M}}^{(l)} - \mu \nabla g(\mathbf{X}_{\mathcal{M}}^{(l)})) > g(\mathbf{X}_{\mathcal{M}}^{(l)}) - \frac{\mu}{2}\varrho^{(l)}$ **do**
      $\mu \leftarrow \beta \cdot \mu$
    end while
    $\mathbf{X}_{\mathcal{M}}^{(l+1)} = \mathbf{X}_{\mathcal{M}}^{(l)} - \mu \nabla g(\mathbf{X}_{\mathcal{M}}^{(l)})$
    Project $\mathbf{X}_{\mathcal{M}}^{(l+1)}$ into $[0, 1]$
    Perform projection (5.7) to obtain $\mathbf{z}(\mathbf{X}_{\mathcal{M}})^{(l+1)}$
    $l \leftarrow l + 1$
  end while
  $\alpha_{j+1} = \eta \cdot \alpha_j$
end for
Output : $\mathbf{X}_{\mathcal{M}}^{(l)}$

and then projecting $\mathbf{X}_{\mathcal{M}}$ between 0 and 1. Moreover, iterations corresponding to α_{j+1} are initialized by the final estimate found for α_j. The final algorithm to solve (5.3) is summarized in Algorithm 3.

Since the problem (5.4) is non-convex, its initialization is important to find an appropriate minimizer. We propose to initialize the algorithm with the solution obtained by the greedy approach we used in Chapter 3. In this way, the algorithm is more likely to end up with a good local minimum.

5.2 Numerical experiments

In this section, we present some experiments to evaluate the performance of the proposed method. First, in Subsection 5.2.1 the numerical setup is explained. Then the influence of the initialization is studied in Subsections 5.2.2. Afterwards, in Subsection 5.2.3 we show the effect of the regularization based on the sensors distances. Finally, the effect of the smoothness of the spatial gain in the performance is presented in Subsection 5.2.4.

5.2.1 Numerical setup

Synthetic data are generated in a 1D space, where the range of the sensor locations is normalized between 0 and 1. We consider a prior on the spatial gain and noise to be generated

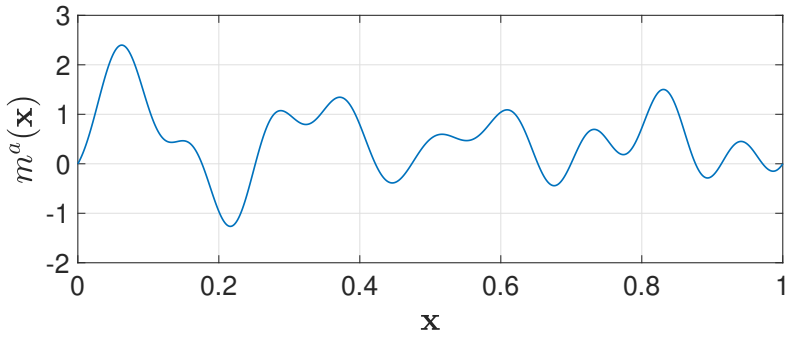


FIGURE 5.1: The mean of the spatial gain $m^a(\mathbf{x})$.

from $\mathcal{GP}\big(m(\mathbf{x}), C(\mathbf{x}, \mathbf{x}')\big)$, with a square exponential covariance function

$$C(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp^{\frac{-(\mathbf{x}-\mathbf{x}')^2}{2\rho^2}} .$$

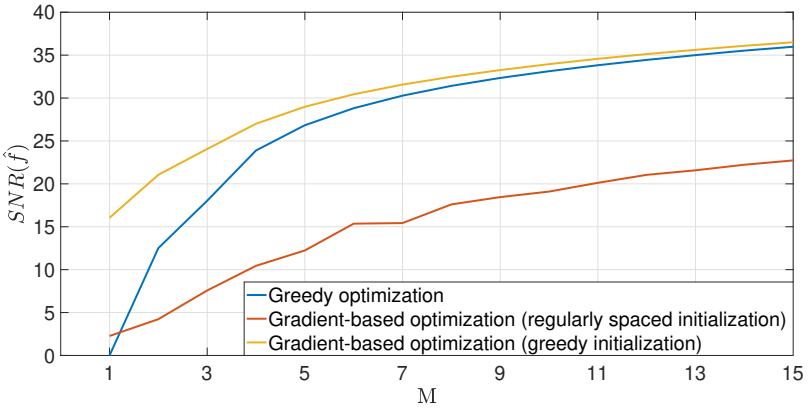
The mean of the noise is set to be 0. The mean of the gain is given by

$$m^a(\mathbf{x}) = \sum_{i=1}^{5} \gamma_i \sin^{d_i}(w_i \pi \mathbf{x}), \tag{5.12}$$

