



**HAL**  
open science

# Binary decision for observations with unknown distribution : an optimal and invariance-based framework

Sabrina Bourmani

## ► To cite this version:

Sabrina Bourmani. Binary decision for observations with unknown distribution : an optimal and invariance-based framework. Signal and Image Processing. Ecole nationale supérieure Mines-Télécom Atlantique, 2020. English. NNT : 2020IMTA0173 . tel-02878810

**HAL Id: tel-02878810**

**<https://theses.hal.science/tel-02878810>**

Submitted on 23 Jun 2020

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

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

# THESE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE MINES-TELECOM ATLANTIQUE  
BRETAGNE PAYS DE LA LOIRE - IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 601  
*Mathématiques et Sciences et Technologies  
de l'Information et de la Communication*  
Spécialité : *Signal, Image, Vision*

Par

**Sabrina BOURMANI**

**Binary decision for observations with unknown distribution:  
An optimal and invariance-based framework.**

Thèse présentée et soutenue à IMT Atlantique, le 20 février 2020  
Unité de recherche : UMR CNRS 6285 LabSTICC  
Thèse N° : 2020IMTA0173

## Rapporteurs avant soutenance :

Abdourrahmane Atto      Maître de conférences, HDR à Polytech Annecy-Chambéry.  
Lang White                Professeur à l'Université d'Adelaïde.

## Composition du Jury :

Présidente : Geneviève Baudoin      Professeur à l'ESIEE.  
Examineurs : François-Xavier Socheleau      Maître de conférences, HDR à IMT atlantique.  
Lang White      Professeur à l'Université d'Adelaïde.  
Dir. de thèse : Dominique Pastor      Professeur à IMT atlantique.

## Invité :

Alain Hillion      Professeur émérite à IMT Atlantique.



This thesis has been prepared at the Signal and Communications Department of IMT Atlantique (UMR CNRS 6285 Lab-STICC).



# Acknowledgements

I would like to express my sincere gratitude to my supervisors: Dominique Pastor and François-Xavier Socheleau, first for giving me the opportunity to do this PhD thesis. Then, for their guidance, patience, and advice throughout all these past three years. They both have been exemplary in terms of support at all levels. They shared with me their immense enthusiasm for science and research, and for that, I will be eternally grateful.

I would like to thank Prof. Geneviève Baudoin for accepting to be president of the jury for my thesis defense, as well as Prof. Lang White and Dr. Abdourrahmane Atto for accepting to be rapporteurs. A special thanks to the former director of the Lab-STICC - UMR6285, Pr. Alain Hillion for accepting our invitation to be our guest the day of the defense.

I also thank each and every person in the SC department of IMT Atlantic with whom I interacted or collaborated closely or remotely. Namely, Martine Besnard whose encouragements helped me regain confidence in times of doubts.

Last but not least, I am extremely grateful to my parents for their trust, sacrifice and unconditional love. I am also very thankful to my second mother for her unwavering caring, and my sisters and friends for their empathy and indulgence.



# Abstract

In many applications, engineers and scientists must face the question of determining the presence or absence of some signal of interest, observed in noise and possibly with interference of other unwanted signals. This question can be answered through the mathematical study of *decision problems*.

During my studies, I took interest in decision problems where signals are assumed to be stochastic, but with unknown distributions. When this kind of hypothesis is assumed, in general, standard literature does not allow to seek optimal solutions. However, Random Distortion Testing (RDT) framework, developed a few years ago in our laboratory, has introduced optimality criteria suitable for such assumptions. Hence, we took interest in the philosophy behind the RDT framework, and we follow the same guidelines about not trying to know or estimate the unknown distributions of the signal, in the process of finding solutions for these decision problems. Apart from our optimality purposes, we also have an invariance based perspective in how we intend to solve this type of decision problems. Indeed, when there are uncertainties about the signal of interest, we can try to derive solutions that are invariant towards them. These are the two key notions we consider throughout our investigations.

In this manuscript, first, we apply the RDT framework for a distributed decision problem to test its suitability to such decision scenarios where the signal of interest is random with unknown distribution and where the observations are collected by a network of sensors instead of just one sensor. Then, we generalise the theoretical material of the RDT framework to when the noise is not necessary Gaussian while still considering the signal of interest random of unknown distribution, we called it the Generalised Random Distortion Testing (GRDT) problem. Finally, we adopt an asymptotic outlook to circumvent the limitations of the RDT and the developed GRDT approach. Although the considered decision scenarios concern the simple case where the signal of interest is deterministic, this last part of the thesis allows to think ahead and start considering the eventual perspective of the generalisation of the GRDT in the asymptotic scope.





# Résumé long

## Mots clés

Décision binaire, Invariance, Optimalité, Signaux aléatoires, Distribution inconnue.

## Introduction

### Contexte

Dans de nombreuses applications de traitement de signal comme le radar et le sonar, la sismologie, les télécommunications ou la surveillance médicale, les données et les signaux mesurés sont généralement observés dans du bruit, en présence, éventuellement, de signaux interférents ou indésirables. Sur la base de ces observations, le but est de décider si un ou plusieurs signaux d'intérêt sont présents.

Ces signaux peuvent être soit des conséquences d'un événement d'intérêt (e.g. vocalises de baleines, présence de cibles ou d'obstacles, ...) soit d'anomalies dont nous souhaitons être informés afin de prendre des mesures adéquates (e.g. arythmie cardiaque, tremblement de terre imminent, ...). La théorie de la décision est un outil utile pour répondre à ce genre de problématiques, dans de nombreux domaines d'application tels que l'économie, la médecine, l'électronique, la défense, la météorologie, etc.

Le choix du détecteur que nous décidons d'utiliser pour résoudre le problème de décision dépend d'une multitude de facteurs. Premièrement, la possibilité ou volonté d'inclure de l'information à priori sur différents éléments du problème, l'importance de garantir l'optimalité selon un certain critère ou encore la robustesse de la solution proposée face à de l'incertitude. Parfois, plusieurs propriétés difficiles à atteindre simultanément sont conjointement visées ce qui mène à des compromis. Par exemple, dans de nombreux scénarios de détection, l'ambition est de maximiser la probabilité de détection tout en minimisant la probabilité de fausse alarme.

La nature de l'application considérée est ce qui détermine quelle propriété cibler et par conséquent quel type d'approche utiliser. Par exemple, les tests Uniformément Plus Puissants (UPP), c'est-à-dire les tests ayant la probabilité de détection la plus élevée pour une probabilité de fausse alarme fixée, peuvent être utilisés pour la détection de catastrophes naturelles, la détection de cibles en sonar et radar, tandis que le risque bayésien et le critère de probabilité minimale d'erreur (MPE) peuvent être utilisés dans les communications numériques. Par conséquent, pour les applications centrées sur la sécurité, une certaine prudence et conservatisme sont essentiels et l'optimalité est de mise car l'effet d'une détection manquée ou d'une fausse détection peut être

catastrophique. Alors que dans d'autres applications, comme celles liées au contrôle de qualité, le but est d'être robuste en réalisant une bonne détection en moyenne.

## Motivation

Outre le domaine d'application, la nature des signaux à détecter doit également être prise en compte dans le choix de la solution. Les signaux qui se trouvent normalement dans la nature ou qui sont produits par des dispositifs, peuvent être classés en deux catégories: les signaux déterministes et les signaux stochastiques. Dans la littérature, il existe de nombreuses méthodes pour la détection de signaux déterministes, cependant cette hypothèse peut s'avérer trop restrictive, et parfois non réaliste dans certaines applications. En effet, certains signaux déterministes par nature peuvent être "randomisés" par l'environnement, le canal de transmission ou le récepteur, pour ensuite rentrer dans la catégorie du modèle stochastique. Par ailleurs, des signaux déterministes inconnus peuvent être supposés aléatoires car trop peu d'information est disponible à leur sujet. Par conséquent, une large gamme de signaux peut être modélisée comme étant des processus aléatoires dans différentes applications et rentrer dans la catégorie des signaux stochastiques.

Malheureusement, même si considérer les signaux stochastiques, plutôt que déterministes, est déjà plus réaliste pour certaines applications, supposer que leur distribution est connue, peut également être contraignant. Cela peut être le cas pour les applications où les signaux d'intérêt sont pollués par des signaux d'interférence inconnus. Par conséquent, nous considérons un modèle alternatif, où le signal est supposé aléatoire de distribution inconnue, mais sans intention de la connaître ou de l'estimer. Nous nommons ce modèle l'approche conditionnelle.

De notre point de vue, les approches de décision binaires de la littérature, pour les signaux déterministes et stochastiques, sont basées sur des modèles inconditionnels. La raison en est que le critère qui peut être considéré dans le processus d'obtention des solutions est intégral, c'est-à-dire que les fonctions de densité de probabilité impliquées ne sont pas conditionnelles. Alors que lorsque les signaux sont aléatoires de distribution inconnue, la fonction de densité de probabilité (d.d.p) de l'observation est conditionnelle et les critères qui peuvent être associés au problème ne sont pas intégraux. Pour plus de concision, considérons une observation  $\mathbf{Y} = \Theta + \mathbf{X}$  tel que  $\Theta$  est le signal aléatoire de distribution inconnue et  $\mathbf{X}$  un bruit additif. Pour un modèle inconditionnel, la stratégie de détection se basera sur la fonction de densité de probabilité  $f_{\mathbf{Y}}(\mathbf{y})$ . Évidemment, pour ce faire,  $f_{\mathbf{Y}}(\mathbf{y})$  doit être connu, à un ou plusieurs paramètres près. Au contraire, lorsque  $f_{\mathbf{Y}}(\mathbf{y})$  est inconnu, le modèle devient conditionnel car la fonction de densité de probabilité utilisé dans le processus de décision devient conditionnel:  $f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y})$ . Lorsque  $\mathbb{P}_{\Theta}$ , la distribution du signal, est connue, il est possible de basculer vers le modèle inconditionnel standard et considérer les fonctions de densités de probabilité inconditionnelles:  $f_{\mathbf{Y}}(\mathbf{y}) = \int f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y})\mathbb{P}_{\Theta}d\theta$  pour résoudre le problème de décision. Par conséquent, connaître ou non  $\mathbb{P}_{\Theta}$  est ce qui détermine le type d'approche; conditionnel ou inconditionnel.

En d'autres termes, dans la théorie de la décision, les signaux à détecter pour le modèle conditionnel sont aléatoires de distribution inconnue, et pour le modèle inconditionnel, les signaux sont aléatoires ou déterministes de distribution connue. Nous

---

considérant que les signaux déterministes sont un cas particulier des signaux aléatoires puisqu'ils suivent une distribution de Dirac.

La communauté scientifique du traitement de signal a mené une vaste investigation en ce qui concerne le modèle inconditionnel, et plusieurs solutions peuvent être trouvées dans la littérature, en fonction de la quantité d'informations disponibles sur le signal d'intérêt, le bruit et les signaux interférents possibles. Cependant, pour notre modèle alternatif: « le modèle conditionnel », il y a clairement peu de travaux qui s'y intéressent. C'est dans ce contexte qu'intervient ma thèse: « Décision binaire pour des observations de distribution inconnue: Une approche optimale basée sur l'invariance ».

Notre terminologie est quelque peu différente de celle utilisée dans la théorie de l'estimation, où les modèles « conditionnels / inconditionnels » ne sont liés qu'à la nature du signal « déterministe / aléatoire ». Cependant, cette terminologie est particulièrement appropriée pour le contexte de nos travaux pour mettre en lumière l'intérêt de l'approche adoptée. En effet, cela montre qu'il existe une lacune dans la littérature et que les chercheurs n'ont pas cherché à trouver des solutions optimales quand les signaux sont aléatoires avec des distributions inconnues. Ainsi, nous ciblons les problèmes de détection où les signaux sont non seulement aléatoires mais aussi de distribution inconnue *sans intention de les estimer*. Par conséquent, nous considérerons des problèmes de décision dans le cadre des modèles conditionnels.

Cette terminologie particulière nous a aidé à classer les méthodes existantes dans la littérature. La première catégorie concerne le modèle inconditionnel pour lequel une panoplie de détecteurs existe dans la théorie de la détection. À savoir, le test de probabilité d'erreur minimale, les tests Uniformément Plus Puissants, le test de rapport de vraisemblance généralisé, le test de Wald, le test de Rao, l'approche bayésienne, les tests robustes, etc. En ce qui concerne la catégorie du modèle conditionnel, à notre connaissance, seul la méthode du RDT, pour «Random Distortion testing» existe. Cette méthode a été élaboré par [Pastor and Nguyen, 2013a] dans le laboratoire d'IMT Atlantique il y a quelques années, proposant une solution optimale selon un certain critère pour ce type de problèmes.

Nos travaux sont en concordance avec la philosophie du RDT et l'esprit du modèle conditionnel, puisque nous voulons trouver des détecteurs, sans avoir à supposer de connaissance sur la distribution des signaux, tout en recherchant l'optimalité. L'accent a été mis sur le modèle conditionnel afin de développer une formulation plus générale du cadre RDT.

En plus de notre objectif d'éviter d'avoir à faire des hypothèses contraignantes sur les distributions des signaux, notre intention est d'utiliser les propriétés d'invariance des problèmes de détection, pour trouver une solution, plus spécifiquement un test statistique, qui vérifie un certain critère d'optimalité sans connaître la distribution du signal.

Notre intérêt découle de la pénurie de travaux dans la littérature s'intéressant aux problèmes de détection de signaux aléatoires avec des distributions inconnues. Par conséquent, les enjeux des directions de recherche envisagées sont considérables.

Le manuscrit peut être divisé en deux parties. La première partie comporte les chapitres de l'état de l'art en lien avec la théorie de la décision. Plus spécifiquement, le premier chapitre présente les approches de la littérature permettant de résoudre des problèmes de détection de signaux quand ces derniers suivent le modèle inconditionnel. Le deuxième chapitre expose la notion d'invariance dans le contexte des problèmes de décision, et le troisième chapitre introduit l'approche RDT, comme étant, à notre connaissance, la seule méthode de la littérature qui permet de trouver une solution optimale à un problème de décision quand le signal d'intérêt est supposé aléatoire de distribution inconnue, *sans intention de la connaître*. Quant à la deuxième partie du manuscrit, elle se compose des chapitres contenant nos contributions. Ainsi, le quatrième chapitre concerne l'application de l'approche RDT dans un contexte de décision distribuée. Dans le cinquième chapitre, l'approche RDT est généralisée dans le cas d'un bruit additif pas nécessairement Gaussien. Enfin, le sixième et dernier chapitre présente une approche asymptotique pour la recherche de tests optimaux et plus spécifiquement les tests Asymptotiquement Uniformément Plus Puissant. Notre contribution était de démontrer une formulation asymptotique du théorème de Karlin-Rubin puis l'appliquer à un problème spécifique de décision binaire.

## Partie I: État de l'art

### Chapitre 1: Méthodes inconditionnelles de la littérature

Nous présentons d'abord quelques notions clés de la théorie de la détection comme l'optimalité et la robustesse. Ensuite, nous présentons diverses méthodes issues de la littérature qui entrent dans le cadre du modèle inconditionnel.

#### **Optimalité:**

Lorsque nous voulons résoudre un problème de décision, l'idéal serait de trouver une solution, plus spécifiquement un test statistique, qui ait des propriétés d'optimalité. Le principe d'optimalité repose fortement sur ce que nous pouvons appeler les critères d'optimalité. Il peut y avoir un seul critère ou plusieurs critères à optimiser conjointement. Ainsi, il faut extraire des spécificités de l'application et du problème de décision leurs exigences afin de bien déterminer ces critères. Tout au long de cette thèse, l'optimalité est présentée comme un concept fondamental afin de trouver des solutions aux problèmes de décision qui nous intéressent. Nous l'associons généralement ici à un test obtenu après la maximisation de la probabilité de détection tout en limitant la probabilité de fausse alarme ou toute probabilité analogue. Il convient de noter cependant que ce n'est pas la seule optimalité qui existe, mais nous pouvons dire que c'est ce genre d'optimalité qui nous intéresse. Cette optimalité au sens de Neyman-Pearson [Neyman and Pearson, 1933] est très appropriée pour les applications de type radar et sonar. Mais il existe aussi d'autres critères d'optimalité par exemple le critère de probabilité minimum d'erreur (MPE), le risque bayésien, etc.

Il faut souligner que les tests statistiques aux propriétés d'optimalité ne sont pas toujours envisageables. Parfois, le problème ne contient pas suffisamment d'informations pour dériver des tests optimaux et le signal et/ou le bruit et/ou l'interférence, si présente, ne sont pas suffisamment connus. Dans cette thèse, très souvent, c'est ce genre de scénarios qui est considéré. Et c'est ainsi que nous nous sommes intéressés aux méthodes invariantes. Cependant, l'invariance n'est pas la seule alter-

native.

En effet, une autre façon de résoudre un problème de décision est de dériver des tests robustes tels que les tests de Huber. La robustesse peut être ciblée soit parce que l'optimalité est difficilement atteignable pour les raisons que nous avons mentionnées précédemment, soit parce que le problème lui-même le requiert. Dans le second scénario, la robustesse est une nécessité pour gérer les problèmes d'inadéquation de sous-espace lorsque le signal d'intérêt est en présence de phénomènes physiques qui ne sont pas modélisés.

### Robustesse:

Considérer qu'un problème de décision est parfaitement défini, et que le seul manque de connaissances qu'il contient est encapsulé dans les paramètres inconnus présentés, est beaucoup trop optimiste. Dans la pratique, il est évident que les sources d'imprécision peuvent être nombreuses. La première source d'imprécision peut être la présence d'interférences inattendues, inconnues ou difficilement caractérisables qui peuvent faire dévier le signal d'intérêt de son modèle nominal. Une autre source d'imprécision non trivial est celle de l'environnement de propagation du signal qui peut induire en erreur si elle n'est pas prise en compte parfaitement dans le processus de modélisation. Enfin, la source d'erreur la plus simple est lorsque les instruments de mesure introduisent des incertitudes. Ainsi, la conception de tests robustes devrait garantir une certaine fiabilité vis à vis des écarts autour du modèle nominal.

En 1981, Huber a développé une méthodologie pour les tests d'hypothèses robustes [Huber, 1981]. Il a considéré que sous chaque hypothèse, une multitude de fonctions de densité de probabilité sont possibles et le but est de trouver la solution en considérant que l'observation suit la fonction de densité de probabilité la moins favorable sous chaque hypothèse. Cette approche, considérée comme assez pessimiste, est appelé la méthode du Minimax.

### Quelques méthodes inconditionnelles de la littérature

- **Le test de probabilité d'erreur minimale (MPE)**

Le test de probabilité d'erreur minimale est basé sur l'hypothèse que la probabilité d'occurrence de chaque hypothèse est connue de telle sorte que  $\mathbb{P}(\mathcal{H}_0) + \mathbb{P}(\mathcal{H}_1) = 1$ . Selon l'application, cette hypothèse peut être plus ou moins raisonnable. Cette croyance préalable sur les hypothèses participe à la dérivation du seuil auquel le rapport de vraisemblance est comparé. Similairement aux tests Uniformément Plus Puissants, aux tests de rapport de vraisemblance généralisés, etc., cette méthode appartient également à la catégorie des tests qui reposent sur des rapports de vraisemblance pour concevoir des détecteurs. Considérons le problème des hypothèses binaires suivant:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Y} = \boldsymbol{\theta}_0 + \mathbf{X} & \text{avec: } \mathbf{Y} \sim f_{\mathbf{Y}}^{\boldsymbol{\theta}_0} \\ \mathcal{H}_1 : \mathbf{Y} = \boldsymbol{\theta}_1 + \mathbf{X} & \text{avec: } \mathbf{Y} \sim f_{\mathbf{Y}}^{\boldsymbol{\theta}_1} \end{cases} \quad (1)$$

avec  $\boldsymbol{\theta}_0$  et  $\boldsymbol{\theta}_1$  le signal respectivement sous l'hypothèse nulle et sous l'hypothèse alternative. Le bruit est modélisé par  $\mathbf{X}$ , l'observation par  $\mathbf{Y}$  et  $f_{\mathbf{Y}}^{\boldsymbol{\theta}_i}$  est la fonction de densité de probabilité de  $\mathbf{Y}$  paramétrée par  $\boldsymbol{\theta}_i$ .

La probabilité d'erreur est définie comme:

$$P_e = \mathbb{P}(\mathcal{H}_0 | \mathcal{H}_1)\mathbb{P}(\mathcal{H}_1) + \mathbb{P}(\mathcal{H}_1 | \mathcal{H}_0)\mathbb{P}(\mathcal{H}_0) \quad (2)$$

La probabilité  $\mathbb{P}(\mathcal{H}_i | \mathcal{H}_j)$  indique la probabilité de choisir  $\mathcal{H}_i$  alors que  $\mathcal{H}_j$  est vraie et la probabilité  $\mathbb{P}(\mathcal{H}_i)$  est la probabilité d'occurrence de l'hypothèse  $\mathcal{H}_i$ . Le test (MPE) est alors:

$$\mathcal{T}_e(\mathbf{y}) = \begin{cases} 1 & \text{si } \frac{\mathbb{P}(\mathbf{y}|\mathcal{H}_1)}{\mathbb{P}(\mathbf{y}|\mathcal{H}_0)} > \frac{\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_1)} = \lambda_e \\ 0 & \text{sinon} \end{cases} \quad (3)$$

avec  $\lambda_e$  le seuil auquel le rapport de vraisemblance est comparé.