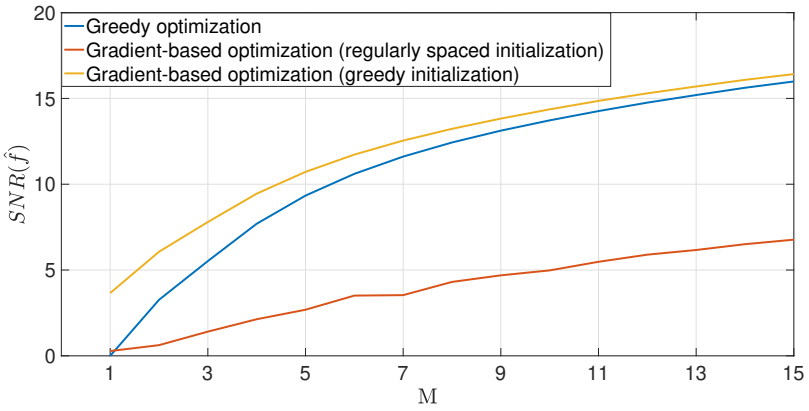
where, γ_i, d_i and w_i are the i^{th} elements of the vectors $\mathcal{G} = [0.1, 0.3, 0.5, 0.7, 0.9]$, $\mathcal{D} = [1, 1, 3, 1, 2]$, and $\mathcal{W} = [5, 6, 7, 8, 9]$, respectively. The reason that we choose (5.12) for the mean of the spatial gain is to control the behaviour of the spatial gain in terms of the number of local optimizer points, as well as their positions in the spatial grid. With the selected parameters, the mean is depicted in Fig. 5.1. The smoothness parameters ρ_n and ρ_a, and the variances σ_n and σ_a as well as the size of the spatial grid for greedy initialization take different values for each experiment. Also, through this section, we used 50 MC realizations of the spatial gain and the noise for different experiments. Finally, concerning the algorithm parameters, we set $\alpha_0 = 1$, $Q = 50$, $\eta = 0.5$, $\mu_0 = 1$, and $\beta = 0.5$.

5.2.2 Influence of the initialization

In this part, we study the influence of the initialization on the performance of the proposed method. We set the size of the spatial grid to be 100. Two different values of the uncertainty on the spatial gain are considered : $\sigma_a = 1$ and $\sigma_a = 3$. The noise variance σ_n is accordingly set such that the SNR becomes 0 dB. The smoothness of the uncertainty on the spatial gain (ρ_a) is set to $\rho_a = 0.001$, corresponding to an uncertainty with almost no spatial correlation. Since in this experiment we want to add several sensors, we have selected a small value for ρ_a which leads to using more sensors. The spatial smoothness of the noise ρ_n is set to $\rho_n = 0.01\rho_a$. By having M sensors, the number of the sensor pairs becomes $\frac{M(M-1)}{2}$, and if we assume the distances between each two pairs of the sensors not to be greater than 10^{-3}, then we can set the lower bound ϵ on $\|\mathbf{X}_{\mathcal{M}}\|_2^2$ to be $\epsilon = \frac{M(M-1)}{2} \times 10^{-3}$.



(a) $\sigma_a = 1$



(b) $\sigma_a = 3$

FIGURE 5.2: Influence of the initialization on the gradient-based optimization approach. Here, the output SNR *vs.* the number of sensors is presented for two different values of σ_a.

The true value of the output SNR, *i.e.* $\text{SNR}(\hat{\mathbf{f}}|\mathbf{X}_M) = (\sigma_s^2 \hat{\mathbf{f}}^T \mathbf{a}_M \mathbf{a}_M^T \hat{\mathbf{f}})/(\hat{\mathbf{f}}^T \mathbf{C}_{MM}^n \hat{\mathbf{f}})$ with the estimated extraction vector $\hat{\mathbf{f}}(\mathbf{X}_\mathcal{M}) = \mathbf{C}^n(\mathbf{X}_\mathcal{M}, \mathbf{X}_\mathcal{M})^{-1}\mathbf{m}^a(\mathbf{X}_\mathcal{M})$ versus the number of desired sensors M is depicted in Fig. 5.2. If no prior information is used for the initialization, one can initialize the sensor locations regularly-spaced between 0 and 1. On the contrary, one can use the previously proposed greedy approach, where the sensors are added one by one, to initialize the sensor locations. Firstly, one can see that the greedy initialization leads to a better extraction of the source $s(t)$ than using regularly-spaced initial locations for the sensors before applying our proposed method to adjust the sensor locations. In Fig. 5.2-a, we can see that for $\sigma_a = 1$ the difference of output SNRs using regularly-spaced initialization and greedy initialization, varies between 12dB for a single sensor and 13dB for 15 sensors. Also considering $\sigma_a = 3$ in Fig. 5.2-b, these values are 4dB for a single sensor and 10dB for 15 sensors. So, we can clearly see the improvement of the proposed method by using the greedy initialization.

Moreover, the proposed method (the yellow curve) for adjusting the sensor locations leads to improving the SNR compared to the greedy approach (the blue curve). For instance, by

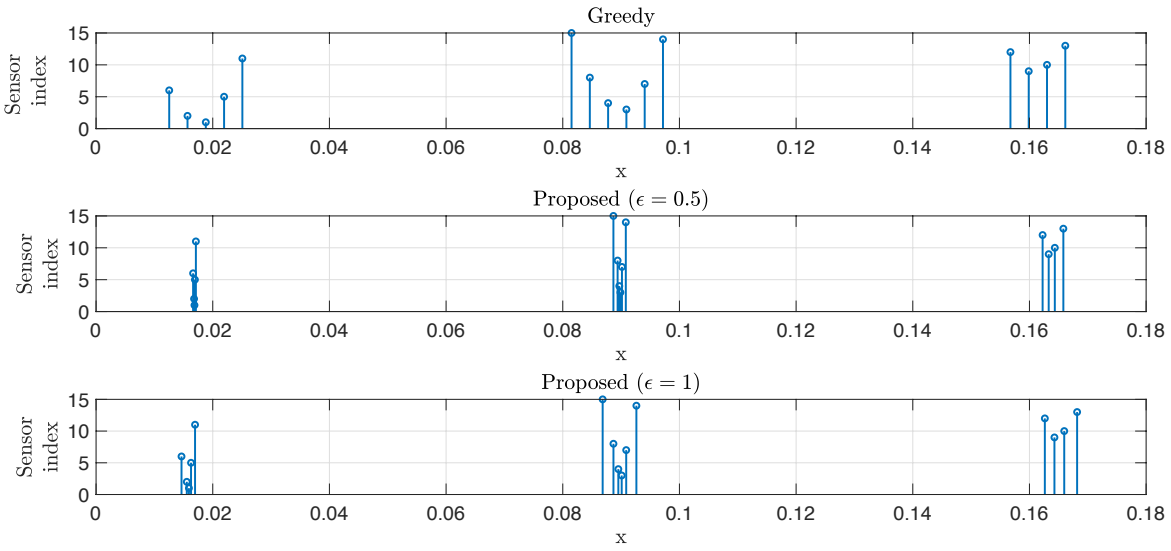


FIGURE 5.3: Effect of the regularization parameter ϵ to control the sensor distances. Top : initial sensors localisation, middle and bottom : final sensors localisation for $\epsilon = .5$ and $\epsilon = 1$, respectively.

considering $\sigma_a = 1$ in Fig. 5.2-a, and by adding 3 sensors, the greedy method improves the SNR up to 17dB while this amount is 24dB for the proposed method (with greedy initialization), which is 7dB better than the greedy method. Indeed, this result is expected since the proposed method tackles the optimization of the sensor locations all at the same time instead of one after the other as in the greedy method. It is also worth noting that the output SNR is worse by applying the proposed method with a regularly-spaced initialization than by just choosing the sensor locations by the greedy method with no additional adjustment. This can be due to the fact that our problem is non-convex, and as such, by having a bad initialization, the gradient-based approach can converge to a bad local minimum.

5.2.3 Regularizing sensors distances

Figure 5.3 shows the effect of regularizing sensor distances and how it can help to control the average distances between pairs of sensors. In this part, all the parameters are set as in the previous section with $\sigma_a = 1$, except that here we consider a tighter grid of size 320. Also, the number of desired sensors is set to be $M = 15$. For the proposed method, two different values for the lower bound are considered : $\epsilon = 0.5$ and $\epsilon = 1$. We note that we intentionally choose the parameters such that we have most of the information between $0 < \mathbf{x} < 0.2$ and the positions of the sensors suggested by the algorithm not being so scattered in the space. In this way, we can better visualize and compare the effect of ϵ in controlling the distances between the sensors. In Fig. 5.3 we can see that the sensors are located in three clusters, one cluster is the range $\mathbf{x} \in [0.01, 0.03]$, the second one is in $\mathbf{x} \in [0.08, 0.1]$, and the third one is in $\mathbf{x} \in [0.15, 0.17]$.

Now, let's start with the result obtained from the greedy method by looking at the first sub-figure. In this sub-figure, we can see that there are regular spaces between the sensors approximately equal to $1/320 \sim 0.003$, which is exactly according to the size of the grid, *i.e.* 320. It means that, if there is some important information within the grid in each cluster, this information will be lost as the greedy approach can not observe within the grid, and to solve this issue, the only solution is to increase the size of the grid which causes a high computational cost due to the combinatorial search that should be done in the greedy method.

In the second and the third sub-figures we have demonstrated the results obtained from the gradient-based approach for two different values of the parameter, $\epsilon = 0.5$ and $\epsilon = 1$, to show the effect of this parameter in tuning the average distances between the sensors. By comparing the second and the third sub-figures, we can see that after increasing ϵ from 0.5 to 1, in each cluster the average distances between the sensors are increased. To better compare the results obtained from these two experiments, we have calculated the average distances between the sensors. The value of $||\mathbf{DX}||$ for the second sub-figure ($\epsilon = 0.5$) is reported equal to 0.8509, while this value is 0.8647 for the third sub-figure where we set $\epsilon = 1$.

We also mention that by comparing the results of the gradient-based approach in the second and the third sub-figures with the results from the greedy approach in the first sub-figure, we can see that by using the gradient-based approach we no longer have the problem of the grid size, and we can easily observe the whole space without increasing the size of the grid and, thus, without increasing the computational cost.

The final SNR values for each approach, from top to bottom, are 29.22 dB (initial SNR), 32.22 dB ($\epsilon = 0.5$), and 31.53 dB ($\epsilon = 1$), respectively. As we can see, since the gradient-based method can search for the information within the grid for sensor placement, we obtained better output SNR. On the other hand, increasing ϵ leads to only a slight decrease of the output SNR while increasing the average distance between the sensors, which is a positive aspect of the proposed method in case of having limitations regarding the minimum distances between the sensors.