- **Les tests Uniformément Plus Puissants (UPP)**

Un test Uniformément Plus Puissant est un test statistique qui garantit la puissance, c'est-à-dire la probabilité de détection, la plus élevée, parmi tous les tests de même niveau  $\gamma$ , c'est-à-dire avec une probabilité de fausse alarme inférieure à  $\gamma$ . Quand le problème de décision est simple, le théorème de Neyman-Pearson donne l'expression du test UPP. Soit  $\mathbf{Y}$  un vecteur aléatoire de fonction de densité de probabilité  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  avec  $\theta \in \Theta$ . Le test UPP de niveau  $\gamma$  pour tester par exemple  $\mathcal{H}_0 : \theta = \theta_0$  contre  $\mathcal{H}_1 : \theta = \theta_1$  est, selon le théorème de Neyman-Pearson:

$$\mathcal{T}_{NP}(\mathbf{y}) = \begin{cases} 1 & \text{si } \frac{f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} > \lambda_{NP} \\ 0 & \text{sinon} \end{cases} \quad (4)$$

où le seuil  $\lambda_{NP} \geq 0$  est calculé de telle sorte que  $\mathbb{E}_{\theta_0}[\mathcal{T}_{NP}(\mathbf{Y})] = \gamma$  et  $0 \leq \gamma \leq 1$ .

Quand le problème de décision est composite le théorème de Karlin-Rubin donne l'expression du test UPP. Soit  $\mathbf{Y}$  un vecteur aléatoire de pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  où  $\theta \in \Theta \subset \mathbb{R}$  et soient  $\Theta_1$  et  $\Theta_0$  deux sous-ensembles de  $\Theta$  tels que  $\Theta = \Theta_0 \cup \Theta_1$  et  $\Theta_0 \cap \Theta_1 = \emptyset$ . Le problème de décision considéré est:

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1 \end{cases} \quad (5)$$

Le rapport de vraisemblance est défini comme:

$$\Lambda_{\theta_1, \theta_0}(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} \quad (6)$$

avec  $f_{\mathbf{Y}}^{\theta_0} \neq 0$ , pour  $\theta_1 \in \Theta_1$  et  $\theta_0 \in \Theta_0$ . Soit le rapport de vraisemblance tel que  $\Lambda_{\theta_0, \theta_1} = h_{\theta_0, \theta_1}(V)$ , avec  $V : \mathbb{R}^N \mapsto \mathbb{R}$ . Le test  $\mathcal{T}_{KR}$  est défini comme:

$$\mathcal{T}_{KR}(\mathbf{y}) = \begin{cases} 1 & \text{si } V(\mathbf{y}) > \lambda_{KR} \\ 0 & \text{sinon} \end{cases} \quad (7)$$

avec  $\lambda_{KR}$  déterminé de telle sorte que:  $\mathbb{E}_{\theta_0}[\mathcal{T}_{KR}(\mathbf{Y})] = \gamma$ . Selon le théorème de Karlin-Rubin, si pour  $\theta_1 > \theta_0$ ,  $h_{\theta_1, \theta_0}$  augmente strictement en  $V$ , alors le test  $\mathcal{T}_{KR}$  est UPP de niveau  $\gamma$  pour tester  $\mathcal{H}_1$  contre  $\mathcal{H}_0$ .

Bien que les tests UPP garantissent une certaine optimalité, ils nécessitent la connaissance de tous les paramètres du problème. En pratique, les paramètres peuvent être inconnus, ce qui peut mener à l'utilisation d'autres méthodes de l'état de l'art, comme les méthodes se basant sur le principe du maximum de vraisemblance; par exemple le test du rapport de vraisemblance généralisé (GLRT), le test de Wald, ou encore le test de Rao.

- **Le test du rapport de vraisemblance généralisé (GLRT)**

Le test du rapport de vraisemblance généralisé ou (GLRT) pour "Generalised Likelihood Ratio Test" consiste à calculer le ratio de la fonction de densité de probabilité (pdf) de l'observation sous chaque hypothèse, "le rapport de vraisemblance", puis à remplacer les paramètres inconnus par leurs estimations du maximum de vraisemblance (MLE) pour "Maximum Likelihood Estimate". La statistique de test résultante est ensuite comparée à un seuil. Nous considérons un problème de détection composite où  $\mathbf{Y}$  est un vecteur aléatoire de fonction de densité de probabilité (pdf)  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  et où  $\theta \in \Theta \subset \mathbb{R}$ . Pour considérer un cas simple, nous supposons que le paramètre  $\theta$  est inconnu sous l'hypothèse alternative et connu sous l'hypothèse nulle tel que:

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0 \end{cases} \quad (8)$$

Le test du (GLRT) [Kay, 2009] est:

$$L_G(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\hat{\theta}_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} > \lambda_G \quad (9)$$

ou  $\hat{\theta}_1 = \underset{\theta_1}{\operatorname{argmax}} f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})$  est le MLE de  $\theta_1$ , c'est-à-dire  $\theta$ , lorsque l'hypothèse alternative  $\mathcal{H}_1$  est vraie. Le seuil  $\lambda_G$  est défini de sorte à garantir une certaine probabilité de fausse alarme, mais malheureusement il n'y a aucune garantie de respecter la contrainte due à l'utilisation des MLE. Cependant, en pratique, le GLRT est connu pour donner de très bons résultats même s'il ne présente aucun gage d'optimalité.

- **L'approche Bayésienne**

En décision statistique binaire, l'approche bayésienne implique d'avoir un modèle paramétrique d'observation et d'attribuer des probabilités à priori, souvent appelées "prior", à des paramètres inconnus. Les "priors" sont des distributions de probabilité qui reflètent la quantité d'informations que nous avons ou que nous pensons avoir sur les paramètres inconnus. L'approche Bayésienne est légèrement différente du risque Bayésien rappelé précédemment, où seule la probabilité d'occurrence des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  est utilisée et où l'optimalité selon un certain critère est assurée. Alors qu'avec l'approche Bayésienne, aucune optimalité ne peut être assurée et les "priors" sont attribués à tout paramètre inconnu sans que ce soit l'occurrence des hypothèses. De plus, l'approche bayésienne considère les paramètres inconnus comme des réalisations de variables aléatoires et les intègre dans l'expression des pdf.

- **Le test UBCP de Wald**

En 1943, Wald a présenté dans [Wald, 1943] les tests UBCP pour (Uniformly Best Constant Power) ou "Uniformément Meilleure Puissance Constante" pour les problèmes de test d'hypothèses composite. Différent du test classique de Wald, il peut être considéré comme la source d'inspiration pour les auteurs de [Pastor and Nguyen, 2013b] autour du framework RDT pour "Random Distortion Testing". Le test UBCP est donc très similaire au test RDT, mais a aussi une certaine ressemblance avec les tests UPP précédemment présentés. Wald définit



la classe de  $\mathcal{K}_\gamma$  te que:  $\mathcal{K}_\gamma = \{\mathcal{T} : \sup_{\theta \in \overline{\mathcal{R}^T}} \mathbb{P}[\mathcal{T}(\theta) = 1] \leq \gamma\}$  avec  $0 \leq \gamma \leq 1$  et  $\overline{\mathcal{R}^T}$  le complémentaire de la région critique d'un test  $\mathcal{T}$ . En d'autres termes, la classe de tests  $\mathcal{K}_\gamma$  est composée de tests de niveau  $\gamma$ . Soit  $k$  un paramètre réel et positif. Soit  $\mathcal{S}$  une famille de surfaces définies comme:  $\mathcal{S} = \{S_k : \theta \in \mathbb{R}^N \text{ tel que } \phi(\theta) = k\}$ , ou  $\phi(\theta)$  est une fonction de  $\theta$ . Un test  $\mathcal{T} \in \mathcal{K}_\gamma$  est UBCP sur  $\mathcal{S}$  si:

- Pour tout  $\theta$  et  $\theta'$  appartenant à la même surface  $S_k \in \mathcal{S}$ :  $\beta_{\mathcal{T}}(\theta) = \beta_{\mathcal{T}}(\theta')$ .
- Pour tout autre test  $\mathcal{T}' \in \mathcal{K}_\gamma$  qui vérifie le premier point i.e.  $\beta_{\mathcal{T}'}(\theta) = \beta_{\mathcal{T}'}(\theta')$ :  $\beta_{\mathcal{T}}(\theta) \geq \beta_{\mathcal{T}'}(\theta)$  pour tout  $\theta \in S_k$  et toutes les surfaces  $S_k$ .

La fonction  $\phi(\theta)$  peut représenter l'énergie du signal au carré par exemple. Le premier point de la définition spécifie la puissance constante sur une surface dans  $\mathcal{S}$  et le deuxième point définit le critère d'optimalité sur toute surface  $S_k$ . La notion de famille de surfaces  $\mathcal{S}$  peut être considérée comme un ensemble de points équivalents qui constituent des partitions, définis par rapport aux propriétés d'invariance du problème.

## Chapitre 2: Invariance dans la théorie de la détection

À notre connaissance, la notion d'invariance a commencé à voir le jour à la fin des années 1930, pendant la seconde guerre mondiale, avec des chercheurs tels que Wald, Scharf, Lehman, Borovkov et d'autres. En effet, en raison des recherches scientifiques intensives dans le domaine de la théorie de la détection pour la défense, il a été rapidement remarqué que dans certains scénarios, lorsqu'il n'existe pas de test Uniformément Plus Puissant, la théorie de la détection avait des limites difficiles à surmonter avec les approches standard. C'est ainsi que le concept d'invariance est apparu. Ainsi, dans la théorie de la détection, lorsqu'il est difficile de trouver un test optimal selon un certain critère, restreindre la classe de tests peut être envisagé. Une façon de construire la classe de tests est d'utiliser l'invariance inhérente au problème de détection. L'invariance peut être considérée comme une symétrie naturelle du problème à partir de laquelle nous pouvons extraire des informations pour restreindre la classe de tests dans laquelle une solution optimale est recherchée. Ainsi, l'invariance est une notion qui aide à obtenir l'optimalité lorsque les approches standard, telles que celles présentées précédemment, ne parviennent pas à la fournir.

Le principe d'invariance est fondamentalement basé sur un groupe approprié de transformations qui laisse le problème invariant, ce groupe résulte de la nature des propriétés de symétrie du problème. Ensuite, un maximal invariant, c'est-à-dire une statistique avec des propriétés de maximalité et d'invariance qui opère une réduction sur les données, est dérivé en conséquence. Le maximal invariant extrait sera utilisé comme statistique de test dans l'expression du test statistique et ainsi de la solution.

### Le groupe de transformation

La première notion importante liée à l'invariance est le groupe de transformation. La définition suivante d'un groupe est d'après [Eaton, 1983, Définition 6.1, p. 185].

**Définition.** *Un groupe  $(G, \circ)$  est un ensemble  $G$  avec une opération binaire  $\circ$  telle que les propriétés suivantes sont valables pour tous les éléments de  $G$ :*

- $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ .

- Il y a un élément unique dans  $G$ , noté  $e$ , tel que  $g \circ e = e \circ g = g$  pour tout  $g \in G$ . L'élément  $e$  est l'élément identité de  $G$ .
- Pour chaque  $g \in G$ , il y a un élément unique dans  $G$ , noté  $g^{-1}$  tel que  $g \circ g^{-1} = g^{-1} \circ g = e$ . L'élément  $g^{-1}$  est l'inverse de  $g$ .

En ce qui concerne la notion de groupe dans le contexte de l'invariance et de la théorie de la décision, l'opération  $\circ$  est implicite et n'est pas exprimée par un opérateur de sorte que l'on écrit par exemple  $g_1 g_2$  au lieu de  $g_1 \circ g_2$ . Comme nous l'avons dit précédemment, un groupe de transformation devrait émaner de la nature même du problème de décision considéré.

### Le maximal invariant

La définition suivante d'un maximal invariant est commune à toutes les références qui traitent du sujet, par exemple [Borovkov, 1998; Lehmann and Romano, 2005b; Scharf, 1991] etc.

**Définition.** Pour  $\mathbf{y} \in \mathbb{R}^n$ , une statistique  $\mathcal{M}(\mathbf{y})$  est considérée comme étant un maximal invariant associé au groupe de transformation  $G$  si:

- $\mathcal{M}[g(\mathbf{y})] = \mathcal{M}(\mathbf{y})$ , pour tout  $g \in G$ .
- $\mathcal{M}(\mathbf{y}_1) = \mathcal{M}(\mathbf{y}_2) \Rightarrow \mathbf{y}_2 = g(\mathbf{y}_1)$ , pour  $g \in G$ .

### Les tests invariants

S'il y a des symétries dans l'espace d'observation, représentés par le groupe  $G$ , il est naturel de limiter l'attention aux tests qui sont également invariants. Ainsi, il faut trouver le "meilleur" test invariant après avoir trouvé le groupe de transformations  $G$  qui laisse le problème invariant ainsi que la statistique d'un maximal invariant nécessaire dans l'élaboration d'un test invariant. Tout d'abord les tests invariants sont définis:

**Définition.** Soit  $G$  un groupe de transformation. Pour  $\mathbf{y} \in \mathbb{R}^n$ , un test d'hypothèse  $\mathcal{T}(\mathbf{y})$  est dit  $G$ -invariant si:  $\mathcal{T}[g(\mathbf{y})] = \mathcal{T}(\mathbf{y})$ , pour tous les  $g \in G$ .

Il est important de savoir que chaque test invariant  $G$  peut être écrit en fonction d'un maximal invariant:  $\mathcal{T}(\mathbf{y}) = \mathcal{T}[\mathcal{M}(\mathbf{y})]$ , tous deux liés au même groupe de transformation  $G$ .

## Chapitre 3: Le modèle conditionnel: l'approche RDT

### Le problème RDT:

Soit  $\mathbf{Y}, \Theta, \mathbf{X}$  trois vecteurs aléatoires réels de dimension  $N$ . Le signal d'intérêt  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  est supposé aléatoire de distribution inconnue. Nous rappelons que  $\mathcal{M}(\Omega, \mathbb{R}^N)$  est l'ensemble des vecteurs aléatoires réels de dimension  $N$ . Le bruit  $\mathbf{X}$  est supposé additif Gaussien tel que  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$  et  $\mathbf{C}$  sa matrice de covariance est définie positif. Le problème du test de distorsion aléatoire désigné par (RDT) [Pastor

and Nguyen, 2013a] est présenté comme:

$$\text{RDT:} \begin{cases} \text{Observation : } \mathbf{Y} = \Theta + \mathbf{X} \begin{cases} \Theta \text{ et } \mathbf{X} \text{ indépendants,} \\ \Theta \in \mathcal{M}(\Omega, \mathbb{R}^N), \\ \mathbf{X} \sim \mathcal{N}(0, \mathbf{C}), \end{cases} \\ \text{Événement nul: } [\nu_{\mathbf{C}}(\Theta - \theta_0) \leq \tau], \\ \text{Événement Alternatif: } [\nu_{\mathbf{C}}(\Theta - \theta_0) > \tau]. \end{cases} \quad (10)$$

avec  $\theta_0 \in \mathbb{R}^N$  un signal déterministe utilisé comme référence pour mesurer la distance au signal d'intérêt. la tolérance  $\tau \geq 0$  est utilisée comme paramètre pour modéliser l'écart par rapport au signal de référence  $\theta_0$  avec l'aide de la norme de Mahalanobis:  $\nu_{\mathbf{C}} : \mathbb{R}^N \rightarrow [0, \infty)$  défini par rapport à  $\mathbf{C}$  et affecte à n'importe quel  $\mathbf{x} \in \mathbb{R}^N$  le nombre réel positif  $\nu_{\mathbf{C}}(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}$ . Le problème (RDT) consiste à décider si  $\nu_{\mathbf{C}}(\Theta - \theta_0) \leq \tau$ , l'hypothèse nulle ou  $\nu_{\mathbf{C}}(\Theta - \theta_0) > \tau$ , l'hypothèse alternative, est vrai. Une telle formulation des hypothèses permet une meilleur gestion des incertitudes sur le signal. Ainsi, l'incorporation d'une tolérance  $\tau$ , pour modéliser ces incertitudes, permet une certaine souplesse vis-à-vis du manque de connaissance du signal d'intérêt  $\Theta$ . Le problème (RDT) est invariant sous l'action d'un groupe  $\mathcal{G}$ . Les orbites associées à  $\mathcal{G}$  peuvent être définies comme  $\Upsilon_{\rho} = \{\mathbf{y} \in \mathbb{R}^N : \nu_{\mathbf{C}}(\mathbf{y} - \theta_0) = \rho\}$  avec  $\rho \geq 0$ .

### Un test optimal

Les auteurs du framework RDT [Pastor and Nguyen, 2013b] définissent un critère d'optimalité relativement nouveau;  $\gamma$ -MCCP pour Maximal Constant Conditional Power:

**Définition.** Pour  $\tau \geq 0$  et  $\gamma \in (0, 1)$ , un test  $\mathcal{T}^*$  est dit  $\gamma$ -MCCP si:

- [**Level**]  $\mathcal{T}^*$  est de niveau  $\gamma$  (i.e. de probabilité de fausse alarme inférieure ou égale à  $\gamma$ ).
- [**MCCP**] Pour tout  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  et pour  $\mathbb{P}_{\nu_{\mathbf{C}}(\Theta - \theta_0)}$ -presque tout  $\rho > \tau$ , le test  $\mathcal{T}^*$  a une fonction de puissance conditionnelle constante pour tout  $\Theta \in \Upsilon_{\rho}$  et pour tout test  $\mathcal{T}$  de niveau  $\gamma$  avec une fonction de puissance conditionnelle constante sur  $\Upsilon_{\rho}$  nous avons:

$$\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}] \leq \mathbb{P}[\mathcal{T}^*(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}]$$

Selon les auteurs du RDT, un test  $\mathcal{T}$  a une fonction de puissance conditionnelle constante pour  $\Theta \in \Upsilon_{\rho}$  si  $\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}] = \beta_{\mathcal{T}}(\theta), \forall \theta \in \Upsilon_{\rho}$ , avec  $\beta_{\mathcal{T}}(\theta) = \mathbb{P}[\mathcal{T}(\mathbf{Y}) = 1]$ .

Le théorème suivant représente le résultat théorique fondamentale du framework RDT où le test optimal selon le critère  $\gamma$ -MCCP est donné pour le problème (RDT).

**Théorème.** Pour tout  $\gamma \in (0, 1)$  et tout  $\tau \geq 0$ , le test  $\mathcal{T}_{\lambda_{\gamma}(\tau)}$  défini comme:

$$\mathcal{T}_{\lambda_{\gamma}(\tau)}(\mathbf{y}) = \begin{cases} 1 \text{ si } \nu_{\mathbf{C}}(\mathbf{y} - \theta_0) > \lambda_{\gamma}(\tau) \\ 0 \text{ si } \nu_{\mathbf{C}}(\mathbf{y} - \theta_0) \leq \lambda_{\gamma}(\tau) \end{cases} \quad (11)$$

avec  $\lambda_{\gamma}(\tau)$  la solution unique en  $x$  de  $1 - \mathcal{R}(\tau, x) = \gamma$ , est  $\gamma$ -MCCP avec  $\mathbb{P}[\mathcal{T}_{\lambda_{\gamma}(\tau)}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}] = 1 - \mathcal{R}(\rho, \lambda_{\gamma}(\tau))$  pour tout  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  et  $\mathbb{P}_{\nu_{\mathbf{C}}(\Theta - \theta_0)}$ -presque tout  $\rho \geq 0$ .

$\mathcal{R}(\tau, \cdot)$  est la fonction de répartition de la racine carrée d'une variable aléatoire qui suit une distribution du  $\chi^2$  non centrée avec  $N$  degrés de liberté et  $\tau^2$  le paramètre de non-centralité.

### Conclusion

En conclusion, le cadre RDT a proposé des tests optimaux relativement nouveaux;  $\gamma$ -MCCP pour le problème de décider si la norme de Mahalanobis de la différence entre un signal d'intérêt et un signal de référence est supérieure ou inférieure à une tolérance  $\tau$ . Ce qui se trouve être intéressant autour de ce framework est que le signal d'intérêt  $\Theta$  est supposé aléatoire de distribution inconnue. De plus, le cadre RDT rassemble à la fois les avantages des méthodes paramétriques et ceux des méthodes non paramétriques. Plus spécifiquement; l'optimalité et la robustesse.