5.2.4 Effect of the smoothness parameter ρ_a

In this part, we study the performance for different smoothness levels of the uncertainty of the spatial gain (ρ_a). We consider an almost difficult situation compared to the previous experiment and we increase the uncertainty parameter of the spatial gain to $\sigma_a = 5$. We also considered SNR to be 10 dB to set σ_n. Note that, here, we consider a large enough SNR so that we are able to study the difficult situations in terms of the smoothness of the noise and the spatial gain. The rest of the parameters are set similarly as in Subsection 5.2.2. In Fig. 5.4, the output SNR versus the degree of the spatial gain smoothness, *i.e.* ρ_a, is depicted. A large ρ_a indicates situations where we are dealing with smooth conditions such as representing the maternal tissues of a pregnant women to extract the fetal ECG. A small ρ_a corresponds to more complicated situations where the source signal is attenuated in a heterogeneous environment, *e.g.* an acoustic signal passing through the ocean where there are many obstacles of different

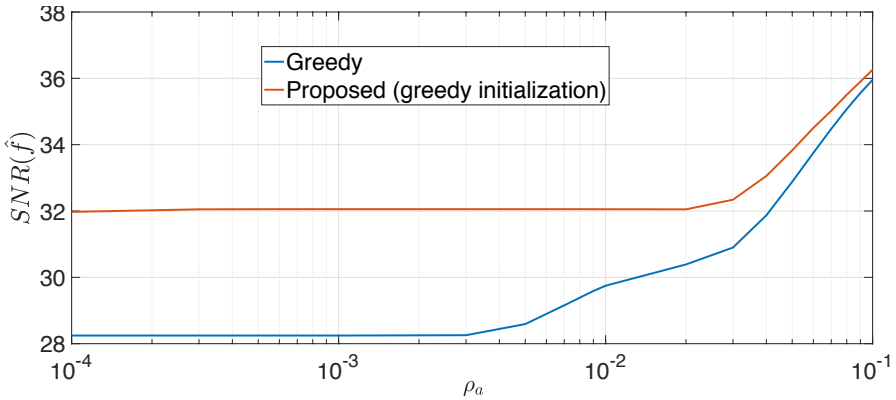


FIGURE 5.4: Effect of the smoothness parameter of spatial gain ρ_a.

types on the way of the source signal. We can see that as the signal becomes more non-smooth (*i.e.* if ρ_a decreases), the performance of the greedy approach deteriorates much faster than the proposed method. This is due to the presence of highly informative sensor positions in between grid points, which cannot be chosen by the greedy approach.

5.3 Conclusion

In this chapter, we addressed optimal sensor placement for signal extraction by maximizing output signal to noise ratio. In contrast to the greedy approach proposed in Chapter 3, our new proposed method adjusts all the sensors locations at once instead of choosing them one at a time. To this end, a gradient-based method is proposed to search for the sensor locations over the whole space. A constraint, controlling the average distances between the sensors, is also considered to avoid choosing too close sensors (*e.g.*, depending on their size). Due to the non-convexity of the cost function, the proposed algorithm is initialized with the solution of the greedy approach. Experimental results demonstrate that the proposed method provides about 3 dB improvements over the greedy approach. Also, thanks to the new constraint, the proposed method is shown to be able to control the average distances between the sensors. In future works, an explicit constraint on each distance between pair of sensors can be studied as well as global optimization algorithms to avoid converging to a local optimum.

Chapitre 6

Conclusions and perspectives

6.1 Conclusions

In this thesis, we have studied the problem of optimal sensor placement for signal extraction from noisy measurements. This problem is crucial if the data are collected by a limited number of sensors. To this end, we proposed different criteria, as well as a novel optimization approach to target the problem. First, in Chapter 3, the average output SNR of the linearly extracted signal has been proposed as a quality criterion to predict sensor locations. Such a criterion includes the uncertainty on the spatial gain of the source to be extracted, providing a suitable solution for the optimal sensor placement problem. Numerical simulations have shown the superior efficiency and accuracy of the proposed method in the source extraction problem when compared to the classical sensor placement criteria such as entropy and mutual information.

Since the SNR is a continuous random variable in our setting, we derived its pdf in Chapter 4. We then 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 in Chapter 3 can be seen as a special case of the new study. Also, to reduce the computational cost of evaluating the criterion, we proposed a sequential approach where new sensor locations are chosen in batches. We showed how to update this sequential version when some information on the the gains of the already placed batches of sensors is available. Numerical results demonstrated a consistent superiority of the proposed criterion compared with the classical kriging and the average SNR criteria in terms of the output SNR and robustness against the uncertainty on the model of the spatial gain.

Finally, in Chapter 5, our last contribution was focused on presenting a new optimization approach to solve the sensor placement problem. In contrast to the proposed greedy approach, the new method adjusts all the sensors locations at once instead of choosing them one at a time. To this end, a gradient-based method is proposed to search for the sensor locations over the whole space. A constraint, controlling the average distances between sensors, was also considered to avoid choosing too close sensors (*e.g.*, depending on their size). Due to the non-convexity of the cost function, the proposed algorithm is initialized with the solution of the greedy approach. Experimental results demonstrated that the proposed method provides about 3 dB improvements over the greedy approach. Also, thanks to the new constraint, the

proposed method is shown to be able to control the average distances between the sensors.

6.2 Perspectives

There are several future research topics regarding the proposed approaches. Some of these perspectives are complementary works that are along with the works done so far in this thesis, and we categorize them as *short-term perspectives*. Meanwhile, there are other interesting works that can be done on this topic that require more fundamental studies besides what we have done in this thesis, and we consider them as *long-term perspectives*. In the following, we explain some of these future works, separately.

• Short-term perspectives

— **Noise uncertainty :** In this thesis, we did not take into account the uncertainty on the noise. As a perspective, we can improve our proposed method by taking into account the uncertainty of the noise and derive the pdf of the SNR based on the distribution of both the spatial gain and the distribution of the noise, for instance based on the *Wishart distribution* [Wis28].

— **Hyperparameters estimation :** In this thesis, estimation of the hyperparameters of the models of the noise and the spatial gain is not considered. This will be an important problem to be tackled in further studies. For instance, full-Bayesian inference can be helpful to fill this gap in our proposed methods and provide a good estimation of the hyperparameters of our models [Bis06] (Chapter 10).

— **High dimensional settings :** In this thesis, we 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 optimization algorithm for maximizing our criterion. As an example, the gradient-based technique proposed in Chapter 5 can be extended for our proposed robust criterion J_P in Chapter 4. Also, to extend our proposed method to 2-D and 3-D spatial spaces, we can take advantage of the *Wishart distribution* as a generalization to multiple dimensions of the gamma distribution [Wis28].

• Long-term perspectives

— **Spatial gain estimation :** In the sequential approach we used in this thesis, we have stated that information on the gains of the already placed sensors can be retrieved with independent component analysis (ICA) or sparse component analysis (SCA). As

a future work, such an approach where sensor placement is coupled with the blind source separation (BSS) techniques can be considered.

— **Multiple source extraction :** In this thesis, we only considered that we want to extract a single source from a set of measurements. However, extending or proposed method to the extraction of several sources requires to perfectly study the BSS techniques [CJ10a].

— **Trade-off between the SNR improvement and the complexity :** As mentioned in this thesis, although by adding new sensors we probably have improvement in the average output SNR, *i.e.* $\overline{SNR}(\hat{\mathbf{f}})$, 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) [Aka98] 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.

— **Linear convolutive mixture model :** In this thesis, we consider a fast propagation for the source signal, and we used a linear instantaneous model to represent the sensor measurements. However, in case of the acoustic signals such as the PCG signals, it is essential to take into account the propagation delay. To this end, a linear convolutive mixture model must be used instead of the instantaneous model. In this case, although in the Fourier domain, we also have a linear representation, however, the criteria for estimating the filter (and possibly a huge number of parameters) are also complicated. So, extending the proposed criterion to this situation is an interesting challenge that can be taken into account as a perspective.

— **Dynamic design :** In many real-world applications, we are usually dealing with dynamic environments. For instance, in underwater acoustic communication, the spatial field between the acoustic source signal and the receiver sensors is always being changed due to the movement of the living creatures in the ocean, or the movement of the source, *e.g.* a whale which emits sound waves. Consequently, the corresponding spatial gain will be changed over time. Therefore, implementing our proposed method into a dynamic application requires to design a dynamic system, *e.g.* by using *Kalman filter* which is wildly used in many real-time applications [Kal60], or in case of a mobile source, we have to consider the changes in the system [Bas88] ; [BN93]. In addition, linear dynamical systems are well introduced in [Bis06] (Chapter 13.3), which can be useful to pursue this goal.

Annexe A

Proof of (4.23) : the cdf of $w(\mathrm{X}_{\mathcal{M}})$

From (4.21) we have

$$\begin{aligned} g_{\boldsymbol{w}}(w) =& g_{\Gamma_1}(w) * g_{\Gamma_2}(w) * \cdots * g_{\Gamma_M}(w) \\ =& \frac{1}{\prod_{i=1}^{M} d_i} g_{\nu_1}(\frac{w}{d_1}; k_1, \lambda_1) * g_{\nu_2}(\frac{w}{d_2}; k_2, \lambda_2) * \cdots * g_{\nu_M}(\frac{w}{d_M}; k_M, \lambda_M). \end{aligned}$$

Let's define $h(w) \triangleq g_{\Gamma_2}(w) * \cdots * g_{\Gamma_M}(w)$. So we have

$$g_{\boldsymbol{w}}(w) = g_{\Gamma_1}(w) * h(w) = \int_{-\infty}^{+\infty} h(\tau) g_{\Gamma_1}(w-\tau) d\tau. \tag{A.1}$$

On the other hand, the cdf of $g_{\boldsymbol{w}}(w)$ is

$$G_{\boldsymbol{w}}(\theta) = \int_{w=-\infty}^{\theta} g_{\boldsymbol{w}}(w) dw. \tag{A.2}$$

So, by replacing (A.1) in (A.2), we have

$$\begin{aligned} G_{\boldsymbol{w}}(\theta) &= \int_{w=-\infty}^{\theta} \int_{\tau=-\infty}^{+\infty} h(\tau)\ g_{\Gamma_1}(w-\tau)\ d\tau\ dw \\ &= \int_{\tau=-\infty}^{+\infty} h(\tau) \int_{w=-\infty}^{\theta} g_{\Gamma_1}(w-\tau)\ dw\ d\tau \\ &= \int_{\tau=-\infty}^{+\infty} h(\tau) G_{\Gamma_1}(w-\tau) d\tau \\ &= G_{\Gamma_1}(\theta) * h(w) \\ &= G_{\Gamma_1}(\theta) * g_{\Gamma_2}(\theta) * \cdots * g_{\Gamma_M}(\theta) \\ &= \frac{1}{\prod_{i=2}^{M} d_i} \times G_{\nu_1}\Big(\frac{\theta}{d_1}; k_1, \lambda_1\Big) * g_{\nu_2}\Big(\frac{\theta}{d_2}; k_2, \lambda_2\Big) * \cdots * g_{\nu_M}\Big(\frac{\theta}{d_M}; k_M, \lambda_M\Big). \end{aligned}$$

We note that since all the variables Γ_i for $i = 1,\ 2, \ldots,\ M$ play the same role, the above formula can be obtained as the convolution product of the cdf function of any Γ_i term and $h(w)$ being the convolution production of the other terms (different of Γ_i).