## Partie II: Contributions

### Chapitre 4: RDT distribué

Dans ce chapitre nous présentons notre première contribution, où nous appliquons le framework RDT pour un scénario de détection distribué où les observations sont collectées par un réseau de  $p$  capteurs. Tout comme dans le chapitre précédent, les signaux que nous voulons détecter sont supposés aléatoires de distributions inconnues, en présence de bruit Gaussien additif en plus d'interférence de distribution également inconnue. Le choix de telles hypothèses peut être très intéressant en pratique car il est moins contraignant que de supposer le signal d'intérêt de distribution connue.

#### Énoncé du problème:

Nous considérons  $p$  capteurs, situés dans une région d'intérêt, et un centre de fusion (FC) chargé de la centralisation des données pour la prise de décision. Nous voulons que les capteurs collaborent pour décider si une source émet un signal ou non. Chaque capteur recueille un vecteur observation  $\mathbf{Y}_i$  qui est, sous l'hypothèse alternative, la somme d'un signal  $\mathbf{S}_i$  émanant d'une source, un bruit Gaussien additif  $\mathbf{X}_i$ , et de l'interférence mal connus  $\boldsymbol{\xi}_i$ . Sous l'hypothèse nulle, le vecteur d'observation  $\mathbf{Y}_i$  de chaque capteur ne contient que le bruit additif  $\mathbf{X}_i$  et l'interférence  $\boldsymbol{\xi}_i$ . Après avoir collecté un vecteur observation par chaque capteur et l'avoir envoyé au (FC) non traité ou prétraité, une décision doit être prise et le centre de fusion choisit quelle hypothèse est vraie c'est-à-dire si une source émet ou non. Le vecteur observation reçu par le  $i^{\text{ème}}$  capteur ( $i = 1, 2, \dots, p$ ) noté  $\mathbf{Y}_i$  est un élément de  $\mathcal{M}(\Omega, \mathbb{R}^N)$ ; l'ensemble des vecteurs aléatoires réels de dimensions  $N$ . En conséquence, le modèle d'observation est:

$$\mathbf{Y}_i = \epsilon \mathbf{S}_i + \boldsymbol{\xi}_i + \mathbf{X}_i \text{ avec } \begin{cases} h_0 : \epsilon = 0 \\ h_1 : \epsilon = 1 \end{cases}$$

où  $\epsilon$  est une variable aléatoire de Bernoulli telle que  $\epsilon \in \{0, 1\}$ , supposée indépendante de  $\mathbf{S}_i, \boldsymbol{\xi}_i$  et  $\mathbf{X}_i$ . Elle symbolise la présence ou l'absence du signal source. Le signal  $\mathbf{S}_i$  émis par une source, l'interférence  $\boldsymbol{\xi}_i$  et le bruit  $\mathbf{X}_i$  sont tous des éléments de  $\mathcal{M}(\Omega, \mathbb{R}^N)$ ,  $\forall i \in \{1, 2, \dots, p\}$ . Nous supposons que le bruit  $\mathbf{X}_i$  et le signal  $\mathbf{S}_i$  sont indépendants et que  $\mathbf{X}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$  où  $\boldsymbol{\Sigma}_i$  est une matrice de covariance définie positive de dimension  $N \times N$ . Le signal  $\mathbf{S}_i$  ainsi que l'interférence  $\boldsymbol{\xi}_i$  sont supposés aléatoires de distribution inconnue.

Les capteurs sont supposés opérer une transformation linéaire sur les données

pour effectuer une réduction dimensionnelle. Chaque transformation de capteur est modélisée par une matrice de rang complet  $\mathbf{A}_i \in \mathbb{R}^{n \times N}$  avec  $n < N$ . En pratique,  $\mathbf{A}_i$  peut être une matrice de compression, une matrice de projection pour la détection de sous-espace etc. Nous notons  $\mathbf{Z}_i \in \mathcal{M}(\Omega, \mathbb{R}^n)$  l'observation après transformation à chaque capteur  $i$  et nous avons:

$$\mathbf{Z}_i = \mathbf{A}_i \mathbf{Y}_i = \epsilon \mathbf{A}_i \mathbf{S}_i + \mathbf{A}_i \boldsymbol{\xi}_i + \mathbf{W}_i \quad (12)$$

avec  $\mathbf{W}_i \sim \mathcal{N}(0, \mathbf{C}_i)$  et  $\mathbf{C}_i = \mathbf{A}_i \boldsymbol{\Sigma}_i \mathbf{A}_i^T$  la matrice de covariance définie positive du bruit après transformation. La méconnaissance de la distribution du signal d'intérêt et de l'interférence est bornée par les deux propriétés suivantes concernant le signal et l'interférence, en les supposant presque sûrement (p.s.) vraies:

$$\begin{cases} \mathcal{P}_0 : \forall i \quad \nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\xi}_i) \leq \tau \quad \forall i & (p.s.) \\ \mathcal{P}_1 : \forall i \quad \nu_{\mathbf{C}_i}(\mathbf{A}_i \mathbf{S}_i + \mathbf{A}_i \boldsymbol{\xi}_i) > \tau' \geq \tau \quad \forall i & (p.s.) \end{cases} \quad (13)$$

ce qui signifie que la probabilité que  $\mathcal{P}_0$  et  $\mathcal{P}_1$  se produisent est presque sûrement égale à 1.  $\tau \in [0, \infty)$  la tolérance est introduite pour modéliser l'effet de l'interférence qui est peu connu. De même, le paramètre  $\tau' \in [\tau, \infty)$  est introduit pour modéliser la seule connaissance disponible sur le signal d'intérêt  $\mathbf{S}_i$ . La norme de Mahalanobis  $\nu_{\mathbf{C}_i}$ , permet de mesurer une distance tout en tenant compte de la matrice de covariance du bruit  $\mathbf{C}_i$ . L'introduction des propriétés  $\mathcal{P}_0$  et  $\mathcal{P}_1$  est un moyen de borner le manque de connaissances sur la distribution du signal et de l'interférence. Dans cet esprit, les deux propriétés introduisent juste assez d'informations pour nous permettre de résoudre le problème grâce au matériel théorique du RDT.

Nous définissons le signal  $\boldsymbol{\Theta}_i \in \mathcal{M}(\Omega, \mathbb{R}^N)$  comme la somme de l'interférence  $\boldsymbol{\xi}_i$  perçue par le capteur  $i^{\text{ème}}$  et un signal d'intérêt  $\mathbf{S}_i \in \mathcal{M}(\Omega, \mathbb{R}^N)$  si présent:

$$\boldsymbol{\Theta}_i = \epsilon \mathbf{S}_i + \boldsymbol{\xi}_i \quad (14)$$

Lorsque nous supposons  $\mathcal{P}_0$  et  $\mathcal{P}_1$  vraies, presque sûrement, nous avons prouvé que le problème de décision, au niveau de chaque capteur, peut être reformulé tel que:

$$\begin{cases} \text{Observation} : \mathbf{Z}_i = \mathbf{A}_i \boldsymbol{\Theta}_i + \mathbf{W}_i \\ \mathcal{H}_0 : \nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\Theta}_i) \leq \tau \quad \forall i \\ \mathcal{H}_1 : \nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\Theta}_i) > \tau' \geq \tau \quad \forall i \end{cases} \quad (15)$$

### Un test optimal pour la configuration centralisée:

L'envoi de données brutes au centre de fusion, c'est-à-dire l'observation telle qu'elle est reçue par les capteurs, est ce que nous appelons une configuration centralisée [Varshney, 2012]. Elle est connue pour être optimale et est généralement utilisée pour éviter la perte d'informations, même si elle nécessite une forte transmission de données, contrairement aux configurations distribuées où le trafic de données est plus léger. Pour les configurations distribuées, lorsqu'un centre de fusion est disponible, les données sont prétraitées à l'échelle locale avant d'être envoyées au centre de fusion où la décision est prise. Par ailleurs, quand aucun centre de fusion n'est utilisé, les capteurs prennent

leur décision directement ou après s'être consultés. Quant à nous, nous considérons un réseau de capteurs  $p$  avec un centre de fusion. Nous étudions le cas où toutes les observations  $\mathbf{Z}_i$  sont envoyées au centre de fusion afin que la statistique de test puisse être calculée sans perte de performance. Ensuite, nous examinerons la configuration distribuée pour éviter des coûts de communication élevés. Les propriétés  $\mathcal{P}_0$  et  $\mathcal{P}_1$  qui limitent notre méconnaissance autour du signal et de l'interférence sont reformulées au niveau du centre de fusion tel que:

$$\begin{cases} \mathcal{P}_0^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A}\boldsymbol{\xi}) \leq p\tau & (p.s.) \\ \mathcal{P}_1^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi}) > p\tau' \geq p\tau & (p.s.) \end{cases} \quad (16)$$

avec  $\mathbf{C} = \text{blkdiag}(\mathbf{C}_1, \dots, \mathbf{C}_p) \in \mathbb{R}^{np \times np}$  et  $\mathbf{C}_i = \mathbf{A}_i \boldsymbol{\Sigma}_i \mathbf{A}_i^T$ , ainsi que:

$$\begin{cases} \mathbf{S} = [\mathbf{S}_1^T, \dots, \mathbf{S}_p^T]^T, \quad \mathbf{S} \in \mathbb{R}^{Np} \\ \boldsymbol{\xi} = [\boldsymbol{\xi}_1^T, \dots, \boldsymbol{\xi}_p^T]^T, \quad \boldsymbol{\xi} \in \mathbb{R}^{Np} \\ \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{A}_p \end{pmatrix}, \quad \mathbf{A} \in \mathbb{R}^{np \times Np} \end{cases} \quad (17)$$

Nous démontrons que, en supposant vraies, presque surement, les propriétés  $\mathcal{P}_1^{\text{FC}}$  et  $\mathcal{P}_0^{\text{FC}}$ , le problème de décision peut être présenté, du point de vue de centre de fusion, comme étant:

$$\begin{cases} \text{Observation} : \mathbf{Z} = \mathbf{A}\boldsymbol{\Theta} + \mathbf{W} \\ \mathcal{H}_0^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A}\boldsymbol{\Theta}) \leq p\tau \\ \mathcal{H}_1^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A}\boldsymbol{\Theta}) > p\tau' \geq p\tau \end{cases} \quad (18)$$

avec:

$$\begin{cases} \mathbf{Z} = [\mathbf{Z}_1^T, \dots, \mathbf{Z}_p^T]^T, \quad \mathbf{Z} \in \mathbb{R}^{np} \\ \mathbf{W} = [\mathbf{W}_1^T, \dots, \mathbf{W}_p^T]^T, \quad \mathbf{W} \in \mathbb{R}^{np} \\ \boldsymbol{\Theta} = [\boldsymbol{\Theta}_1^T, \dots, \boldsymbol{\Theta}_p^T]^T, \quad \boldsymbol{\Theta} \in \mathbb{R}^{Np} \end{cases} \quad (19)$$

Selon le théorème principale du framework RDT, le test optimal  $\mathcal{T}_{FC} : \mathbb{R}^{np} \rightarrow \{0, 1\}$  selon le framework RDT, pour la configuration centralisé et le problème de décision présenté dans (18), est défini pour chaque  $\mathbf{z} \in \mathbb{R}^{np}$  par:

$$\mathcal{T}_{FC}(\mathbf{z}) = \begin{cases} 1 & \text{if } \nu_{\mathbf{C}}(\mathbf{z}) > \lambda_{FC} \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Le seuil  $\lambda_{FC}$  est la solution unique dans  $\eta$  à l'équation:  $Q_{np/2}(p\tau, \eta) = \gamma$ , où  $\gamma \in (0, 1)$  est le niveau du test.  $Q_{np/2}$  est la fonction de Marcum généralisée définie pour toute paire  $(a, b) \in [0, \infty) \times [0, \infty)$  telle que  $Q_{np/2}(a, b) = 1 - \mathbb{F}_{\mathcal{X}_{np}^2(a^2)}(b^2)$  où  $\mathbb{F}_{\mathcal{X}_{np}^2(a^2)}$  est la fonction de répartition d'une distribution du Chi2 non centrée:  $\mathcal{X}_{np}^2(a^2)$  avec  $np$  degré de liberté et  $a^2$  le paramètre de non-centralité.

### Un test optimal pour la configuration distribuée:

La seconde configuration que nous considérons dans ce problème est une configuration distribuée avec un centre de fusion. Ainsi, un centre de fusion est en communication avec  $p$  capteurs. Le but est de minimiser la quantité d'informations échangées entre le centre de fusion et les capteurs. C'est pourquoi, nous avons l'intention d'envoyer au centre de fusion le minimum vital, en termes de quantité de données: un scalaire. Ce

scalaire doit contenir autant d'informations que possible ou au moins réduire au minimum les pertes de performances. Nous continuons à supposer que les deux propriétés  $\mathcal{P}_0^{\text{FC}}$  et  $\mathcal{P}_1^{\text{FC}}$  sont presque sûrement vraies. Nous considérons que nous voulons résoudre le même problème de décision (18) que pour la configuration centralisée, cependant avec une contrainte additive sur la communication entre le centre de fusion et les capteurs.

Le problème, tel qu'il est exposé, est très différent des approches classiques, où les signaux à détecter sont supposés déterministes avec des paramètres inconnus qui sont estimés ou attribués a priori. En effet, dans cette approche, aucune hypothèse sur les signaux ni la distribution des interférences n'est faite. Cela étend le problème de détection à la classe substantielle de tous les vecteurs aléatoires de distribution inconnue et une énergie limitée.

L'étape suivante consiste à trouver un test optimal selon un certain critère. À notre connaissance, avec ces hypothèses, seul le cadre du RDT peut fournir un test optimal. C'est dans ce contexte, que notre contribution autour de ce quatrième chapitre apparaît. Nous montrons que l'envoi de la statistique RDT locale (qui est un scalaire), au centre de fusion, non seulement réduit le coût de communication, mais garantit également les mêmes performances que la configuration centralisée. Plus précisément, nous montrons que l'envoi des observations brutes au centre de fusion qui dérive ensuite la statistique de test globale, sur la base de ces observations, équivaut à envoyer les statistiques de test locales calculées individuellement par chaque capteur au centre de fusion où seule une addition de ces statistiques est opérée. Ceci est formalisé à travers la proposition suivante:

**Proposition.** *Si  $\mathcal{T}_{\text{FCD}} : [0, \infty)^p \rightarrow \{0, 1\}$ , est défini pour tout  $p$ -uplets  $(x_1, \dots, x_p)$  de réels non négatifs, comme étant:*

$$\mathcal{T}_{\text{FCD}}(x_1, \dots, x_p) = \begin{cases} 1 & \text{si } \sum_{i=1}^p x_i > \lambda_{\text{FC}} \\ 0 & \text{sinon} \end{cases} \quad (21)$$

alors, avec la même notation que ci-dessus:

$$\mathcal{T}_{\text{FC}}(\mathbf{Z}) = \mathcal{T}_{\text{FCD}}(\nu_{\mathbf{C}_1}(\mathbf{Z}_1), \dots, \nu_{\mathbf{C}_p}(\mathbf{Z}_p)) \quad (22)$$

Ainsi le test optimal selon le critère d'optimalité RDT pour la configuration distribuée est le même test que pour la configuration centralisée. La seule différence concerne la quantité d'information envoyée par les capteurs vers le centre de fusion.

### Conclusion:

Dans ce chapitre, nous voulions élargir la portée du framework RDT et l'étendre à un contexte multi-capteurs pour tester ses limites. Il s'est avéré que le cadre RDT s'adapte naturellement à ce contexte et donne la possibilité de réduire les coûts de communication. En effet, grâce à son extensibilité, le framework RDT s'adapte à la fois à une configuration centralisée ainsi qu'à une configuration distribuée avec centre de fusion. Ainsi un test optimal selon le critère RDT a été dérivé pour les deux configurations. Ce test a été démontré être le même pour les deux configurations. Cependant, la quantité d'information échangé entre les capteurs et le centre de fusion a été diminué d'un facteur de  $n$  en passant d'une configuration centralisée à une configuration distribuée avec centre de fusion. Nous avons aussi démontré qu'en termes de performance, plus

spécifiquement, la probabilité de détection pour un niveau de fausse alarme fixé, les deux tests étaient équivalents. Pour conclure, l'utilisation du framework RDT permet la résolution de problèmes de décision, classiquement insolubles avec des approches standard comme la méthode de Neyman-Pearson, l'approche bayésienne, etc.

## Chapitre 5: RDT généralisé

Notre deuxième contribution est une généralisation du framework RDT ou test de distorsion aléatoire [Pastor and Nguyen, 2013a] présenté dans la partie état de l'art. En raison de l'hypothèse concernant la distribution du signal d'intérêt supposée inconnue, le framework RDT peut s'adapter à un large éventail de scénarios de détection. Néanmoins, le problème que nous considérons ici, que nous appelons test de distorsion aléatoire généralisé (GRDT), et qui permet une généralisation du RDT, contient une hypothèse contraignante en moins, à savoir sur le bruit supposé de distribution quelconque pas nécessairement Gaussien. De plus, la norme de Mahalanobis dans l'énoncé du problème, est remplacée par n'importe quel maximal invariant  $\mathcal{M}(\cdot)$  associé au groupe de transformation  $\mathcal{G}$  qui laisse le problème invariant. Notre contribution apparaît à travers un théorème qui donne les conditions pour qu'un test soit optimal selon un critère analogue à celui du RDT, dans le cadre du problème GRDT.

### Énoncé du problème

Soit  $\mathbf{Y}$ ,  $\Theta$ ,  $\mathbf{X}$  trois vecteurs aléatoires réels de dimension  $N$ . Nous supposons que le signal  $\Theta$  est aléatoire de distribution inconnue. Le bruit  $\mathbf{X}$  admet une fonction de densité de probabilité  $f_{\mathbf{X}}$  connue, et l'observation  $\mathbf{Y}$  est telle que  $\mathbf{Y} = \Theta + \mathbf{X}$ . Nous notons  $\mathcal{M}(\cdot)$ , un maximal invariant lié à un groupe de transformation  $\mathcal{G}$  qui laisse le problème invariant tel que  $\theta \in \mathbb{R}^N \mapsto \mathcal{M}(\theta) \in \mathbb{R}$ . Un paramètre  $\tau \geq 0$  est appelé tolérance, de manière analogue au framework RDT. le problème GRDT consiste à décider si l'événement  $[\mathcal{M}(\Theta) \leq \tau]$  ou  $[\mathcal{M}(\Theta) > \tau]$  est vérifié et peut être présenté comme suit:

$$\text{GRDT: } \begin{cases} \text{Observation : } \mathbf{Y} = \Theta + \mathbf{X} & \begin{cases} \Theta \text{ et } \mathbf{X} \text{ indépendants,} \\ \mathbf{X} \sim f_{\mathbf{X}}, \end{cases} \\ \text{Événement nul: } [\mathcal{M}(\Theta) \leq \tau], \\ \text{Événement alternatif: } [\mathcal{M}(\Theta) > \tau]. \end{cases} \quad (23)$$

L'utilisation du maximal invariant  $\mathcal{M}(\cdot)$  peut paraître à première vue trop abstrait. Néanmoins, cette formulation du problème peut se révéler très concrète. En effet,  $\mathcal{M}(\theta)$  peut représenter par exemple le SNR;  $\mathcal{M}(\theta) = \frac{\|\theta\|^2}{\|\mathbf{X}\|^2}$ , l'énergie du signal d'intérêt  $\mathcal{M}(\theta) = \|\theta\|^2$ , sa projection dans un sous-espace, etc. En considérant un maximal invariant, ceci sous-entend qu'il existe un groupe de transformation  $\mathcal{G}$  qui laisse le problème GRDT invariant. Ainsi la notion d'orbite peut être introduite de manière naturelle. Les orbites de  $\mathcal{G}$  peuvent être définies comme  $\Upsilon_{\rho} = \{\mathbf{y} \in \mathbb{R}^N : \mathcal{M}(\mathbf{y}) = \rho\}$  avec  $\rho \in \mathbb{R}$ . La tolérance  $\tau$ , permet de modéliser le manque de connaissances autour du signal d'intérêt en limitant les incertitudes.

### Un test optimal

Pour résoudre un tel problème tout en visant l'optimalité, les approches standard de la littérature, comme celles citées dans la première partie, ne peuvent malheureusement pas être envisagées et ce en raison du manque d'informations sur la distribution du signal. Ci-dessous, nous utiliserons le critère d'optimalité  $\gamma$ -MCCP (Maximal Constant



Conditional Power) introduit par le cadre RDT, car il est très approprié dans le cas de signaux aléatoires de distribution inconnue. Cependant, nous l'adaptions au problème GRDT étant donné le bruit  $\mathbf{X}$  quelconque et tout maximal invariant associé  $\mathcal{M}(\cdot)$ .

**Définition.** Soit  $\tau \geq 0$  et  $\gamma \in (0, 1)$ , un test  $\mathcal{T}^*$  est  $\gamma$ -MCCP si:

- **[Level]**  $\mathcal{T}^*$  est de niveau  $\gamma$  (i.e. de probabilité de fausse alarme inférieure ou égale à  $\gamma$ ).
- **[MCCP]** Pour tout  $\Theta$  et pour  $\mathbb{P}_{\mathcal{M}(\Theta)}$ -presque tout  $\rho > \tau$ , le test  $\mathcal{T}^*$  a une fonction de puissance conditionnelle constante pour tout  $\Theta \in \Upsilon_\rho$  et pour tout test  $\mathcal{T}$  de niveau  $\gamma$  avec une fonction de puissance conditionnelle constante sur  $\Upsilon_\rho$ , nous avons:

$$\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \mathcal{M}(\Theta) = \rho] \leq \mathbb{P}[\mathcal{T}^*(\Theta + \mathbf{X}) = 1 \mid \mathcal{M}(\Theta) = \rho]$$

Sur la base de cette définition, qui introduit le critère d'optimalité typique du RDT dans le cadre du problème GRDT, le théorème suivant présente les conditions suffisantes pour qu'un test  $\mathcal{T}$  soit  $\gamma$ -MCCP pour le problème GRDT.

**Théorème.** Soit  $\mathcal{T}^*$  un test de niveau  $\gamma$  pour le problème GRDT tel que pour tout  $\Theta$ ,  $\mathcal{T}^*$  a une fonction de puissance conditionnelle constante pour  $\Theta \in \Upsilon_\rho$  et pour  $\mathbb{P}_{\mathcal{M}(\Theta)}$ -presque tout  $\rho > \tau$ . Si, pour tout  $\rho > \tau$ , il existe deux variables aléatoires  $\Theta_0$  et  $\Theta_1$ , avec  $\Theta_0 \in \Upsilon_\tau$  et  $\Theta_1 \in \Upsilon_\rho$  presque sûrement, tel que le test  $\mathcal{T}^*$  est plus puissant de niveau  $\gamma$  pour tester:

$$\begin{cases} \mathcal{H}_0: & \mathbf{Y} = \Theta_0 + \mathbf{X} \\ \mathcal{H}_1: & \mathbf{Y} = \Theta_1 + \mathbf{X} \end{cases} \quad (24)$$

alors le test  $\mathcal{T}^*$  est  $\gamma$ -MCCP pour le problème GRDT.

## Conclusion

Ce résultat nous a conduit à la question suivante: existe-t-il d'autres paires de bruit/maximal invariant, autres que le bruit Gaussien et la norme de Mahalanobis, pour lesquels le test  $\gamma$ -MCCP existe. Malheureusement, nous n'avons toujours pas la réponse à cette question. Néanmoins, cela nous a mis sur la piste intéressante du chapitre suivant où nous avons considéré des aspects asymptotiques de l'optimalité lorsque les critères d'optimalité du genre  $\gamma$ -MCCP ne sont pas possibles. L'approche asymptotique que nous considérons est basée sur le critère d'optimalité UPP (Uniformément Plus Puissant). Comme nous le verrons dans la suite, les tests Asymptotiquement Uniformément Plus Puissants (AUPP) seront présentés et appliqués à un exemple de problème de décision.

## Chapitre 6: Approche asymptotique: Le théorème de Karlin-Rubin Asymptotique

Dans ce chapitre, le but était de généraliser le framework RDT, présenté dans le troisième chapitre, dans une perspective asymptotique, compte tenu des limites du GRDT présenté dans le chapitre précédent. Dans un premier lieu, nous voulions investiguer l'aspect asymptotique, dans le cas des modèles inconditionnels, pour ensuite

passer aux modèles conditionnels avec l'approche RDT. Cependant, à cause du manque de temps nous avons travaillé sur l'aspect asymptotique uniquement dans le cas des modèles inconditionnels. Ainsi, nous avons démontré une formulation asymptotique du théorème de Karlin-Rubin pour des tests Asymptotiquement Uniformément Plus Puissants (AUPP). Ensuite, nous l'appliquons, conjugué au principe d'invariance, dans le cadre de la résolution d'un problème de décision spécifique.

### Théorème de Karlin-Rubin Asymptotique

**Théorème.** Soit  $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^N$  un vecteur aléatoire absolument continu de pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$ , où  $\theta \in \Theta \subset \mathbb{R}$ , et soit  $\Theta_1$  et  $\Theta_0$  deux sous-ensembles de  $\Theta$  tel que:  $\Theta = \Theta_0 \cup \Theta_1$  et  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Soit:

$$\Lambda_{\theta_0, \theta_1}(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} \mathbb{1}_{[f_{\mathbf{Y}}^{\theta_0} f_{\mathbf{Y}}^{\theta_1} \neq 0]} + \infty \mathbb{1}_{[f_{\mathbf{Y}}^{\theta_0} = 0] \cap [f_{\mathbf{Y}}^{\theta_1} \neq 0]} \quad (25)$$

le rapport de vraisemblance tel que:  $\Lambda_{\theta_0, \theta_1} = h_{\theta_0, \theta_1}(V)$ , avec  $V : \mathbb{R}^N \mapsto \mathbb{R}$ .

Soit  $(\mathbf{Y}_n)_{n \in \mathbb{N}^*}$  be a sequence of random vectors where every  $\mathbf{Y}_n$  a une pdf  $f_{\mathbf{Y}_n}^{\theta}(\mathbf{y})$ .

Si pour tout  $\theta_0 < \theta_1$ ,  $h_{\theta_0, \theta_1}$  est strictement croissant en  $V$  et  $\mathbf{Y}_n \xrightarrow{\mathcal{L}} \mathbf{Y}$ , alors le test

$$\mathcal{T}'(\mathbf{y}) = \begin{cases} 1 & \text{si } V(\mathbf{y}) \geq \lambda \\ 0 & \text{sinon,} \end{cases} \quad (26)$$

est AUPP parmi tous les tests appartenant à  $\Psi_{\mathbf{Y}}$  pour tester

$$\begin{cases} \mathcal{H}'_0 : \theta \leq \theta_0 \\ \mathcal{H}'_1 : \theta > \theta_0 \end{cases} \quad (27)$$

Ce théorème fournit l'expression du test Asymptotiquement Uniformément Plus Puissant: (AUPP) pour un problème de test d'hypothèses composite, ainsi que les conditions que le problème doit respecter pour que ce test existe.

**Application: Tester la présence d'un signal déterministe inconnu dans un sous-espace cône**

#### • Énoncé du problème

Soit  $\boldsymbol{\theta} \in \mathbb{R}^N$  un signal déterministe inconnu observé dans un bruit additif Gaussien  $\mathbf{X} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$  de variance inconnue  $\sigma^2$ . Le SNR est supposé connu. Le vecteur observation est modélisé:

$$\mathbf{Y} = \boldsymbol{\theta} + \mathbf{X}. \quad (28)$$

Soit  $\mathbf{H}$  une matrice  $N \times n$  qui génère un sous-espace  $\langle \mathbf{H} \rangle$  de rang  $n$  avec  $n \leq N$  et  $\mathbf{P}_{\mathbf{H}}$  sa matrice de projection correspondante, i.e.  $\mathbf{P}_{\mathbf{H}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ . Étant donné une tolérance  $0 \leq \tau \leq 1$ , soit  $\mathcal{C}_{\mathbf{H}}$  un sous-espace cône défini comme  $\mathcal{C}_{\mathbf{H}} = \left\{ \mathbf{y} \in \mathbb{R}^N : \rho = \frac{|\mathbf{P}_{\mathbf{H}} \mathbf{y}|^2}{|\mathbf{y}|^2} > \tau \right\}$ .

Le problème de décision peut donc être formulé:

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{\theta} \notin \mathcal{C}_H \\ \mathcal{H}_1 : \boldsymbol{\theta} \in \mathcal{C}_H. \end{cases} \quad (29)$$

Le problème équivaut à vouloir tester si  $\boldsymbol{\theta}$  appartient au cône  $\mathcal{C}_H$  ou non. Plus explicitement, le but est de tester si une proportion suffisante de l'énergie du signal se trouve dans le sous-espace  $\langle \mathbf{H} \rangle$ . Une telle formulation du problème peut être pertinente lorsque le signal à détecter obéit au modèle de sous-espace linéaire et lorsqu'il est corrompu par une interférence peu connue ou difficile à caractériser  $\boldsymbol{\xi}$ , c'est-à-dire  $\boldsymbol{\theta} = \mu \mathbf{H} \boldsymbol{\phi} + \boldsymbol{\xi}$ , où  $\mu$  est une variable aléatoire dans  $\{0, 1\}$  modélisant la présence ou l'absence possible du signal, et où  $\boldsymbol{\phi} \in \mathbb{R}^n$  est le vecteur coordonnées inconnu de ce signal dans le sous-espace  $\langle \mathbf{H} \rangle$ .  $\boldsymbol{\xi}$  peut être défini comme un vecteur dont l'énergie se situe principalement en dehors du sous-espace dans lequel réside le signal d'intérêt, c'est-à-dire  $\|\mathbf{P}_H \boldsymbol{\xi}\|^2 \leq \tau \|\boldsymbol{\xi}\|^2$ . Ce problème de décision binaire peut être pertinent par exemple dans les applications de type apprentissage automatique où le but est de vérifier sur un ensemble de données si un modèle linéaire supposé correspond aux données analysées avec une précision suffisante  $\tau$ .

- **Invariances du problème**

Une manière de résoudre ce problème de décision est de dériver un test qui soit Uniformément Plus Puissant (UPP). En effet, garantir qu'un test a la plus grande puissance (probabilité de détection) parmi la classe de tests de même niveau est un critère d'optimalité fort. Cependant, comme  $\boldsymbol{\theta}$  n'est pas parfaitement connu, aucun test UPP ne peut être calculé. Dans ce genre de cas, notre attention se limite aux tests invariants par rapport à un groupe de transformation. Un groupe de transformations qui laisse le problème invariant est:

$$\mathcal{G} = \{g : g(\mathbf{y}) = \kappa(\mathbf{U}_H \mathbf{Q} \mathbf{U}_H^T + \mathbf{U}_{H^\perp} \mathbf{R} \mathbf{U}_{H^\perp}^T) \mathbf{y}\}, \quad (30)$$

où  $\kappa \in \mathbb{R}$ ,  $\mathbf{Q}$  et  $\mathbf{R}$  sont des matrices orthogonales de dimension  $n \times n$  et  $(N - n) \times (N - n)$  respectivement.  $\mathbf{U}_H$  est défini tel que  $\mathbf{P}_H = \mathbf{U}_H \mathbf{U}_H^T$  et  $\mathbf{U}_{H^\perp}$  tel que  $\mathbf{P}_H = \mathbf{I}_N - \mathbf{U}_{H^\perp} \mathbf{U}_{H^\perp}^T$ .

Étant donné l'invariance de notre problème, il est possible de trouver un test  $\mathcal{T}$  invariant qui soit fonction d'un maximal invariant. Un maximal invariant associé à  $\mathcal{G}$  est:

$$\mathbf{y} \in \mathbb{R}^N \mapsto M(\mathbf{y}) = \frac{N - n}{n} \times \frac{\|\mathbf{P}_H \mathbf{y}\|^2}{\|(\mathbf{I}_N - \mathbf{P}_H) \mathbf{y}\|^2} \in [0, \infty) \quad (31)$$

- **Un test optimal: AUPPI**

Si un test Uniformément Plus Puissant Invariant (UPPI) existe, sa statistique peut être obtenue en calculant le rapport de vraisemblance du maximal invariant  $M(\mathbf{y})$ . Puisque  $M(\mathbf{y})$  peut être exprimé comme un rapport de deux variables aléatoires de distribution Chi-2 non centrée, il suit une distribution de Fisher  $\mathcal{F}$  doublement non centrée [Johnson et al., 1995, Ch. 30]. Par conséquent le rapport de vraisemblance s'écrit:

$$L_M(\mathbf{y}) = \frac{\mathcal{F}(n, N - n, N\zeta\rho_1, N\zeta(1 - \rho_1), M(\mathbf{y}))}{\mathcal{F}(n, N - n, N\zeta\rho_0, N\zeta(1 - \rho_0), M(\mathbf{y}))} \quad (32)$$

où  $\rho_0 \leq \tau$  (resp.  $\rho_1 > \tau$ ) représente le pourcentage réel de l'énergie de  $\boldsymbol{\theta}$  dans le sous-espace  $\langle \mathbf{H} \rangle$  sous l'hypothèse  $\mathcal{H}_0$  (resp.  $\mathcal{H}_1$ ). Le paramètre  $\zeta$  représente le SNR. Le test UPPI peut être dérivé en invoquant le théorème de Karlin Rubin qui stipule que la comparaison d'une statistique de test (Le maximal invariant) à un seuil, conduit à un test UPPI si cette statistique a un rapport de vraisemblance strictement croissant [Scharf, 1991]. Malheureusement,  $L_M(\mathbf{y})$  n'est pas une fonction strictement croissante de  $M(\mathbf{y})$  pour chaque paire  $(\rho_1 > \rho_0, \rho_0)$  et tout ensemble de paramètres  $(n, N, \zeta)$ . Par conséquent, le test UPPI n'existe pas. Dans ce genre de scénarios, le test " sous-optimal " de rapport de vraisemblance généralisé (GLRT) devient très souvent l'option choisie bien qu'il ne garantisse pas toujours l'optimalité. C'est à ce stade que notre résultat théorique; "Théorème de Karlin-Rubin asymptotique" peut trouver toute son utilité. En effet, nous pouvons dériver un test qui soit asymptotiquement UPPI quand  $N$  est grand. Ainsi, d'après le théorème de Karlin-Rubin asymptotique, un test AUPPI de taille  $\alpha$  (probabilité de fausse alarme) pour notre problème de décision binaire est le test  $\mathcal{T}_A$ , pour  $\alpha \in (0, 1)$ , défini pour tout  $\mathbf{y} \in \mathbb{R}^N$ :

$$\mathcal{T}_A(\mathbf{y}) = \begin{cases} 1 & \text{si } M(\mathbf{y}) \geq \lambda_A \\ 0 & \text{sinon} \end{cases} \quad (33)$$

avec  $\lambda_A$  tel que

$$F(n, N - n, N\zeta\tau, N\zeta(1 - \tau), \lambda_A) = 1 - \alpha. \quad (34)$$

## Conclusion

Bien que l'approche de ce chapitre soit fortement influencée par les obstacles que nous avons rencontré sur notre chemin dans la généralisation du framework RDT, elle a fini par nous conduire à des résultats asymptotiques très intéressants dans le cas des modèles inconditionnels. En effet, le champ des possibilités d'utilisation du théorème de Karlin-Rubin asymptotique est large et de nombreuses applications pourraient en bénéficier, comme par exemple la détection en milieu marin [Socheleau et al., 2015], [Leroy et al., 2016].

## Conclusion générale

Le sujet de ma thèse portait sur la détection des signaux aléatoires lorsque leur distribution est inconnue, en exploitant les invariances inhérentes au problème et en ciblant l'optimalité. La première partie présente les différentes approches de l'état de l'art pour la détection de signaux suivant le modèle inconditionnel ainsi que le modèle conditionnel. La notion d'invariance a aussi été mise en lumière compte tenu de son importance dans la procédure de résolution et dans les pistes emmprintées.

La deuxième moitié du manuscrit, contient nos principales contributions. Le chapitre 4 examine l'adéquation du framework RDT à un contexte distribué, c'est-à-dire lorsque les vecteurs d'observation ne sont pas collectés par un capteur mais par un réseau de capteurs. Le scénario de détection considéré comporte un signal stochastique de distribution inconnue, en présence d'interférence inconnue et d'un bruit Gaussien additif. L'objectif étant de décider, sur la base de plusieurs vecteurs observation, si le

signal d'intérêt est présent. Un tel problème avec si peu d'informations sur la distribution du signal d'intérêt et de l'interférence est insoluble selon les approches standard lorsque l'optimalité est recherchée. Cependant, un test optimal existe au sens du critère d'optimalité du RDT. En effet, en limitant le manque de connaissance à la fois du signal d'intérêt et de l'interférence, l'optimalité peut être atteinte. Le résultat le plus intéressant de ce chapitre concerne l'équivalence de performance en termes de probabilité de détection entre une configuration centralisée et une configuration distribuée. De plus, une baisse de communication peut être obtenu quand la configuration utilisée est distribuée plutôt que centralisée, sans aucune perte de performance. Ce travail a été réalisé en collaboration avec Prashant Khanduri et Pramod K. Varshney du département de génie électrique et informatique de l'Université de Syracuse, New York.

Dans le chapitre 5, notre but était de généraliser le framework RDT aux problèmes où le bruit n'est pas nécessairement Gaussien et où le maximal invariant associé n'est pas non plus la norme de Mahalanobis. Le signal d'intérêt est toujours supposé de distribution inconnue. À notre connaissance, un tel scénario de décision n'a pas été traité dans la littérature standard tout en visant l'optimalité. Nous avons considéré le même critère d'optimalité que celui du framework RDT. Ensuite, nous avons démontré un résultat qui donne les conditions suffisantes pour qu'un test soit optimal, selon le critère d'optimalité RDT, pour le problème général GRDT. Cette approche est conforme à l'esprit du modèle conditionnel où, tout comme le framework RDT, la distribution inconnue du signal n'est ni recherchée ni estimée. Ce qui conduit à des critères d'optimalité non intégraux comme mentionné dans l'introduction. Nous avons travaillé sur ce chapitre en collaboration avec Guillaume Ansel du département signal et communication de IMT Atlantique Brest.

Enfin, le chapitre 6 traite de la notion d'optimalité dans un contexte asymptotique, pour le cas où le signal d'intérêt suit le modèle inconditionnel. Ainsi, les tests AUPP sont définis, puis une formulation asymptotique du théorème de Karlin-Rubin est démontrée. Ce théorème est utilisé dans un problème de décision pour tester la présence d'un signal déterministe inconnu dans un sous-espace cône. De façon similaire à l'approche RDT, le problème de décision a été formulé en incorporant une tolérance pour limiter les incertitudes sur le signal. Cette application a mis en évidence l'utilité de l'approche asymptotique pour certains problèmes de décision, lorsqu'une optimalité conventionnelle n'est pas atteignable.