Annexe B

MATLAB code to generate matrix D

Matlab Code

```
num_comb = M*(M-1)/2;
x_perm = combnk(x_M,2);
[ismemb , ind_perm ] = ismember(x_perm,x_M);

D = zeros(num_comb,M);
D(:,ind_perm(:,1)) = 1;
D(:,ind_perm(:,2)) = -1;
```

Bibliographie

[Aka98] H. Akaike. "Information theory and an extension of the maximum likelihood principle". In : *Selected Papers of Hirotugu Akaike* 1 (1998), p. 199-213 (cf. p. 93).

[AS72] M. Abramowitz et I. A. Stegun. *Handbook of Mathematical Functions*. United States Department of Commerce, National Bureau of Standards (NBS), 1972 (cf. p. 54).

[Atk07] A. N. ; Tobias R. D. Atkinson A. C. ; Donev. "Optimum experimental designs, with SAS". In : *Oxford University Press* (2007), p. 511 (cf. p. 18).

[Atk88] A. C. Atkinson. "Recent developments in the methods of optimum and related experimental designs". In : *International Statistical Review / Revue Internationale de Statistique* 56.2 (1988), 99–115 (cf. p. 15).

[Atk96] A. C. Atkinson. "The usefulness of optimum experimental designs". In : *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), 59–76 (cf. p. 15).

[Bas88] M. Basseville. "Detecting changes in signals and systems—A survey". In : *Prentice Hall information and system sciences series* 24.3 (1988), p. 309-326 (cf. p. 93).

[Ber+05] J. W. Berry et al. "Sensor Placement in Municipal Water Networks". In : *Journal of Water Resources Planning and Management* 131.3 (2005) (cf. p. 3).

[Ber99] D. P. Bertsekas. *Nonlinear Programming*. Belmont, MA : Athena Scientific, 1999 (cf. p. 83, 84).

[Bis06] M. Ch. Bishop. "Pattern Recognition and Machine Learning". In : *Information Science and Statistics, Springer* 2 (2006) (cf. p. 22, 92, 93).

[BN93] M. Basseville et I. V. Nikiforov. "Detection of Abrupt Changes : Theory and Application". In : *Prentice Hall information and system sciences series* (1993) (cf. p. 93).

[BV04] S. Boyd et L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004, 384–396 (cf. p. 15, 18).

[Car98] J.-F. Cardoso. "Blind signal separation : statistical principles". In : *Proceedings of the IEEE* 86.10 (1998), p. 2009-2025 (cf. p. 8).

[CJ10a] P. Comon et C. Jutten, éd. *Handbook of Blind Source Separation*. Elsevier, 2010 (cf. p. 93).

[CJ10b] P. Comon et C. Jutten, éd. *Handbook of Blind Source Separation Independent Component Analysis and Applications*. Academic Press, 2010 (cf. p. 8).

[CL04] S. Y. Chen et Y. F. Li. "Automatic sensor placement for model-based robot vision". In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34.1 (2004), p. 393-408 (cf. p. 3).

[Com94] P. Comon. “Independent component analysis, a new concept ?” In : *Signal processing* 36.3 (1994), p. 287-314 (cf. p. 58).

[Cra46] Harald Cramér. “Mathematical Methods of Statistics”. In : *Princeton Univ. Press* (1946) (cf. p. 17).

[Cre90] N. Cressie. “The origins of kriging”. In : *Mathematical Geology* 22.3 (1990), p. 239-252 (cf. p. 6, 8, 34).

[Cre91] N. A. C. Cressie. “Statistics for Spatial Data”. In : (1991) (cf. p. 23).

[Cur+91] C. Currin et al. “Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments”. In : *Journal of the American Statistical Association* 86.416 (1991), p. 953-963 (cf. p. 23).

[CV95] K. Chaloner et I. Verdinelli. “Bayesian experimental design : A review”. In : *Statistical Science* 10.3 (1995), 273–304 (cf. p. 18).

[Dan91] C. K. Daniel. “Sensor placement for on-orbit modal identification and correlation of large space structures”. In : *Journal of Guidance, Control, and Dynamics* 14.2 (1991), p. 251-259 (cf. p. 3).

[Dar45] Georges Darmois. “Sur les limites de la dispersion de certaines estimations”. In : *Rev. Int. Inst. Statist* 13 (1945), 9–15 (cf. p. 17).

[Duc19] John Duchi. “Lecture Notes for Statistics 311/Electrical Engineering 377, Chapter 16”. In : *Stanford* (2019), p. 210-215 (cf. p. 17).

[Gar58] John J Gart. “An extension of the Cramér–Rao inequality”. In : *Ann. Math. Stat* 29 (1958), 367–380 (cf. p. 17).

[H71] Theil H. “Best Linear Unbiased Estimation and Prediction”. In : *Principles of Econometrics. New York : John Wiley Sons.* (1971), 119–124 (cf. p. 19).

[Her+00] H. J. Hermens et al. “Development of recommendations for SEMG sensors and sensor placement procedures”. In : *Journal of Electromyography and Kinesiology* 10.5 (2000), p. 361-374 (cf. p. 3).