# Contents

<b>Acknowledgements</b>	<b>3</b>
<b>Abstract</b>	<b>5</b>
<b>Résumé long</b>	<b>7</b>
<b>Contents</b>	<b>29</b>
<b>Acronyms</b>	<b>31</b>
<b>Notations</b>	<b>33</b>
<b>Introduction</b>	<b>35</b>
<b>1 The unconditional approaches in detection theory</b>	<b>39</b>
1.1 Hypothesis testing problems . . . . .	39
1.1.1 Specificities of the hypotheses . . . . .	40
1.1.2 Statistical tests . . . . .	40
1.1.3 Optimality and robustness . . . . .	42
1.2 Optimality oriented approaches for the unconditional model . . . . .	44
1.2.1 The minimum probability of error test . . . . .	44
1.2.2 The Bayes risk . . . . .	45
1.2.3 Uniformly most powerful tests . . . . .	46
1.2.4 The holy trinity . . . . .	48
1.2.5 The Bayesian approach . . . . .	51
1.2.6 Wald's UBCP tests . . . . .	52
<b>2 Invariance in detection theory</b>	<b>55</b>
2.1 Invariant detection problems . . . . .	56
2.1.1 The group of transformation . . . . .	56
2.1.2 Invariant hypothesis-testing problems . . . . .	58
2.2 Maximal invariant and orbits . . . . .	59
2.3 Invariant tests . . . . .	60
2.4 Limitations of the standard invariance . . . . .	61
2.5 A more general formulation of invariance . . . . .	62

<b>3</b>	<b>A conditional approach: Random Distortion Testing framework</b>	<b>65</b>
3.1	The Deterministic Distortion Testing problem . . . . .	66
3.1.1	Problem statement . . . . .	66
3.1.2	Main theoretical results . . . . .	67
3.2	The Random Distortion Testing problem . . . . .	69
3.2.1	Problem statement . . . . .	69
3.2.2	Main results . . . . .	69
3.3	Conclusions . . . . .	71
<b>4</b>	<b>Distributed Random Distortion Testing</b>	<b>73</b>
4.1	Problem statement . . . . .	73
4.2	Reformulation of the hypothesis testing problem . . . . .	74
4.3	Multi-sensors configurations . . . . .	75
4.3.1	The centralised configuration . . . . .	75
4.3.2	The distributed configuration . . . . .	76
4.4	RDT for multi-sensors configurations . . . . .	77
4.4.1	Theoretical material . . . . .	77
4.4.2	Optimality for centralised configuration . . . . .	79
4.4.3	Optimality for the distributed configuration . . . . .	81
4.5	Numerical results . . . . .	82
4.5.1	Lower bounds on detection probabilities . . . . .	82
4.5.2	The constraint of level for the false alarm probability . . . . .	84
4.6	Conclusion . . . . .	84
<b>5</b>	<b>General Random distortion Testing</b>	<b>87</b>
5.1	Problem statement . . . . .	87
5.2	Preliminary definitions . . . . .	88
5.2.1	Power function and size . . . . .	88
5.2.2	Conditional power function . . . . .	89
5.3	Theoretical results . . . . .	89
5.3.1	Preliminary results . . . . .	90
5.3.2	An optimal test . . . . .	90
5.4	Conclusion . . . . .	92
<b>6</b>	<b>An asymptotic approach: Asymptotic Karlin-Rubin's Theorem</b>	<b>93</b>
6.1	An asymptotic formulation of the Neyman-Pearson theorem . . . . .	93
6.2	An asymptotic formulation of the Karlin-Rubin theorem . . . . .	96
6.3	Testing the presence of a deterministic signal in a subspace cone . . . . .	97
6.3.1	Problem statement . . . . .	97
6.3.2	Invariance under group action . . . . .	98
6.3.3	Uniformly Most Powerful Invariant test . . . . .	98
6.3.4	Asymptotically Uniformly Most Powerful Invariant test . . . . .	99
6.3.5	Unknown SNR . . . . .	99
6.3.6	Numerical results . . . . .	101
6.3.7	Asymptotic optimality of the GLRT . . . . .	101
6.4	Conclusions . . . . .	104

---

<b>Conclusions and perspectives</b>	<b>105</b>
6.5 Conclusions about our contributions . . . . .	105
6.6 Perspectives . . . . .	106
<b>A Proof of equivalence between the initial problems (4.1) and (4.8), and the RDT formalised problems (4.5) and (4.11).</b>	<b>109</b>
A.1 Proof of Proposition 4.2.1 . . . . .	109
A.2 Proof of Proposition 4.3.1 . . . . .	111
<b>B Proof of the Asymptotic Neyman Pearson Theorem 6.1.1</b>	<b>113</b>
<b>C Proof of the Asymptotic Karlin Rubin Theorem 6.2.1</b>	<b>115</b>
<b>D Proof of Theorem 3.1.2</b>	<b>119</b>
<b>E Proof of lemma 3.2.1</b>	<b>123</b>
<b>F Proof of Theorem 3.2.2</b>	<b>125</b>
<b>G Proof of lemma 5.3.1</b>	<b>127</b>
<b>H Proof of Theorem 5.3.4</b>	<b>129</b>
<b>Bibliography</b>	<b>133</b>





# Acronyms

<b>AMP</b>	Asymptotically Most Powerful
<b>AUMP</b>	Asymptotically Uniformly Most Powerful
<b>AUMPI</b>	Asymptotically Uniformly Most Powerful Invariant
<b>DDT</b>	Deterministic Distortion Testing
<b>DRDT</b>	Distributed Random Distortion Testing
<b>FC</b>	Fusion Center
<b>GLRT</b>	Generalised Likelihood Ratio Test
<b>GRDT</b>	General Random Distortion Testing
<b>MCCP</b>	Maximal Constant Conditional Power
<b>MCP</b>	Maximal Constant Power
<b>MLE</b>	Maximum Likelihood Estimator
<b>MP</b>	Most Powerful
<b>MPE</b>	Minimum Probability of Error
<b>PDF</b>	Probability Density Function
<b>RDT</b>	Random Distortion Testing
<b>SNR</b>	Signal-to-Noise Ratio
<b>UBCP</b>	Uniformly Best Constant Power
<b>UMP</b>	Uniformly Most Powerful
<b>UMPI</b>	Uniformly Most Powerful Invariant



# Notations

$x$	scalar
$\mathbf{x}$	vector
$\mathbf{X}$	random variable
$\mathbb{R}$	set of real numbers
$\mathbb{N}$	set of natural numbers
$(\Omega, B, \mathbb{P})$	probability space where $B$ is a Borel subset of $\mathbb{R}^N$ and $\mathbb{P}$ a probability measure on $B$ .
$\mathcal{M}(\Omega, \mathbb{R}^N)$	set of $N$ -dimensional real random vectors defined on $(\Omega, B)$
$\mathbf{I}_B$	indicator function that equals to 1 $\forall x \in B$ and 0 otherwise
$\mathbf{A}^T$	transpose
$\ \mathbf{x}\ $	Euclidean norm of $\mathbf{x}$ such that $\ \mathbf{x}\  = \sqrt{\sum_i \mathbf{x}_i^2}$
$\nu_{\mathbf{C}}(\mathbf{x})$	Mahalanobis norm such that $\nu_{\mathbf{C}}(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}$
$\mathbb{E}[\mathbf{X}]$	mathematical expectation
$\mathcal{N}(\mu, \sigma^2)$	normal or Gaussian distribution with $\mu$ the mean and $\sigma^2$ the variance
$Q_{N/2}$	generalised Marcum function
$\chi_N^2(a^2)$	non-central chi-square distribution with $N$ degrees of freedom and non-centrality parameter $a^2$
$\mathbf{X} \sim \mathcal{L}$	random variable $\mathbf{X}$ that follows the distribution $\mathcal{L}$
$\mathbb{P}$ -almost surely	stands for a predicate that is true except on a set of null measure regarding the probability measure $\mathbb{P}$ (also written $\mathbb{P}$ -a.s.)



# Introduction

## Motivation and context of the study

In many signal processing applications like radar and sonar, seismology, telecommunications, or medical monitoring, measured data and signals are usually observed in noise and possibly interference. On the basis of these observations, one would want to decide whether one or several useful signals are present. These signals can either be consequences of some event of interest (*e.g.* whales vocalisations, presence of targets or obstacles, ...) or of anomalies we wish to be aware of in order to take adequate measures (*e.g.* cardiac arrhythmia, imminent earthquake, ...). Statistical decision theory - also called detection theory and hypothesis testing - is a useful tool to answer such question, and has many fields of application such as in economics, medicine, electronics, defence, meteorology, etc.

The choice of the detector we decide to use to solve the hypotheses testing problem depends on a manifold of factors. Firstly, the willingness to include prior information about the hypothesis, the importance of guaranteeing optimality according to a certain criterion and the robustness of the solution. Sometimes we must make compromise when several properties that are not concurrently reachable are targeted. For instance, in many detection scenarios the ambition is to maximise the probability of detection while minimising the probability of false alarm.

The nature of the application in hands is what determines which property to target and consequently what kind of approach to use. For example Uniformly Most Powerful (UMP) tests (Subsection 1.2.3.2), *i.e.* tests with the highest probability of detection for a fixed false alarm probability, can be used for the detection of natural disasters, sonar and radar applications, while the Bayesian risk (Subsection 1.16) and the Minimum Probability of Error (MPE) (Subsection 1.2.1) criterion can be used in digital communications. Therefore, for safety centered applications a certain conservatism and caution is vital because the effect of a miss detection or a false detection can be catastrophic. Whereas other applications, like quality inspection ones, rather aim at good detection on average. Thus, properties like optimality and robustness can be defined on the basis of average performance or the worst case performance for example.

In addition to the application field, the nature of the signals to be detected must also be taken into consideration. Signals that are commonly found in nature or produced by devices can fall into two categories: deterministic signals or stochastic

signals. Despite plenty of methods in the signal processing literature, the deterministic assumption may turn out to be too restrictive, and sometimes not realistic in some applications. As a matter of fact, some deterministic signals by nature can be randomised by the environment, the transmission channel or the receiver, and then come under the stochastic model. Add to that unknown deterministic signals that one chooses to assume random because too little information is available, and also signals that are innately random. Therefore, a wide range of signals can be modeled as random processes in different applications and fit into the the scope of stochastic signals with known distribution.

Unfortunately, even if considering the signals stochastic, rather than deterministic, could be far more realistic, assuming that their distribution is known can also be limiting. This can be the case in applications where signals of interest are polluted by unknown interference signals. Hence, we consider an alternative model, where the signal is assumed random with unknown distribution, but with no intention of knowing or estimating it. We call it the conditional approach. This alternative model seems to be the best option to avoid limitations of the unconditional model.

From our point of view, all the binary decision approaches that can be found in literature, for both deterministic and stochastic signals, are based on unconditional models. The reason is that their solution's criteria are always integral i.e., the involved probability density functions are not conditional. Whereas when signals are random with unknown distribution, the observation probability density function (PDF) is conditional and all the related criteria in the problem are non-integral. For more conciseness, let us consider an observation  $\mathbf{Y} = \Theta + \mathbf{X}$  such that  $\Theta$  is the random signal of unknown distribution and  $\mathbf{X}$  an additive Gaussian noise. An unconditional model will base the whole detection strategy on the PDF  $f_{\mathbf{Y}}(\mathbf{y})$ . Obviously, to do so,  $f_{\mathbf{Y}}(\mathbf{y})$  needs to be known, up to some parameter. Contrariwise, when  $f_{\mathbf{Y}}(\mathbf{y})$  is unknown, the model becomes conditional because the involved PDF that can be used in the decision process becomes conditional:  $f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y})$ . When  $\mathbb{P}_{\Theta}$ , the signal's distribution, is known, one can switch to the standard unconditional model and use non-conditional PDF's  $f_{\mathbf{Y}}(\mathbf{y}) = \int f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y})\mathbb{P}_{\Theta}d\theta$  to solve the decision problem. Consequently, knowing or not  $\mathbb{P}_{\Theta}$  is what determines the type of approach; conditional or unconditional.

In other words, in decision theory, signals to be detected in the conditional model are random with unknown distribution, and signals are random with a known distribution in the unconditional model. Indeed, deterministic signals can be considered as a particular case of random signals where it follows a Dirac distribution.

The signal processing community has led a large investigation about the unconditional model, and several solutions can be found in the literature, depending on the available quantity of information about the signal of interest, the noise and potential interference. However, for our alternative model: « the conditional model », there is clearly a scantiness of interest. It is within this context that my thesis: « Binary decision for observations with unknown distribution: An optimal and invariance-based framework » takes place.

Our terminology is somewhat different from the one used in estimation theory, where « conditional/unconditional » models are only linked to the nature of the signal

« deterministic/random ». However, it is particularly appropriate for the context of our research to pinpoint the path of our reasoning. Indeed, it shows the existence of a gap in the literature and that researchers have not much explored scenarios where signals are random with unknown distributions. Thus, we target cases where signals are not only random but also have unknown distributions and family of distributions *with no intention of finding them*. Therefore, we will consider decision problems within the scope of conditional models.

This particular terminology will help us classifying existing methods and drawing a straight line inside the state of the art detection theory. The first category concerns the unconditional model for which a panoply of detectors can be found in literature. Namely, the minimum probability of error test, the Bayes risk, Uniformly most powerful tests, the Generalised likelihood ratio test, Wald test, Rao's test, the Bayesian approach, robust tests, Wald's UBCP tests etc. With regards to the conditional model category, as far as we know, only the Random Distortion Testing framework (RDT), that has been elaborated by [Pastor and Nguyen, 2013a] in our laboratory a few years ago, proposes an optimal solution according to a certain criterion in this category.

Our works are in line with the RDT philosophy and the conditional model spirit, since we want to find detectors, without having to assume knowledge about signals distributions, while still seeking optimality. The focus has been made on the conditional model in order to develop a more general formulation of the RDT framework.

In addition to our aim at avoiding having to make constraining assumptions about signals distributions, our intention is to use symmetry properties of detection problems, namely invariance, to still find a solution, more specifically a statistical test, that verifies a certain criterion of optimality without any knowledge about the signal distribution.

Our methodology clearly stems from the shortage of works in the literature that take interest in detection problems of random signals with unknown distributions. Hereafter, the stakes of our outlook for such detection scenarios are quite significant.

## Plan and contributions

The present manuscript is composed of six chapters. The first three chapters are mainly bibliographic, aside from a small contribution in the end of the second chapter. The three last chapters present our works.

Chapter 1 presents the existing methods for solving detection problems when an unconditional model is considered, along with some fundamental definitions and properties in the field of statistical signal processing and, more specifically, decision theory.

Chapter 2 concerns the state of the art about invariance and the necessity of exploiting it in statistical tests to solve problems hardly solvable otherwise, or at least with no pledge of optimality. We present several notions in relation with invariance,



like the group of transformation, the maximal invariants, orbits and so on. Some limitations of the standard concept of invariance are also discussed, and an alternative definition of invariance is presented, more in line with the conditional model. This more adapted formulation of invariance can be considered as the first small contribution of my thesis.

In chapter 3, the entire RDT theoretical material is dissected, particularly the different steps of the proof, since it is the only method in literature that considers the conditional model with random signals of unknown distribution. Finally, we end up discovering the limits of this framework for applications where the noise is not necessary white Gaussian.

Chapter 4 is an application of the RDT framework in a context of distributed detection. It will show the compatibility of the RDT framework to such configurations thanks to its natural stretchability. The signal to be detected is of course assumed random with unknown distribution and  $p$  sensors are in charge of collecting the observations. The detection is operated with the use of the optimal test according to the RDT framework. This chapter can be considered as our first contribution.

Chapter 5 is our second contribution. Through Chapter 5, the RDT framework is generalised to a scenario where the noise's distribution is assumed not necessary to be Gaussian. The signal of interest is still assumed random with unknown distribution. The tested hypotheses or, more rigorously, the tested events are stated on the basis of a new invariant statistic, namely a maximal invariant. This maximal invariant is problem dependent, and does not necessarily take the pleasant form of a Mahalanobis norm as in the RDT framework. This leads to the formulation of the Generalised Random Distortion Testing (GRDT) problem. Some optimality properties are presented for the new GRDT framework.

Chapter 6 investigates scenarios where asymptotic optimality is reachable in the scope of the standard unconditional models. Asymptotic optimality is presented, in the case where the signal of interest is deterministic. By relying on the notion of Uniformly Most Powerful (UMP) tests, Asymptotically Uniformly Most Powerful (AUMP) tests are defined. Then an asymptotic version of Neyman-Pearson's theorem is proven first, before presenting an asymptotic formulation of Karlin-Rubin theorem and demonstrating it with the help of the first asymptotic theorem. This theoretical material is employed for a specific decision problem along with a state of the art method, the GLRT, so as to evaluate the practicality of our approach. This work has led to the following publications:

S. Bourmani, F.-X. Socheleau, D. Pastor, "Asymptotic Karlin-Rubin's Theorem with Application to Signal Detection in a Subspace Cone", IEEE 2019 27th European Signal Processing Conference (EUSIPCO).

S. Bourmani, F.-X. Socheleau, D. Pastor, "Théorème de Karlin Rubin asymptotique appliqué à la détection d'un signal dans un sous-espace cône", XXVIIème Colloque francophone de traitement du signal et des images, GRETSI 2019, Aug 2019, Lille, France.

# Chapter 1

## The unconditional approaches in detection theory

In this chapter, we present some fundamentals of detection theory in relation to the unconditional model. First, we recall several key notions of detection theory (statistical tests, optimality, robustness ...). Then, we present various methods from the literature that fall within the scope of the unconditional model. Further in this manuscript, the main hypothesis testing problems we intend to study under this model considers random signals with unknown distributions. However, in this chapter the state of the art concerns the deterministic case, because we consider that detection problems where signals are random with known distributions are just an extension of the deterministic particular case. To the best of our knowledge, there are no methods or solutions in the literature for hypothesis testing problems where signals are assumed random with unknown distribution, family of distribution and neighbourhood of distributions, while still achieving optimality, aside from the Random Distortion Testing (RDT) framework developed in our laboratory.

Before presenting the existing methods used when an unconditional model is considered, some definitions and properties about detection problems will be introduced.

### 1.1 Hypothesis testing problems

The most basic detection problems we can encounter are those where we need to decide whether a signal of interest embedded in noise is present or absent. The null hypothesis  $\mathcal{H}_0$  is hence the presence of noise only, while the alternative hypothesis  $\mathcal{H}_1$  is the presence of the signal in addition to noise. The recommended solution depends, on the one hand, on the properties of the noise and the signal, as well as all the available knowledge about them. On the other hand, it depends on the form of both hypotheses.

### 1.1.1 Specificities of the hypotheses

First, the problem can be binary or multiple. When the problem is binary, it only presents two alternatives:  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . As for the multiple decision problems, we normally wish to decide which hypothesis is true among  $M$  hypotheses:  $\{\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{M-1}\}$ , with  $M \geq 3$ . In this case, the  $M$ -ary hypothesis testing problem is more of a classification problem than a detection problem. Hence, a detection problem can be seen as a particular case of classification problems, also called discrimination problems. Classification problems are commonly found in communications where  $M$  symbols must be distinguished at the receiver, or in pattern recognition. For our works we only consider the case of binary decision problems.

The hypotheses can also be simple or composite. When a hypothesis is simple, the probability distribution of observations is completely specified with no unknown parameters. For example, if we consider a deterministic signal  $\boldsymbol{\theta} \in \mathbb{R}^d$  and  $\boldsymbol{\theta}_i \in \mathbb{R}^d$  its value under the hypothesis  $\mathcal{H}_i$ , then the hypotheses are  $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  versus  $\mathcal{H}_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ . In contrast, for composite hypotheses, the distributions are not completely specified which leads to an ensemble of possible models for one or both hypotheses. This uncertainty over the distributions results from the existence of disjointed parameter sets, take  $\Theta_0$  and  $\Theta_1$  where evolves the parameter  $\boldsymbol{\theta} \in \Theta$  such that  $\Theta = \Theta_0 \uplus \Theta_1$ . The hypotheses can be formulated as:  $\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0$  versus  $\mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1$ . In a nutshell, simple hypotheses permit to decide which state is true between one of two states, corresponding to the hypotheses, while composite hypotheses allow a multitude of states for each hypothesis. Throughout this manuscript, we are more interested in composite hypothesis testing problems.

Finally, composite hypothesis testing problems can be either one sided or two-sided. For instance, the one-sided hypotheses can be stated as:  $\mathcal{H}_0 : \theta \leq \theta_0$  versus  $\mathcal{H}_1 : \theta > \theta_0$  with  $\theta, \theta_0 \in \mathbb{R}$ , while the two-sided hypotheses can be formulated as:  $\mathcal{H}_0 : \theta = \theta_0$  versus  $\mathcal{H}_1 : \theta \neq \theta_0$ . From an optimality point of view, a one-sided problem is more likely to have an optimal solution than a two-sided problem. Consequently, seldom do we target tests with standard optimality criterion like Uniformly Most powerful (UMP) tests (Definition 1.2.1) for two-sided problem because it relies on the notion of increasing power function that stems strongly from the shape of the parameter sets. In brief, two-sided problems imply nested parameter sets and, hence, classical optimality is hardly pursued. Inherent properties of invariance (Definition 2.1.3) can be exploited to target an invariance-based optimality (see Chapter 2), while still considering a scalar parameter.

### 1.1.2 Statistical tests

A statistical test is a decision rule that allows to reject or accept the null hypothesis, based on an observation vector  $\mathbf{y} \in \mathbb{R}^N$ . Hence, it returns the index of the hypothesis considered to be true. A statistical test can either be randomised or non-randomised.

**Definition 1.1.1.** *A non-randomised test is a mapping  $\mathcal{T} : \Gamma \mapsto \{0, 1\}$ , for  $\mathbf{y} \in \mathbb{R}^N$*

and  $\Gamma \in \mathbb{R}^N$  such that

$$\mathcal{T}(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \in \mathcal{R}^{\mathcal{T}} \\ 0 & \text{if } \mathbf{y} \in \overline{\mathcal{R}^{\mathcal{T}}} \end{cases} \quad (1.1)$$

with  $\mathcal{R}^{\mathcal{T}} \cup \overline{\mathcal{R}^{\mathcal{T}}} = \Gamma$  and  $\mathcal{R}^{\mathcal{T}} \cap \overline{\mathcal{R}^{\mathcal{T}}} = \emptyset$ .  $\mathcal{R}^{\mathcal{T}}$  is called the critical region of the test  $\mathcal{T}$ .

When the test is non-randomised, it means that  $\mathbb{P}[\mathbf{y} \in \partial\mathcal{R}^{\mathcal{T}}] \neq 0$  where  $\partial\mathcal{R}^{\mathcal{T}}$  is the boundary of  $\mathcal{R}^{\mathcal{T}}$  with  $\mathcal{R}^{\mathcal{T}}$  an open subset. The shape of the test is as in Eq. (1.1). The boundary can be specified according to some optimality criterion. On the contrary, when  $\mathbb{P}[\mathbf{y} \in \partial\mathcal{R}^{\mathcal{T}}] = 0$ , the test needs to be randomised i.e. the decision must be randomly selected when the observation  $\mathbf{y}$  stands on the boundary of the critical region  $\mathcal{R}^{\mathcal{T}}$  such that:

$$\mathcal{T}(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \in \mathcal{R}^{\mathcal{T}} \\ k & \text{if } \mathbf{y} \in \partial\mathcal{R}^{\mathcal{T}} \\ 0 & \text{if } \mathbf{y} \in \overline{\mathcal{R}^{\mathcal{T}}} \end{cases} \quad (1.2)$$

with  $k \in (0, 1)$ . The value of  $k$  is chosen to respect the problem's constraints, for instance the false alarm probability [Lehmann and Romano, 2005a, Theorem 3.2.1].

All the distribution functions we will use in our works are continuous, and verify the condition:  $\mathbb{P}[\mathbf{y} \in \partial\mathcal{R}^{\mathcal{T}}] = 0$ . Thereby, only non-randomised tests will be considered in this manuscript.

A statistical test can generate four types of decisions as summarised in Table 1.1. Two are correct decisions and two are errors.

Reality Decision	$\mathcal{H}_0$ is true	$\mathcal{H}_1$ is true
Reject $\mathcal{H}_0$	Error of type I	Correct decision of type II
Accept $\mathcal{H}_0$	Correct decision of type I	Error of type II

Table 1.1: Types of decisions

The correct decision of type II corresponds to the probability of detection. It is commonly called the power of the test by statisticians. The power  $\beta_{\mathcal{T}}$  of a test  $\mathcal{T}$  is classically defined as:

$$\beta_{\mathcal{T}}(\mathbf{y}) \triangleq \mathbb{P}[\mathcal{T}(\mathbf{y}) = 1 \mid \mathcal{H}_1] \quad (1.3)$$

The type I error concerns the false alarm probability or also known as the size of a test. It is classically defined as:

$$\alpha_{\mathcal{T}} \triangleq \mathbb{P}[\mathcal{T}(\mathbf{y}) = 1 \mid \mathcal{H}_0] \quad (1.4)$$

The type I correct decision corresponds to  $\mathbb{P}[\mathcal{T}(\mathbf{y}) = 0]$  when the signal of interest is absent and only the noise is present, i.e. when  $\mathcal{H}_0$  is true. Finally, the type II error is commonly known as the probability of miss detection  $P_M = \mathbb{P}[\mathcal{T}(\mathbf{y}) = 0]$  when  $\mathcal{H}_1$  is true. The importance of this type of error can vary considerably according to the nature of the application. For example, for meteorological disaster forecasting,

the probability of miss detection can be of great interest knowing that such a mistake may cost people's lives.

However, throughout this manuscript, the focus is merely on the two first types of decisions because of the kind of optimality we target.

### 1.1.3 Optimality and robustness

#### 1.1.3.1 Optimality

We are tempted to say that the most desirable way to solve a decision problem is to find an optimal test. The principle of optimality relies heavily on what is called optimality criteria. There can be one single criterion or multiple criteria to optimise jointly. Most importantly, one needs to extract the requirements of the application and the decision problem in order to properly specify these criteria. Throughout this thesis, optimality is presented as a key notion in order to find solutions for the decision problems we take interest in. We usually associate it here to some test that is obtained after the maximisation of the probability of detection while limiting the probability of false alarm or any analogous probabilities. It is worth noticing though, that this is not the only optimality that exists, but we can say that it is the sort of optimality we look after. This optimality in the sense of Neyman-Pearson (Theorem 1.2.1 ) is very appropriate for radar and sonar like applications. In Section 1.2, various criteria of optimality will be presented, including Neyman-Pearson's, each giving rise to a specific method of detection, such as the Minimum Probability of Error (MPE), the Bayes risk, etc.

It must be emphasised that statistical tests with optimality properties are not always within reach. Sometimes, the problem does not contain sufficient information in order to derive optimal tests and the signal and/or noise and/or interference, if present, are very poorly known. In this thesis, very often, these are the considered scenarios and the reason why invariance-based methods engaged us. However, this is not the only reasoning we can conduct and seeing things from a different perspective is possible.

Indeed, another desirable way of solving decision problems is to derive robust tests such as Huber's tests that we present in the next subsection. Robustness can either be targeted because optimality is hardly attainable for the reasons we mentioned before, or because the problem itself calls for it. In the second case scenario, robustness is a forced choice to handle issues of mismatching when the signal of interest is in presence of physical phenomena that go unmodelled.

#### 1.1.3.2 Robustness

Considering that a decision problem is perfectly defined, and that the only lack of knowledge in it is encapsulated in the exhibited unknown parameters, is far too optimistic. In practice, it is obvious that sources of imprecision can be plentiful. The

first one that comes to mind is the presence of unexpected, unknown or hardly characterisable interference that can deflect the signal of interest from its nominal model. Another non-trivial problem is when the physical effects of the signal propagation environment are not taken into account faultlessly in the modelling process. Finally, the most basic one is when measuring instruments introduce inaccuracy. All these issues have in common the necessity to design tests that are resilient towards imprecision. Hence, designing robust tests should guarantee a certain insensitivity to deviations around the nominal model. These deviations can be more or less important depending on the characteristics of the application and its environment. In any case, robust tests need to include the degree of conservatism sought, by introducing a tolerance to avoid performance loss when departures from the nominal model occur. In other words, robust tests are optimal tests calculated on the basis of the worst case and robustness means stability of performance when variations of the nominal models happen.

In 1981, Huber developed a methodology for robust hypothesis testing [Huber, 1981]. He considered that under each hypothesis, a multitude of probability density functions are possible and the scheme is to select the best solution for the least-favorable probability density of the observation in the sense of some loss function, as we will see in what follows. This is called the minimax approach and can be considered quite pessimistic. More specifically, we consider an observation  $\mathbf{Y}$  that has a density function  $f_{\mathbf{Y}}^{\theta_i}$  under the hypothesis  $\mathcal{H}_i$ . The actual model  $f_{\mathbf{Y}}^{\theta_i}$  is unknown due to the eventual presence of physical effects. Only the nominal model  $g_i$  is known, such that the actual model  $f_{\mathbf{Y}}^{\theta_i}$  is the combination of some unknown physical effects to the known nominal model  $g_i$ . We consider the problem:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Y} = \boldsymbol{\theta}_0 + \mathbf{X} & \text{with } \mathbf{Y} \sim f_{\mathbf{Y}}^{\theta_0} \text{ and } f_{\mathbf{Y}}^{\theta_0} \in \mathcal{F}_0 \\ \mathcal{H}_1 : \mathbf{Y} = \boldsymbol{\theta}_1 + \mathbf{X} & \text{with } \mathbf{Y} \sim f_{\mathbf{Y}}^{\theta_1} \text{ and } f_{\mathbf{Y}}^{\theta_1} \in \mathcal{F}_1 \end{cases} \quad (1.5)$$

A neighbourhood  $\mathcal{F}_i$  with  $i \in (0, 1)$  can be defined on the basis of various metrics. An example of a particular neighbourhood  $\mathcal{F}_i$  is the contamination model defined as:

**Definition 1.1.2.** [Levy, 2008, Ch.6] Given the nominal probability distribution  $g_i$  and a tolerance  $\kappa_i$  with  $\kappa_i \in [0, 1]$ , the set  $\mathcal{F}_i$  of  $\kappa_i$ -contaminated probability densities centered on  $g_i$  is:

$$\mathcal{F}_i = \{f_{\mathbf{Y}}^{\theta_i} : f_{\mathbf{Y}}^{\theta_i} = (1 - \kappa_i)g_i + \kappa_i h_i\} \quad (1.6)$$

where  $h_i$  is an arbitrary probability density function that models the physical effects and  $f_{\mathbf{Y}}^{\theta_i}$  the actual model's probability density function.

Regardless of the metric chosen to construct the neighbourhoods  $\mathcal{F}_i$  with  $i \in (0, 1)$ , the general form of a neighbourhood  $\mathcal{F}_i$  of a probability density function  $f_{\mathbf{Y}}^{\theta_i}$  can be defined as:

$$\mathcal{F}_i = \{f_{\mathbf{Y}}^{\theta_i} : d(f_{\mathbf{Y}}^{\theta_i}, g_i) \leq \kappa_i\} \quad (1.7)$$

with  $\kappa_i \in \mathbb{R}$  a prespecified level of tolerance that symbolises the uncertainties. The distance  $d(\cdot, \cdot)$  can be chosen to be the Kolmogorov distance, the Lévy distance, the Kulback-Leibler divergence, the total variation, the Prohorov distance etc.

Whatever distance selected, the next step is to determine the least favorable probability density in the neighbourhood for each hypothesis. Such an optimisation is

operated with respect to some criterion, called in this context a loss function or a score function  $\mathcal{L}(\cdot, \cdot, \cdot)$ . The general form of the minimax test is:

$$\min_{\mathcal{T}} \max_{g_0 \in \mathcal{F}_0, g_1 \in \mathcal{F}_1} \mathcal{L}(\mathcal{T}, g_0, g_1) \quad (1.8)$$

For some methods, such as Neyman-Pearson tests (Theorem 1.2.1) and the Bayes Risk (Eq. (1.15)), this function depends on the probability of false alarm or probability of miss detection or both. Then, the decision rule or the statistical test is derived by an optimisation of the detection performance. For a detailed procedure to obtain robust Bayes risk and Neyman-Pearson tests refer to [Levy, 2008, Ch. 6].

Robustness, as stated here, seems reachable, by following Huber's procedure, and by combining it to a detection method. In reality, some convexity/concavity issues can appear due to the additional optimisation aspects that Huber's methodology involves. In addition, the computation complexity that robustness pursuit might require must not be neglected.

## 1.2 Optimality oriented approaches for the unconditional model

In this section, we intend to present several classical methods of detection theory, applicable for unconditional models. First, we recall the minimum probability of error approach, and the Bayes risk and recall the link between them. Secondly, we will bring forward uniformly most powerful tests through two main theorems; the Neyman-Pearson theorem and the Karlin-Rubin theorem. These two theorems will be subject to a contribution later on in Chapter 6 by means of a generalisation to the asymptotic case. Finally, we will introduce the well known generalised likelihood ratio test.

### 1.2.1 The minimum probability of error test

The minimum probability of error test is based on the assumption that the probability of occurrence of each hypothesis is known such that  $\mathbb{P}(\mathcal{H}_0) + \mathbb{P}(\mathcal{H}_1) = 1$ . According to the application at hand, this assumption can be more or less reasonable. This prior belief about the hypotheses participate in the derivation of the threshold with which the likelihood ratio is compared. Analogously to uniformly most powerful tests, generalised likelihood ration tests etc., this approach belongs also to the category of tests that rely on likelihood ratios to design detectors. Let us consider the binary hypotheses problem:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Y} = \boldsymbol{\theta}_0 + \mathbf{X} & \text{with: } \mathbf{Y} \sim f_{\mathbf{Y}}^{\boldsymbol{\theta}_0} \\ \mathcal{H}_1 : \mathbf{Y} = \boldsymbol{\theta}_1 + \mathbf{X} & \text{with: } \mathbf{Y} \sim f_{\mathbf{Y}}^{\boldsymbol{\theta}_1} \end{cases} \quad (1.9)$$

with  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}_1$  the signal respectively under the null hypothesis and the alternative hypothesis. The noise is modeled by  $\mathbf{X}$ , the observation by  $\mathbf{Y}$  and  $f_{\mathbf{Y}}^{\boldsymbol{\theta}_i}$  is the probability density function of  $\mathbf{Y}$  parametrized by  $\boldsymbol{\theta}_i$ .

The probability of error is defined as:

$$\begin{aligned} P_e &= \Pr[\text{decide } \mathcal{H}_0 \text{ and } \mathcal{H}_1 \text{ true}] + \Pr[\text{decide } \mathcal{H}_1 \text{ and } \mathcal{H}_0 \text{ true}] \\ &= \mathbb{P}(\mathcal{H}_0 | \mathcal{H}_1)\mathbb{P}(\mathcal{H}_1) + \mathbb{P}(\mathcal{H}_1 | \mathcal{H}_0)\mathbb{P}(\mathcal{H}_0) \end{aligned} \quad (1.10)$$

The probability  $\mathbb{P}(\mathcal{H}_i | \mathcal{H}_j)$  indicates the probability of choosing  $\mathcal{H}_i$  when in fact  $\mathcal{H}_j$  is true and the probability  $\mathbb{P}(\mathcal{H}_i)$  is the probability of occurrence of the hypothesis  $\mathcal{H}_i$ . The purpose of this approach is to deliver a detector that is optimal according to the minimum probability of error criterion. Therefore:

$$P_e = \mathbb{P}(\mathcal{H}_1) \int_{\overline{\mathcal{R}^1}} \mathbb{P}(\mathbf{y} | \mathcal{H}_1) dy + \mathbb{P}(\mathcal{H}_0) \int_{\mathcal{R}^1} \mathbb{P}(\mathbf{y} | \mathcal{H}_0) dy \quad (1.11)$$

with  $\mathcal{R}^1 = \{\mathbf{y} : \text{such that the decision is } \mathcal{H}_1\}$  the critical region and  $\overline{\mathcal{R}^1} = \{\mathbf{y} : \text{such that the decision is } \mathcal{H}_0\}$  its complement. Considering that the critical region  $\mathcal{R}^1$  and  $\overline{\mathcal{R}^1}$  its complement partition:

$$\int_{\overline{\mathcal{R}^1}} \mathbb{P}(\mathbf{y} | \mathcal{H}_i) dx = 1 - \int_{\mathcal{R}^1} \mathbb{P}(\mathbf{y} | \mathcal{H}_i) dy \quad (1.12)$$

Then:

$$P_e = \mathbb{P}(\mathcal{H}_1) + \int_{\mathcal{R}^1} [\mathbb{P}(\mathcal{H}_0)\mathbb{P}(\mathbf{y} | \mathcal{H}_0) - \mathbb{P}(\mathcal{H}_1)\mathbb{P}(\mathbf{y} | \mathcal{H}_1)] dy \quad (1.13)$$

Finally, the resulting test is:

$$\mathcal{T}_e(\mathbf{y}) = \begin{cases} 1 & \text{if } \frac{\mathbb{P}(\mathbf{y}|\mathcal{H}_1)}{\mathbb{P}(\mathbf{y}|\mathcal{H}_0)} > \frac{\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_1)} = \lambda_e \\ 0 & \text{otherwise} \end{cases} \quad (1.14)$$

with  $\lambda_e$  the threshold with which the likelihood ratio is compared to. When the probability of occurrence of each the hypotheses are equal i.e.  $\mathbb{P}(\mathcal{H}_1) = 1/2$  and  $\mathbb{P}(\mathcal{H}_0) = 1/2$  (in the case binary hypotheses testing of course), the threshold test becomes a common comparison between the probabilities .

The shape of the test is very similar to Neyman-Pearson's test but the criterion of optimality is not the same as we will see in the sequel. Indeed the optimised criterion of optimality here is the probability of error.

The following approach is the Bayes risk. It generalises the minimum probability of error approach by considering in addition to the probabilities of occurrence, costs assigned to each type of decision.

### 1.2.2 The Bayes risk

The Bayes risk method enshrines in the Bayesian paradigm that says that any problem with uncertainty can be solved by using probabilistic models. Although the Bayes risk is quite different because no unknown parameter is assigned probability distributions. Instead, in the Bayes risk method, each type of error/decision (figure 1.1) is assigned a cost. These costs reflect one's prior belief and previous experience about the specificities of the applications. In other words they can be very approximate sometimes even



subjective and judgemental. For example, designing a radar to detect an enemy plane that attempts to harm has not the same stakes as developing detectors for cars to detect the presence of thieves. Indeed, in the first case, a false alarm could lead to the extermination of a plane full of innocents, whereas a false alarm for a car may only lead to strike the alarm and, in the worse case scenario, to wake up the whole neighbourhood. Once costs have been assigned based on the prior knowledge, the Bayes risk is derived:

$$\mathbf{R} = \sum_{i=0}^1 \sum_{j=0}^1 C_{ij} \mathbb{P}(\mathcal{H}_i | \mathcal{H}_j) \mathbb{P}(\mathcal{H}_j) \quad (1.15)$$

with  $C_{ij}$  is the assigned cost for choosing the hypothesis  $i$  while the hypothesis  $j$  is true. Notice that the Bayes risk is a generalisation of the probability of error, and that the MPE criterion is a special case of the Bayes risk criterion when  $C_{00} = C_{11} = 0$  and  $C_{10} = C_{01} = 1$ .

In order to find the corresponding test for the Bayes risk criterion, the expected cost  $\mathbf{R}$  is minimised analogously to the probability of error (see Eq. (1.11), (1.12) and 1.13). Assuming that  $C_{10} > C_{00}$  and  $C_{01} > C_{11}$ , which is a very reasonable assumption, the expression of the test that minimises the Bayes risk  $\mathbf{R}$  is:

$$\mathcal{T}_{\mathbf{R}}(\mathbf{y}) = \begin{cases} 1 & \text{if } \frac{\mathbb{P}(\mathbf{y}|\mathcal{H}_1)}{\mathbb{P}(\mathbf{y}|\mathcal{H}_0)} > \frac{(C_{10}-C_{00})\mathbb{P}(\mathcal{H}_0)}{(C_{01}-C_{11})\mathbb{P}(\mathcal{H}_1)} = \lambda_{\mathbf{R}} \\ 0 & \text{otherwise} \end{cases} \quad (1.16)$$

The threshold  $\lambda_{\mathbf{R}}$  is used similarly to any other likelihood ratio approach, by comparing it to the likelihood ratio to minimise the Bayes risk.

Unfortunately, it is not always possible to obtain information about probabilities of occurrence of each hypothesis to derive the minimum probability of error test and the costs to minimise the Bayes risk. Therefore, if one has an application at hand with no such knowledge, but still needs to compute a detector that guarantees optimality according to a certain criterion, uniformly most powerful tests can be the solution. Usually, uniformly most powerful tests are more often employed in radar and sonar like applications, while the two other approaches are more commonly found in applications like pattern recognition, digital communications etc.

### 1.2.3 Uniformly most powerful tests

In detection theory, when the considered model is unconditional and deterministic, and we have no willingness of following the Bayesian spirit, by any means, a uniformly most powerful test may be the answer. A uniformly most powerful test is a statistical test that guarantees the highest power  $\beta$  (Eq. (1.3)) i.e. probability of detection, among all the tests with the same level  $\gamma$ , i.e. with a size  $\alpha_{\mathcal{T}} \leq \gamma$ , such that  $\alpha_{\mathcal{T}}$  is defined in . (Eq. (1.4)).

**Definition 1.2.1.** *Let  $\mathbf{Y}$  be a random vector with probability density function  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  where  $\theta \in \Theta$ . We consider the following detection problem*

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1, \quad \Theta_0 \cup \Theta_1 = \Theta, \quad \Theta_0 \cap \Theta_1 = \emptyset \end{cases} \quad (1.17)$$

A test  $\mathcal{T}$  is called UMP (Uniformly Most Powerful) of level  $\gamma$  if

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\mathcal{T}(\mathbf{Y})] \leq \gamma \quad (1.18)$$

and if for any other test  $\mathcal{T}'$  satisfying

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \leq \gamma$$

we have,  $\forall \theta \in \Theta_1$

$$\mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \leq \mathbb{E}_\theta[\mathcal{T}(\mathbf{Y})]. \quad (1.19)$$

### 1.2.3.1 The Neyman-Pearson lemma

The well known Neyman-Pearson lemma first appeared in 1933 in a paper of Jerzy Neyman and Egon Sharpe Pearson [Neyman and Pearson, 1933]. It is used to find an optimal test according to the criterion of the maximum probability of detection of a test. Hence, the purpose is to obtain the test with the greatest power for a fixed size. However, this lemma is dedicated to detection problems where the considered hypotheses are simple, which means that each hypothesis corresponds to one distribution only. The problem needs also to be completely known and unknown parameter must be specified by a single value i.e. no unknown parameters can be present in the distributions. The Neyman-Pearson lemma basically says that likelihood ratio tests are optimal, for testing simple hypothesis problems, according to a certain criterion, that is, the best probability of detection for a fixed probability of false alarm.

We recall that only non-randomised tests are considered in this manuscript.

**Theorem 1.2.1.** *Let  $\mathbf{Y}$  be a random vector of probability density function  $f_{\mathbf{Y}}^\theta(\mathbf{y})$  with  $\theta \in \Theta$ . We consider the following detection problem*

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1 \end{cases} \quad (1.20)$$

Let  $\mathcal{T}_{NP}$  be the threshold test of the form:

$$\mathcal{T}_{NP}(\mathbf{y}) = \begin{cases} 1 & \text{if } \frac{f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} > \lambda_{NP} \\ 0 & \text{otherwise} \end{cases} \quad (1.21)$$

where the threshold  $\lambda_{NP} \geq 0$  is computed such that  $\mathbb{E}_{\theta_0}[\mathcal{T}_{NP}(\mathbf{Y})] = \gamma$  and  $0 \leq \gamma \leq 1$ . Hence the test  $\mathcal{T}_{NP}$  is most powerful of level  $\gamma$  for testing  $\mathcal{H}_0 : \theta = \theta_0$  against  $\mathcal{H}_1 : \theta = \theta_1$ .

As mentioned earlier, the Neyman-Pearson lemma provides a most powerful test only for simple hypothesis testing problems. In the same vein, a theorem exists that applies to the composite hypothesis testing case to find a uniformly most powerful test (see Definition 1.2.1). It is the Karlin-Rubin theorem.