[JSW98] D. R. Jones, M. Schonlau et W. J. Welch. “Efficient global optimization of expensive black-box functions”. In : *Journal of Global optimization* 13.4 (1998), p. 455-492 (cf. p. 8).

[Kal60] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In : *transactions of the American Society for mechanical engineering, Series D, Journal of basic engineering* 82 (1960), p. 35-45 (cf. p. 93).

[Kay93] Steven M. Kay. “Fundamentals of Statistical Signal Processing : Estimation Theoryy”. In : *Prentice Hall, signal processing series* 1 (1993), Chapter 4 (cf. p. 17, 19, 29).

[KLQ95] C. Ko, J. Lee et M. Queyranne. “An Exact Algorithm for Maximum Entropy Sampling”. In : *Operations Research* 43.4 (1995), p. 684-691 (cf. p. 23).

[Kra+08] A. Krause et al. “Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks”. In : *Journal of Water Resources Planning and Management* 134.6 (2008) (cf. p. 3).

[Kri51] D. G. Krige. “A statistical approach to some basic mine valuation problems on the Witwatersrand”. In : *Journal of the Southern African Institute of Mining and Metallurgy* 52.6 (1951), p. 119-139 (cf. p. 8).

[KSG08] A. Krause, A. Singh et C. Guestrin. “Near-Optimal Sensor Placements in Gaussian Processes : Theory, Efficient Algorithms and Empirical Studies”. In : *Journal of Machine Learning Research* 9 (2008), p. 235-284 (cf. p. 6, 8, 19-21, 24, 25, 34, 37, 38).

[Lin56] D. V. Lindley. “On a measure of the information provided by an experiment”. In : *Annals of Mathematical Statistics* 27 (1956), 986–1005 (cf. p. 22).

[Mal99] Gustave Malécot. “Statistical methods and the subjective basis of scientific knowledge”. In : *translated from Année X 1947 by Daniel Gianola]. Genet. Sel. Evol.* 31 (1999), 269–298 (cf. p. 17).

[MBC79] M. D. McKay, R. J. Beckman et W. J. Conover. “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”. In : *Technometrics* 21.2 (1979), p. 239-245 (cf. p. 23).

[MZ05] M. Meo et G. Zumpano. “On the optimal sensor placement techniques for a bridge structure”. In : *Engineering Structures* 27.10 (2005), p. 1488-1497 (cf. p. 3).

[NA75] A. Nuttall et H. Albert. “Some integrals involving the Q-M function”. In : *Transactions on Information Theory* 21.1 (1975), p. 95 -96 (cf. p. 55).

[Puk87] F. Pukelsheim. “Information increasing orderings in experimental design theory”. In : *International Statistical Review / Revue Internationale de Statistique* 55.2 (1987), 203–219 (cf. p. 15).

[Ram+05] N. Ramakrishnan et al. “Gaussian Processes for Active Data Mining of Spatial Aggregates”. In : *Proceedings of the 2005 SIAM International Conference on Data Mining* (2005), p. 12 (cf. p. 24).

[Rao45] Calyampudi Radakrishna Rao. “Information and the accuracy attainable in the estimation of statistical parameters”. In : *Bulletin of the Calcutta Mathematical Society* 37 (1945), 81–89 (cf. p. 17).

[Rao94] S. Das Gupta (editor) Rao Calyampudi Radakrishna. “Selected Papers of C. R. Rao”. In : *Wiley* (1994) (cf. p. 17).

[RW06] C. E. Rasmussen et C. K. I. Williams. *Gaussian processes for machine learning*. T. 1. MIT press Cambridge, 2006 (cf. p. 12, 13, 29, 52, 58).

[Sac+89] J. Sacks et al. “Design and analysis of computer experiments”. In : *Statistical science* (1989), p. 409-423 (cf. p. 8).

[Sam06] S. et al. Sameni. “Electrode Selection for Noninvasive Fetal Electrocardiogram Extraction using Mutual Information Criteria”. In : *26th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jul 2006, Paris, France.* (2006), 97–104 (cf. p. 6).

[Sha48] Claude E. Shannon. “A Mathematical Theory of Communication”. In : *Bell System Technical Journal* 27.4 (1948), 623–65 (cf. p. 22).

[SJ+18] A. L. D. Siqueira Junior et al. "Respiratory Waveform Estimation from Multiple Accelerometers : An Optimal Sensor Number and Placement Analysis". In : *IEEE journal of biomedical and health informatics* (2018) (cf. p. 3, 58).

[Ste12] M. L. Stein. *Interpolation of spatial data : some theory for kriging*. Springer Science & Business Media, 2012 (cf. p. 13).

[SW87a] M. C. Shewry et H. P. Wynn. "Maximum entropy sampling". In : *Journal of Applied Statistics* 14.2 (1987), p. 165-170 (cf. p. 6, 34).

[SW87b] M. C. Shewry et H. P. Wynn. "Maximum entropy sampling". In : *Journal of applied statistics* 14.2 (1987), p. 165-170 (cf. p. 8).

[SWN13] T. J. Santner, B. J. Williams et W. I. Notz. *The design and analysis of computer experiments*. Springer Science & Business Media, 2013 (cf. p. 8).

[VJV04] F. Vrins, Ch. Jutten et M. Verleysen. "Sensor Array and Electrode Selection for Non-invasive Fetal Electrocardiogram Extraction by Independent Component Analysis". In : *Independent Component Analysis and Blind Signal Separation, Fifth International Conference, ICA, Granada, Spain, September* (2004) (cf. p. 3, 6, 7).