### 1.2.3.2 The Karlin-Rubin theorem

The Karlin-Rubin theorem [Karlin and Rubin, 1956] can be regarded as an extension of the Neyman-Pearson lemma to the case of composite hypothesis testing. Considering such detection problems, is more realistic, due to the number of applications where the probability distribution functions of the observations are not totally known, in other words when the PDFs have unknown parameters that evolve inside known intervals. Therefore, designing detectors for such detection scenarios can be of significant practical interest. However not all composite hypothesis testing problems can be solved using the Karlin-Rubin theorem; the detection problem needs to be one-sided. As for two-sided detection problems, uniformly most powerful invariant tests can be sometimes envisaged, depending on the formulation of the problem (see Chapter 2) if such an optimality is still targeted. As stated before, only binary hypotheses are considered in this manuscript.

**Theorem 1.2.2.** *Let  $\mathbf{Y}$  be a random vector of pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  where  $\theta \in \Theta \subset \mathbb{R}$  and let  $\Theta_1$  and  $\Theta_0$  be two subsets of  $\Theta$  such that  $\Theta = \Theta_0 \cup \Theta_1$  and  $\Theta_0 \cap \Theta_1 = \emptyset$ . We consider the following problem:*

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1 \end{cases} \quad (1.22)$$

Let the likelihood ratio be:

$$\Lambda_{\theta_1, \theta_0}(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} \quad (1.23)$$

with  $f_{\mathbf{Y}}^{\theta_0} \neq 0$ , for  $\theta_1 \in \Theta_1$  and  $\theta_0 \in \Theta_0$ . Let the likelihood ratio be such that  $\Lambda_{\theta_0, \theta_1} = h_{\theta_0, \theta_1}(V)$ , with  $V : \mathbb{R}^N \mapsto \mathbb{R}$ . The test  $\mathcal{T}_{KR}$  defined as:

$$\mathcal{T}_{KR}(\mathbf{y}) = \begin{cases} 1 & \text{if } V(\mathbf{y}) > \lambda_{KR} \\ 0 & \text{otherwise} \end{cases} \quad (1.24)$$

with  $\lambda_{KR}$  determined such that:  $\mathbb{E}_{\theta_0}[\mathcal{T}_{KR}(\mathbf{Y})] = \gamma$ . If for  $\theta_1 > \theta_0$ ,  $h_{\theta_1, \theta_0}$  is strictly increasing in  $V$ , then the tests  $\mathcal{T}_{KR}$  is uniformly most powerful for testing  $\mathcal{H}_1$  against  $\mathcal{H}_0$ .

Although Uniformly Most Powerful (UMP) tests guarantee a certain optimality, they need to know all the parameters of the problem, meaning that the detection problem must be completely known. In practice, parameters can often be unknown, which lead us to the state of the art methods used in such scenarios; the holy trinity i.e. the Generalised Likelihood ratio test (GLRT), the Wald test and the Rao test, as well as the Bayesian approach.

### 1.2.4 The holy trinity

The holy trinity gathers three common techniques in statistical decision theory: the Generalised Likelihood ratio test (GLRT), the Wald test and the Rao test. There are used when no Uniformly Most Powerful test exists, for instance when the problem under consideration is two-sided (see subsection 1.1.1 ) or/and when the detection problem

contains unknown parameters. The (GLRT), or more specifically, the likelihood ratio tests, were introduced by Jerzy Neyman and Egon Sharpe Pearson in 1928, the Wald test by the hungarian mathematician Abraham Wald in 1943 and the Rao test by the Indian-American mathematician and statistician Caleyampudi Radha Krishna Rao, also called the score test, in 1948. These tests are known to have similar asymptotic performance in terms of detection, but with no guarantee of optimality according to a determined criterion. They are based on the same principle of comparing a test statistic to a threshold. The particularity of the score test is that it requires no maximum likelihood estimators (MLE) unlike the other two tests.

#### 1.2.4.1 The Generalised Likelihood Ratio test

The GLRT consists of calculating the ratio of the probability density function (pdf) under each hypothesis, "the likelihood ratio", and then replace the unknown parameters by their maximum likelihood estimates (MLEs). The resulting test statistic is then compared to a threshold. We consider a two sided detection problem where  $\mathbf{Y}$  is a random vector of pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  and where  $\theta \in \Theta \subset \mathbb{R}$ . For simplicity, we assume that the parameter  $\theta$  is unknown under the alternative hypothesis and known under the null hypothesis such that:

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0 \end{cases} \quad (1.25)$$

The (GLRT) test statistic is [Kay, 2009]:

$$L_G(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\hat{\theta}_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} > \lambda_G \quad (1.26)$$

where  $\hat{\theta}_1 = \underset{\theta_1}{argmax} f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})$  is the MLE of  $\theta_1$ , that is  $\theta$  when the alternative hypothesis  $\mathcal{H}_1$  is true. The threshold  $\lambda_G$  is set to guarantee a certain probability of false alarm, but unfortunately there is no guarantee of respecting the constraint due to the use of MLEs. However, in practice, the GLRT is known to give quite good results even if it is with no pledge of optimality. In asymptotic scenarios, it has been found that the GLRT may become optimal in the sense of Uniformly Most Powerful Invariant (UMPI) test [Kay, 1998], meaning that it becomes UMP inside the class of invariant tests. Despite its flaws, when no UMP or UMPI test exists and when the Bayesian approach cannot be considered, the GLRT remains the best option according to the community of signal processing.

#### 1.2.4.2 The Rao test

The GLRT provides a test statistic that is undeniably simple and has interesting asymptotic optimality properties [Kay, 1998] according to the Wilks theorem. However, the Rao test (1948), also called the score test, can be a good alternative to the GLRT, especially when the problem requires, the calculation of maximum likelihood estimates under both hypotheses in opposition to the given GLRT example presented in Eq. (1.26), which can become constraining. In contrast, the Rao test statistic needs

a maximum likelihood estimate only under the null hypothesis. This means that for the Problem (1.25), the Rao test will not even need to derive a maximum likelihood estimate for the test statistic because the signal  $\theta$  is known under  $\mathcal{H}_0$ . Let us take the same binary decision problem as an example. Let  $\mathbf{Y}$  be a random vector of pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  and where  $\theta \in \Theta \subset \mathbb{R}$  is unknown under the alternative hypothesis. We consider the same problem from (1.25).

The Rao statistic test for Problem (1.25) is defined as [Kay, 2009]:

$$T_R(\theta) = \frac{\partial \ln(f_{\mathbf{Y}}^{\theta}(\mathbf{y}))}{\partial \theta} \Big|_{\theta=\theta_0}^T \mathbf{I}^{-1}(\theta_0) \frac{\partial \ln(f_{\mathbf{Y}}^{\theta}(\mathbf{y}))}{\partial \theta} \Big|_{\theta=\theta_0} \quad (1.27)$$

The matrix  $\mathbf{I}(\theta_0)$  represents the Fisher information and  $\mathbf{I}^{-1}(\theta_0)$  its inverse. As we can see in Eq. (1.27), no MLE is derived since  $\theta$  is equal to  $\theta_0$  under  $\mathcal{H}_0$  and  $\theta_0$  is known. If  $\theta_0$  was unknown, a maximum likelihood estimate would need to be derived and  $\theta_0$  would be replaced by  $\hat{\theta}_0$  in 1.27.

The test statistic  $\mathcal{T}_R(\theta)$  is then compared to a threshold, just like the GLRT, to take a decision about which hypothesis is true:  $\mathcal{H}_0$  or  $\mathcal{H}_1$ . The threshold can be calculated so as to ensure a constraint on the false alarm:

$$T_R(\theta) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_R \quad (1.28)$$

The Rao test is known to be equivalent to the GLRT in asymptotic scenarios [Kay, 1998]. Consequently their asymptotic performance are also equivalent. It has been claimed though by [Chandra and Sankhya, 1985] and [Chandra and Joshi, 1983] that Rao's test is locally more powerful than the GLRT and the Wald test that comes next.

### 1.2.4.3 The Wald test

The Wald test is the third test from the holy trinity. Just like the Rao test, the Wald test has asymptotic statistics identical to the GLRT since it is based on the asymptotic distribution of the maximum likelihood estimates. Hence, they all three have the same level of performance asymptotically. However, the Wald test is not, in finite samples, invariant to changes in the representation of the null hypothesis. Let us consider the same binary hypothesis testing problem as for the GLRT and Rao test. Let  $\mathbf{Y}$  be the observation of pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$  and where  $\theta_0 \in \mathbb{R}$  and  $\theta \in \Theta \subset \mathbb{R}$ :

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0 \end{cases} \quad (1.29)$$

Wald test statistic  $T_W(\theta)$  for the two-sided decision Problem 1.29 is defined as [Kay, 2009]:

$$T_W(\theta) = (\hat{\theta}_1 - \theta_0)^T \mathbf{I}(\hat{\theta}_1) (\hat{\theta}_1 - \theta_0) \quad (1.30)$$

with  $\hat{\theta}_1$  the maximum likelihood estimate of  $\theta$  under the alternative hypothesis  $\mathcal{H}_1$ . The matrix  $\mathbf{I}(\hat{\theta}_1)$  represents again the Fisher information.

Analogously to the other two tests of the holy trinity, the test statistic  $T_W(\theta)$  is compared to an appropriate threshold determined on the basis of some constraint:

$$T_W(\theta) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_W \quad (1.31)$$

Even if the GLRT is the most famous test of the holy trinity, the Rao and the Wald tests can induce less computational burden than the GLRT, since they do not necessary need maximum likelihood estimates under both hypotheses. On the other hand, the GLRT has invariance properties that the others two do not have and is insensitive to reparametrisation. In the end, the three tests are asymptotically equivalent, and the best choice will depend on the detection scenario.

### 1.2.5 The Bayesian approach

In statistical binary decision, the Bayesian approach entails having a parametric model of observation and assigning prior probabilities, often called just priors, to unknown parameters. Priors are probability distributions that reflects the quantity of information we have or we believe we have, without necessary an evidence, about the unknown parameters. The Bayesian approach is slightly different from the Bayes risk recalled in Subsection 1.2.2, where only probability of occurrence of hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are used and where optimality according to a certain criterion is ensured. Whereas with the Bayesian approach, no optimality can be ensured and priors are attributed to any unknown parameter without being necessary the hypotheses probability of occurrence. Moreover, the Bayesian approach considers unknown parameters as realisations of random variables, and integrates over them. Let  $\mathbf{Y}$  be some observation vector and  $f_{\mathbf{Y}}^{\theta_i}$  its pdf under the hypothesis  $\mathcal{H}_i$  with  $\theta_i$  the unknown parameter under the same hypothesis:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Y} = \theta_0 + \mathbf{X} & \text{with } \mathbf{Y} \sim f_{\mathbf{Y}}^{\theta_0} \text{ and } \theta_0 \in \Theta_0 \\ \mathcal{H}_1 : \mathbf{Y} = \theta_1 + \mathbf{X} & \text{with } \mathbf{Y} \sim f_{\mathbf{Y}}^{\theta_1} \text{ and } \theta_1 \in \Theta_1 \end{cases} \quad (1.32)$$

We denote  $\mathbb{P}_{\theta_i}$  the prior probability of  $\theta_i$  under the hypothesis  $\mathcal{H}_i$ . Thus, the unconditional probability density functions  $f_{\mathbf{Y}}^{\theta_0}$  and  $f_{\mathbf{Y}}^{\theta_1}$  are completely specified with the expressions [Kay, 2009]:

$$\begin{aligned} f_{\mathbf{Y}}^{\theta_0}(\mathbf{y}) &= \int f_{\mathbf{Y}|\theta_0}(\mathbf{y} | \mathcal{H}_0) \mathbb{P}_{\theta_0} d(\theta_0) \\ f_{\mathbf{Y}}^{\theta_1}(\mathbf{y}) &= \int f_{\mathbf{Y}|\theta_1}(\mathbf{y} | \mathcal{H}_1) \mathbb{P}_{\theta_1} d(\theta_1) \end{aligned} \quad (1.33)$$

with  $f_{\mathbf{Y}|\theta_i}(\mathbf{y} | \mathcal{H}_i)$  the conditional probability density function of  $\mathbf{Y}$  under  $\mathcal{H}_i$ . The decision rule can be obtained, on the basis of  $f_{\mathbf{Y}|\theta_0}(\mathbf{y} | \mathcal{H}_0)$  and  $f_{\mathbf{Y}|\theta_1}(\mathbf{y} | \mathcal{H}_1)$ , and be compared to a threshold, similarly to the Neyman-Pearson test for example, by

deriving a likelihood ratio. More generally, a loss function or a utility function can be written on the grounds of  $f_{\mathbf{Y}|\theta_0}(\mathbf{y} | \mathcal{H}_0)$  and  $f_{\mathbf{Y}|\theta_1}(\mathbf{y} | \mathcal{H}_1)$  to derive the decision rule. However, combining this approach to some optimal test like Neyman-Pearson's, is not a pledge of optimality.

It is also worth noting that the Bayesian approach is not self-sufficient for binary decision and that it is more of an estimation strategy based on the Bayes theorem to subvert uncertainty introduced by unknown parameters. That's why it is also called the Bayesian inference. Furthermore, its philosophical validity and justification has been extensively discussed in [Carlin and Louis, 2010], [Bernardo and Smith, 2009], [Gelman et al., 2013] and [Marin and Robert, 2007], and due to the absence of assured optimality, the signal processing community has not always agreed about this approach. However, it is undeniable that the Bayesian approach is based on a strong and solid theoretical material and that it can be interesting to apply for many applications.

## 1.2.6 Wald's UBCP tests

In 1943, Wald presented in [Wald, 1943] the Uniformly Best Constant Power (UBCP) tests for composite hypothesis testing problems. Completely different from the classical Wald test previously described in Subsection 1.2.4.3, it can be considered as the source of inspiration for the authors of [Pastor and Nguyen, 2013b] about the RDT framework. The UBCP test is hence very similar to the RDT test, but with also some resemblance to the UMP tests with an additional aspect of invariance.

The class of tests  $\mathcal{K}_\gamma$  is defined as:  $\mathcal{K}_\gamma = \{\mathcal{T} : \sup_{\boldsymbol{\theta} \in \overline{\mathcal{R}^{\mathcal{T}}}} \mathbb{P}[\mathcal{T}(\boldsymbol{\theta}) = 1] \leq \gamma\}$  with  $0 \leq \gamma \leq 1$  and  $\overline{\mathcal{R}^{\mathcal{T}}}$  the complementary of the critical region of a test  $\mathcal{T}$  presented in Definition 1.1.1. In other words, the class of tests  $\mathcal{K}_\gamma$  is composed of tests of level  $\gamma$ .

**Definition 1.2.2.** [Wald, 1943] *Let  $k$  be some real positive parameter. Let  $\mathcal{S}$  be a family of surfaces defined as:  $\mathcal{S} = \{S_k : \boldsymbol{\theta} \in \mathbb{R}^N \text{ such that } \phi(\boldsymbol{\theta}) = k\}$ , where  $\phi(\boldsymbol{\theta})$  refers to some analytic function of  $\boldsymbol{\theta}$ . A test  $\mathcal{T} \in \mathcal{K}_\gamma$  is said to have uniformly best constant power on  $\mathcal{S}$  if:*

- For any  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}'$  belonging to the same surface  $S_k \in \mathcal{S}$ , we have  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) = \beta_{\mathcal{T}}(\boldsymbol{\theta}')$ .
- For any other test  $\mathcal{T}' \in \mathcal{K}_\gamma$  that verifies the first item i.e.  $\beta_{\mathcal{T}'}(\boldsymbol{\theta}) = \beta_{\mathcal{T}'}(\boldsymbol{\theta}')$ , we have  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) \geq \beta_{\mathcal{T}'}(\boldsymbol{\theta})$  for all  $\boldsymbol{\theta} \in S_k$  and all surfaces  $S_k$ .

The function  $\phi(\boldsymbol{\theta})$  can represent the squared energy of the signal for example. The first item of Definition 1.2.2 specifies the constant power on a surface in  $\mathcal{S}$  and the second item sets the optimality criterion on any surface  $S_k$ . The notion of family of surfaces  $\mathcal{S}$  can be considered as a set of equivalent points that constitute partitions, defined with respect to the problem's properties of invariance.

In the following chapter, the standard concept of invariance will be presented along with some associated notions. Then, we will present our own conception of

invariance for the conditional model, introduced above. Briefly speaking, our definition is suitable for models where signals of interest are random with unknown distributions. Then, in Chapter 3, the RDT framework [Pastor and Nguyen, 2013b] will be presented and the similarities with Wald's UBCP test will be given.





# Chapter 2

## Invariance in detection theory

In detection theory, when it is hard to find an optimal test according to a certain criterion, narrowing the class of tests can be considered. One way of constructing the class of tests is to use the inherent invariance of the detection problem. Invariance can be seen as natural symmetries of the problem from which we can extract information to restrain the class of tests in order to find a solution more easily. Thus, invariance is a possible method that helps obtaining optimality when standard approaches, such as the ones presented in the introduction and the previous chapter, fail to provide that. In some cases, when the invariance is too fundamental to the problem for not being taken into account in the decision rule, the restricted class of tests becomes the problem's basic class of interest itself.

To the best of our knowledge, the notion of invariance has started to blossom in the late 1930s', during the second world war, with researchers such as Wald, Scharf, Lehman, Borovkov and others. Indeed, due to the intensive scientific researches in the detection theory field during the second world war, it has quickly been noticed that in some scenarios, when no unbiased uniformly most powerful tests exists, detection theory had limitations hard to overcome with the standard approaches. That is how the concept of invariance came to light. However, these last twenty years, not a lot of papers took interest in the subject for decision theory, considering that almost everything has already been studied. Actually, the extend of the utility of invariance is yet to explore, in particular in detection problems where the signal's distribution is unknown and for which optimal tests are not even targeted. It is worth noticing though that in applications that involves image processing, invariance has always been central and continues to be. Which is not necessary the case for our field of interest.

The principle of invariance is primarily based on a suitable group of transformations that arises from the nature of the problem's properties of symmetry, in other words, a group of transformations that leaves the problem invariant. Then a maximal invariant, that is a statistic with properties of maximality and invariance that operates a reduction on data, is derived accordingly. The extracted maximal invariant will be used as a test statistic in the expression of the decision rule or what is more commonly known as the statistical test.

In this chapter, the notion of invariance will be presented in details, along with

some fundamental related results. It will lead us to some severe limitations of the approach due to standard definitions of invariance that misfit a more general framework. For instance, when the signal is random and its distribution is unknown. This chapter will however help us have a glimpse of the purpose of relatively different approaches like the random distortion testing framework (RDT) and its generalisation the (GRDT). Approaches, where invariance will continue to play a major role, although in a different sense than the one usually found in literature.

## 2.1 Invariant detection problems

### 2.1.1 The group of transformation

The first important notion related to invariance we need to mathematically define is the group of transformation. The following definition of a group can be found in [Eaton, 1983, Definition 6.1, p. 185].

**Definition 2.1.1.** *A group  $(G, \circ)$  is a set  $G$  together with a binary operation  $\circ$  such that the following properties hold for all elements in  $G$ :*

- $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ .
- *There is a unique element in  $G$ , denoted  $e$ , such that  $g \circ e = e \circ g = g$  for all  $g \in G$ . The element  $e$  is the identity in  $G$ .*
- *For each  $g \in G$ , there is a unique element in  $G$ , denoted by  $g^{-1}$  such that  $g \circ g^{-1} = g^{-1} \circ g = e$ . The element  $g^{-1}$  is the inverse of  $g$ .*

As regards to a group in the context of invariance and detection theory, the binary operation is implicit and is not expressed by an operator so that we write for example  $g_1 g_2$  instead of  $g_1 \circ g_2$ .