[Wis28] J. Wishart. "The generalised product moment distribution in samples from a normal multivariate population". In : *Biometrika* 20A (1928), 32–52 (cf. p. 92).

[ZP01] M. Zibulevsky et B. A. Pearlmutter. "Blind source separation by sparse decomposition in a signal dictionary". In : *Neural computation* 13.4 (2001), p. 863-882 (cf. p. 58).

Abstract — Many signal processing problems can be cast from a generic setting where a source signal propagates through a given environment to the sensors. Under this setting, we can be interested either in (i) estimating the source signal, or (ii) the spatial field, or even (iii) the resulting field of signals in some regions of the environment. In all these cases, signals are recorded by multiple sensors located at different positions. Due to price, energy or ergonomic constraints, 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. The way to tackle the problem of optimal sensor placement depends on which of three aspects mentioned above we want to address.

In this thesis, we focus on estimating the source signal from a set of noisy measurements collected from a limited number of sensors. Our approach differs from classical kriging based optimal sensor placement approaches, since the latter focus on best reconstruction of the spatial measured field. For solving the problem, we propose a first criterion which maximizes the average signal to noise ratio of the estimated signal. Experimentally, performance obtained by this criterion outperforms the results obtained using kriging-based methods. Since the signal to noise ratio is uncertain in this context, to achieve a robust signal extraction, we propose a second placement criterion based on the maximization of the probability that the SNR exceeds a given threshold. This criterion can be easily evaluated using Gaussian process assumption for the signal, the noise and the environment. Moreover, to reduce the computational complexity of the joint maximization of the criterion with respect to all sensor positions, we propose a greedy algorithm where the sensor positions are sequentially (*i.e.* one by one) selected. Experimental results show the superiority of the probabilistic criterion compared to the average SNR criterion. Finally, for improving the sub-optimal greedy algorithm, we present an optimization approach to locate all the sensors at once. In this purpose, we add a constraint to the problem that can control the average distances between the sensors. To solve our problem, we use an alternating optimization penalty method. In the end, we present experimental results which show the superiority of the proposed algorithm over the greedy one.

Keywords : Optimal sensor placement, Kriging, Signal extraction, Signal to noise ratio, Gaussian processes, Alternating optimization, Penalty method.

GIPSA-lab, Grenoble Images Parole Signal Automatique,
11 rue des Mathématiques, Saint Martin d'Hères Cedex

Résumé — De nombreux problèmes de traitement du signal peuvent être résolus à partir d'un cadre générique où un signal source se propage vers les capteurs à travers un environnement donné. Dans ce cadre, on peut s'intéresser soit à (i) l'estimation du signal source, soit (ii) au milieu de propagation, soit même (iii) au champ de signaux résultant dans certaines régions de l'environnement. Dans tous ces cas, les signaux sont enregistrés par plusieurs capteurs. En raison de contraintes de prix, d'énergie ou d'ergonomie, le nombre de capteurs est souvent limité et il devient crucial de placer à des positions qui contiennent le maximum d'informations. Ce problème correspond à un placement optimal des capteurs et il se pose dans un grand nombre d'applications. La manière d'aborder le problème du placement optimal des capteurs dépend trois aspects mentionnés ci-dessus que nous voulons aborder.

Dans cette thèse, nous nous concentrons sur l'estimation du signal source à partir d'un ensemble de mesures bruitées collectées par un nombre limité de capteurs. Notre approche diffère des approches classiques de placement optimal des capteurs basées sur le krigeage, car ces dernières se concentrent sur la meilleure reconstruction du champ spatial mesuré. Pour résoudre le problème, nous proposons un premier critère qui maximise le rapport signal/bruit moyen du signal estimé. Expérimentalement, les performances obtenues par ce critère sont supérieures à celles obtenues avec les méthodes basées sur le krigeage. Comme le rapport signal/bruit est incertain dans ce contexte, pour obtenir une extraction robuste du signal, nous proposons un second critère de placement basé sur la maximisation de la probabilité que le rapport signal/bruit dépasse un seuil donné. Ce critère peut être facilement évalué en utilisant l'hypothèse de processus gaussien pour le signal, le bruit et l'environnement. De plus, pour réduire la complexité de calcul de la maximisation conjointe du critère par rapport à toutes les positions du capteur, nous proposons un algorithme glouton où les positions du capteur sont sélectionnées séquentiellement (*i.e.* une par une). Les résultats expérimentaux montrent la supériorité du critère probabiliste par rapport au critère minimisant le rapport signal/bruit moyen. Enfin, pour améliorer l'algorithme glouton sous-optimal, nous présentons un algorithme d'optimisation permettant de localiser tous les capteurs en même temps. Dans ce but, nous ajoutons une contrainte pour contrôler les distances moyennes entre les capteurs. Pour résoudre notre problème, nous utilisons une méthode d'optimisation alternée. Enfin, nous présentons des résultats expérimentaux qui montrent la supériorité de l'algorithme proposé sur l'algorithme glouton.

Mots clés : placement optimal de capteurs, krigeage, extraction du signal, rapport signal/bruit, processus gaussien, optimisation alternée, optimisation avec contraintes.

GIPSA-lab, Grenoble Images Parole Signal Automatique,
11 rue des Mathématiques, Saint Martin d'Hères Cedex