As we said before, a group of transformation should normally emanate from the very nature of the detection problem in hands. However, in practice, finding the exact expression of a suitable group of transformation is not straightforward. Sometimes it can be triggered by intuition by taking a look at the problem or by imagining its geometry [Scharf, 1991], sometimes we can easily guess the form of a maximal invariant of the problem from which we can suppose the existence and derive a group of transformation. Either way, we have to make sure that the properties of its definition are verified along with those of the maximal invariant (see subsection maximal invariant). The best way to understand how tricky it can be, is to take a simple example where intuition is clearly helpful:

**Example 2.1.1.** [Scharf, 1991] *Let  $\mathbf{Y}$  denote some normal  $N$ -dimensional observation vector such that:  $\mathbf{Y} = k\mu\mathbf{H} + \mathbf{X}$  and  $\mathbf{X} \sim \mathcal{N}(0, k^2\mathbf{I})$  with  $k \in \mathbb{R}^+$ ,  $\mu \in \mathbb{R}$ ,  $\mathbf{I}$  the identity matrix and  $\mathbf{H}$  some known vector that spans the known subspace  $\langle \mathbf{H} \rangle$ . Such a model*

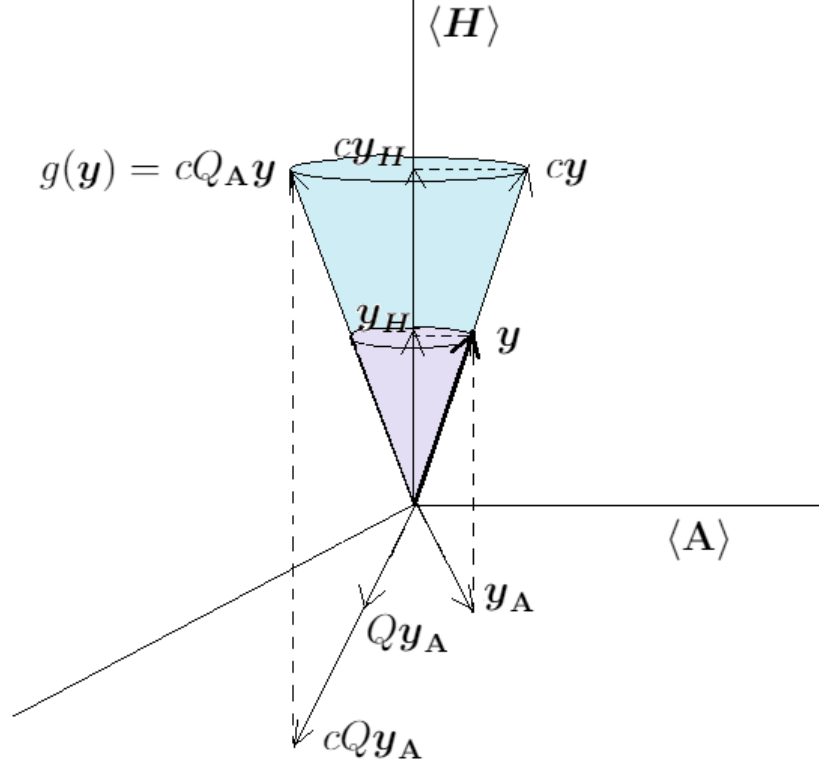


Figure 2.1: The geometry of the invariances of example 2.1.1

corresponds to a source that produces the signal  $\mu\mathbf{H}$  and some channel that introduces a white Gaussian noise  $\mathcal{N}(0, \mathbf{I})$  along with some unknown gain  $k$  and a rotation  $\mathbf{Q}_A$  of unknown angle. The rotation  $\mathbf{Q}_A$  is defined as  $\mathbf{Q}_A = \mathbf{U}_A \mathbf{Q} \mathbf{U}_A^T + \mathbf{P}_H$  with  $\mathbf{P}_A = \mathbf{U}_A \mathbf{U}_A^T$  the projection onto  $\langle \mathbf{A} \rangle$  and  $\mathbf{P}_H$  the projection matrix onto the subspace  $\langle \mathbf{H} \rangle$ .  $\mathbf{Q}$  is an orthogonal matrix that models a rotation in the subspace  $\langle \mathbf{A} \rangle$ . The subspace  $\langle \mathbf{A} \rangle$  is orthogonal to the subspace  $\langle \mathbf{H} \rangle$ . Accordingly, the observation vector is  $\mathbf{Y} \sim \mathcal{N}(k\mu\mathbf{H}, k^2\mathbf{Q}_A\mathbf{I}\mathbf{Q}_A^T)$  and more simply  $\mathbf{Y} \sim \mathcal{N}(k\mu\mathbf{H}, k^2\mathbf{I})$  because  $\mathbf{Q}_A\mathbf{Q}_A^T = \mathbf{Q}_A^T\mathbf{Q}_A = \mathbf{I}$ . Hence the testing hypothesis are:

$$\begin{cases} \mathcal{H}_0 : \mu \leq 0 \\ \mathcal{H}_1 : \mu > 0 \end{cases} \quad (2.1)$$

We assumed that the gain  $k$  and the rotation matrix  $\mathbf{Q}_A$  are unknown. Consequently, we need a group of transformation that is invariant to this lack of information, that is:

$$G = \{g : g(\mathbf{y}) = c(\mathbf{U}_A \mathbf{Q} \mathbf{U}_A^T + \mathbf{P}_H)\mathbf{y}, \text{ with } c \in \mathbb{R}^+\} \quad (2.2)$$

The geometry of the invariances is depicted in Figure 2.1. We can see that the small purple cone represents all the possible vectors  $g(\mathbf{y})$  with  $c = 1$  and  $g \in G$ . If we look at the big blue cone,  $c > 0$  both a rotation and a gain is operated. Similarly, the big cone represents all the possible vectors  $g(\mathbf{y})$  when  $c > 1$  and  $g \in G$ . To understand better how the measurement observations are invariant let us study the transformations that induces  $G$ . The observation vector  $\mathbf{y}$  can be written as the sum of its projection in the subspace  $\langle \mathbf{H} \rangle$  and  $\langle \mathbf{A} \rangle$  respectively  $\mathbf{y}_H$  and  $\mathbf{y}_A$ :  $\mathbf{y} = \mathbf{y}_A + \mathbf{y}_H$ , such that  $\mathbf{y}_A = \mathbf{P}_A\mathbf{y}$  and  $\mathbf{y}_H = \mathbf{P}_H\mathbf{y} = k\mu$ . If we consider  $g \in G$  such that  $g(\mathbf{y}) = c\mathbf{Q}_A\mathbf{y} = c(\mathbf{U}_A \mathbf{Q} \mathbf{U}_A^T + \mathbf{P}_H)\mathbf{y}$ , we can easily isolate two transformations:  $g(\mathbf{y}) =$

$c\mathbf{Q}\mathbf{y}_A + c\mathbf{y}_H$ . We can see in Figure 2.1 the different steps of the transformation  $g$ . First, the observation vector  $\mathbf{y}$  is projected onto  $\langle \mathbf{H} \rangle$  then multiplied by  $c$ . The second step is projecting  $\mathbf{y}$  onto  $\langle \mathbf{A} \rangle$  then operating a rotation  $\mathbf{Q}$  before multiplying by the same factor  $c$ . That is when an addition is made between the  $\langle \mathbf{A} \rangle$  component and the  $\langle \mathbf{H} \rangle$  component to obtain  $g(\mathbf{y})$ .

In Figure 2.1, the geometry of the group of transformations  $G$  is presented here as a cone, that is gathering all the possible observation vectors that will yield the same decision for Problem 2.1.

In this example, the group of transformation is a little bit intuitive, in others it can be too complicated to use intuition to find  $G$ . Next, definitions about invariant families and invariant detection problems will be given.

## 2.1.2 Invariant hypothesis-testing problems

In this subsection, we shall discuss two fundamental concepts of invariance; an invariant family of distributions and an invariant hypothesis testing problem. The first is a necessary condition to the second as we will see below. Hence, the following definition will be useful for definition 2.1.3 about invariant problems and can be found in [Borovkov, 1998, Definition 1, p. 281].

**Definition 2.1.2.** Let  $\mathbf{Y} \sim \{\mathbb{P}_\theta\}$  and  $G$  be a group of measurable transformations  $g$  of the space  $\mathbb{R}^n$  into itself. A family  $\{\mathbb{P}_\theta\}$  is invariant under  $G$  if for each  $g \in G$  and  $\theta \in \Theta$  there exists an element  $\theta_g \in \Theta$  such that:  $\mathbb{P}_{\theta_g}(\mathbf{Y} \in A) = \mathbb{P}_\theta(g\mathbf{Y} \in A)$ , for any Borel set  $A$  of  $\mathbb{R}^n$ .

The assumption about the measurability of  $g \in G$  is made to ensure that whenever  $\mathbf{Y}$  is a random variable,  $g\mathbf{Y}$  is also a random variable. The transformations  $\bar{g}$  of the space  $\Theta$  are defined by the equality  $\bar{g}\theta = \theta_g$  and constitute a group  $\bar{G}$  if there is a one-to-one correspondence between the parameter set  $\Theta$  and the family of distributions  $\{\mathbb{P}_\theta\}_{\theta \in \Theta}$  i.e.  $\mathbb{P}_{\theta_0} \neq \mathbb{P}_{\theta_1}$  if  $\theta_0 \neq \theta_1$ . Thus,  $\bar{G}$  is the corresponding group of transformation for the parameter space  $\Theta$  and is homomorphic to  $G$ .

**Example 2.1.2.** Let us consider the same decision problem (2.1) of example 2.1.1. The observation  $\mathbf{Y}$  follows the distribution  $\mathcal{N}(k\mu\mathbf{H}, k^2\mathbf{I})$ . The parameter  $\theta$  corresponds for the example to  $k\mu\mathbf{H}$  such that  $\mathbf{Y} = \theta + \mathbf{X}$ . Therefore, the family of distributions that needs to be invariant is  $\{\mathbb{P}_\theta\} = \mathcal{N}(\theta, k^2\mathbf{I})$ . Since  $g\mathbf{Y}$  follows the distribution  $\mathcal{N}(ck\mu\mathbf{H}, c^2k^2\mathbf{I})$  i.e.  $\mathcal{N}(c\theta, c^2k^2\mathbf{I})$ , the distribution of  $\mathbf{Y}$  is invariant to  $G$ . Consequently the family  $\{\mathbb{P}_\theta\}$  is invariant under  $G$  and the transformation of the parameters  $(k\mu, k^2\mathbf{I})$  is  $\bar{g}(k\mu, k^2\mathbf{I}) = (ck\mu, c^2k^2\mathbf{I})$ . It is worth noticing that the inverse element of  $G$  is  $g_{inv}(\mathbf{y}) = c^{-1}\mathbf{Q}_A^T\mathbf{y}$  and the identity element is  $g_{id}(\mathbf{y}) = \mathbf{Q}_A^*\mathbf{y}$ , with  $\mathbf{Q}_A^*$  of rotation angle  $2k\pi$  and  $k \in \mathbb{Z}$ .

**Definition 2.1.3.** We say that the problem of testing the hypothesis  $\mathcal{H}_1 = \{\theta \in \Theta_1\}$  against  $\mathcal{H}_0 = \{\theta \in \Theta_0\}$ , where  $\Theta_0 \cup \Theta_1 = \Theta$ , is invariant if the following two conditions are satisfied:

1. The family  $\{\mathbb{P}_\theta\}$  is invariant under  $G$ , the group of transformations.
2. The sets  $\Theta_1$  and  $\Theta_0$  are invariant under  $\bar{g} \in \bar{G}$ , that is  $\bar{g}\Theta_i = \Theta_i$  for  $i = 0, 1$ .

**Example 2.1.3.** We consider the same problem (2.1) from Example 2.1.1. On the basis of Example 2.1.2, the first condition of Definition 2.1.3 is verified. With regards to invariance of the parameter subsets  $\Theta_0 = \mathbb{R}^-$  and  $\Theta_1 = \mathbb{R}^{*+}$ , we know that the expression of the hypothesis for an observation  $g\mathbf{Y}$  is:

$$\begin{cases} \mathcal{H}_0 : ck\mu \leq 0 \\ \mathcal{H}_1 : ck\mu > 0 \end{cases} \quad (2.3)$$

Thus the parameter subsets after transformation are:

$$\begin{aligned} \bar{g}(\Theta_0) &= \mathbb{R}^- = \Theta_0 \\ \bar{g}(\Theta_1) &= \mathbb{R}^{*+} = \Theta_1 \end{aligned} \quad (2.4)$$

Consequently, according to Definition 2.1.3, the problem (2.1) is invariant with respect to  $G$ .

The formulation of the assumption about the invariant family implies that all the considered signals should follow the same parameterized distribution, which can be restrictive when the signals' distributions are assumed unknown. As for the second condition, it is quite understandable that the subsets  $\Theta_0$  and  $\Theta_1$  related to the hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$  we want to test, need to be invariant by the action of the corresponding group of transformation  $\bar{G}$  for the parameter space  $\Theta$ , in order for the hypothesis-testing problem to be invariant.

In the next section, maximal invariants, a notion that we have already spoken of when presenting the group of transformation, will be defined along with the closely interrelated orbits.

## 2.2 Maximal invariant and orbits

The notion of "maximal invariant statistic" may be confused with "sufficient statistic" because of their close meanings. Indeed, both proceed by a reduction of the sample space, but only the use of a maximal invariant statistic involves also shrinking the parameter space. The following definition of a maximal invariant is common to all the references that address the subject, for instance [Borovkov, 1998; Lehmann and Romano, 2005b; Scharf, 1991] etc.

**Definition 2.2.1.** For  $\mathbf{y} \in \mathbb{R}^n$ , a statistic  $\mathcal{M}(\mathbf{y})$  is said to be a maximal invariant to a group of transformation  $G$  if:

- $\mathcal{M}[g(\mathbf{y})] = \mathcal{M}(\mathbf{y})$ , for all  $g \in G$ .
- $\mathcal{M}(\mathbf{y}_1) = \mathcal{M}(\mathbf{y}_2) \Rightarrow \mathbf{y}_2 = g(\mathbf{y}_1)$ , for some  $g \in G$ .

With regards to orbits, they result from the natural equivalence relation between elements of the sample space. A relation defined by the group of transformation  $G$  itself.

**Definition 2.2.2.** *The orbits of  $G$  are the sets of equivalent points that constitute a partition of the sample space, and  $\mathbf{y}_1, \mathbf{y}_2$  are equivalent under  $G$  if  $\exists g \in G$  for which  $\mathbf{y}_2 = g(\mathbf{y}_1)$ .*

In other words, a point  $\mathbf{y}$  traces out an orbit  $\Upsilon$  as all transformations  $g$  of  $G$  are applied to it. Meaning that the orbit  $\Upsilon$  that includes  $\mathbf{y}$ , contains also the entire points  $g\mathbf{y}$  with  $g$  running through all  $G$ . This being said, the second condition of definition 2.2.1 about the invariant property of  $\mathcal{M}$  is equivalent to  $\mathcal{M}$  being constant on each orbit.

## 2.3 Invariant tests

If there are symmetries in both the sample and the parameter space, represented by the groups  $G$  and  $\bar{G}$ , it is natural to restrict attention to tests that are also invariant. Then, finding the "best" invariant test and more specifically a "Uniformly Most Powerful Invariant test" is the final step in an invariant process for solving a detection problem. First let us define invariant tests:

**Definition 2.3.1.** *Let  $G$  be a group of transformation. For  $\mathbf{y} \in \mathbb{R}^d$ , a hypothesis test  $\mathcal{T}(\mathbf{y})$  is said to be  $G$ -invariant if:  $\mathcal{T}[g(\mathbf{y})] = \mathcal{T}(\mathbf{y})$ , for all  $g \in G$ .*

It is important to know that every  $G$ -invariant test can be written as a function of a maximal invariant:  $\mathcal{T}(\mathbf{y}) = \mathcal{T}[\mathcal{M}(\mathbf{y})]$ , both related to the same group of transformation  $G$ . Moreover, Theorem 2.3.1 from [Scharf, 1991] gives a sufficient and necessary condition about the maximal invariant for a test  $\mathcal{T}$  to be invariant.

**Theorem 2.3.1.** *Let  $\mathcal{M}(\mathbf{y})$  be a maximal invariant with respect to  $G$ . Then, a necessary and sufficient condition for  $\mathcal{T}$  to be invariant is that it depends on  $\mathbf{Y}$  only through  $\mathcal{M}(\mathbf{y})$ ; that is there exists a function  $h$  for which  $\mathcal{T}(\mathbf{y}) = h[\mathcal{M}(\mathbf{y})]$  for all  $\mathbf{y}$ .*

The main result of Theorem 2.3.1 is that the restricted class of tests by invariance with respect to  $G$  i.e the class of all  $G$ -invariant tests is in fact the set of tests depending only on the maximal invariant statistic  $\mathcal{M}$ . Now, let us define the Uniformly Most Powerful Invariant test (UMPI).

**Definition 2.3.2.** *A test  $\mathcal{T}$  is Uniformly Most Powerful Invariant of size  $\alpha$  for testing  $\mathcal{H}_1 = \{\boldsymbol{\theta} \in \Theta_1\}$  against  $\mathcal{H}_0 = \{\boldsymbol{\theta} \in \Theta_0\}$  if it is  $G$ -invariant with  $\mathcal{T}(\mathbf{y}) = \mathcal{T}[\mathcal{M}(\mathbf{y})]$  and  $\mathbf{Y} \sim f_{\mathbf{Y}}^{\boldsymbol{\theta}}$  and if for any other  $G$ -invariant test  $\mathcal{T}'$  the following two items are verified:*

- $\sup_{\boldsymbol{\theta} \in \Theta_0} E_{\boldsymbol{\theta}} \mathcal{T}[\mathcal{M}(\mathbf{Y})] = \alpha$  and  $\sup_{\boldsymbol{\theta} \in \Theta_0} E_{\boldsymbol{\theta}} \mathcal{T}'[\mathbf{Y}] \leq \alpha$ .
- $E_{\boldsymbol{\theta}} \mathcal{T}[\mathcal{M}(\mathbf{Y})] \geq E_{\boldsymbol{\theta}} \mathcal{T}'(\mathbf{Y})$  for all  $\boldsymbol{\theta} \in \Theta_1$ .

The first item of theorem 2.3.2 concerns the constraint about the size of an UMPI test  $\mathcal{T}$  and guarantees that the competing tests are all of level  $\alpha$ . The second item concerns the power of the tests such that  $\mathcal{T}$  the UMPI test should have the highest power among all tests of level  $\alpha$ . In practice, in order to find that a  $G$ -invariant test  $\mathcal{T}$  is UMPI, the first option is to derive the likelihood ratio of the maximal invariant statistic while all the problem's parameters are known. Typically when the maximal invariant statistic takes only one value under each hypothesis. Analogously to simple hypothesis testing, the likelihood ratio of the maximal invariant is then compared to a threshold to guarantee a certain size. The second option occurs when the distributions of the maximal invariant under one or both hypotheses holds uncertainty similarly to composite hypothesis testing. In such a case, the maximal invariant-based likelihood ratio needs to be found strictly monotone and compared again to a threshold so that according to the theorem of Karlin Rubin the resulting test is UMPI. Both these two options require the knowledge of the probability density functions of the maximal invariant. Unfortunately, sometimes, finding a maximal invariant statistic and its density is hard to fulfill. That is when the third option can be considered: integrating over the elements  $g$  of the group of transformations  $G$  that leaves the problem invariant to obtain the ratio of the densities of the maximal invariant. This approach is described in [Wijsman, 1967] and [Stein, 1956] and does not require the knowledge of a maximal invariant statistic. It is presented from a statistical point of view in [Eaton, 1989] and [Kariya and Sinha, 1989], and applied to signal detection in [Gabriel and Kay, 2002]. On account of the theoretical complexity of this approach, seldom do we find it in engineering oriented works.

## 2.4 Limitations of the standard invariance

Throughout this chapter, we have presented different aspects of invariance. Thereby, it is clear enough that the standard concept of invariance as it is found in literature holds some limitations. These limitations are mainly related to definitions 2.1.2 and 2.1.3. In Definition 2.1.2 the condition for a family to be invariant imposes that signals have the same distribution, up to some parameter. Admittedly, with a different value of the parameter, but still the totality of the considered signals need to belong to the same family of distribution. In a context where signals are assumed to have unknown distributions with no guarantee of similarity, this assumption hinder the use of invariance and its interesting aspects of reduction and optimality. Regarding the second problematic aspect of the standard definition of an invariant hypothesis-testing problem, it is strongly linked to the parameter set. Indeed, ordinarily the hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$  induce corresponding parameter sets respectively  $\Theta_0$  and  $\Theta_1$ , which need to be invariant by the action of the group of transformation  $G$ . In a lot of scenarios this assumption is very much reasonable, but in others,  $\mathcal{H}_0$  and  $\mathcal{H}_1$  can be based on something other than  $\mathbb{R}$  subsets, for instance probabilistic events as we will see in Chapter 3 and 5.



## 2.5 A more general formulation of invariance

According to the standard notion of invariance that we can find in literature, a decision problem is invariant if the two conditions in Definition (2.1.3) are verified. In brief, a hypothesis testing problem is invariant if the family of distributions of the observations  $\{\mathbb{P}_\theta\}$  is invariant towards a group of transformations  $G$  and if the parameter sets  $\Theta_0$  and  $\Theta_1$  are invariant towards the associated group of transformation  $\overline{G}$  of the parameter set  $\Theta$ . Since the type of problems we target fall within the scope of the conditional model, we will tend to consider decision problems where the signal of interest is assumed random with unknown distribution and no intention to estimate it. Unfortunately, the classical definition of invariance cannot cope with this assumption, first, because the family of distributions cannot be defined like in Definition (2.1.3) as the distribution itself is not known. Secondly, due to the randomness of the signal, the nature of the hypothesis to test changes, and they become probabilistic events. These are the two challenges that makes the classical concept of invariance hard to apply in such contexts. At this point, the question we asked ourselves is: is there a family that can be defined, differently from (2.1.3), but that can still bring together these random signals into some kind of set that verifies invariance?

We consider the family:

$$\begin{aligned} \mathcal{F} &= \{\mathbb{P}_\mu : \mu \in \mathcal{D}\} \text{ such that} \\ \mathcal{D} &= \{ \text{The set of all probability distributions on } (\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N)) \} \end{aligned} \quad (2.5)$$

The  $\sigma$ -algebra of Borel sets on  $\mathbb{R}^N$  is  $\mathcal{B}(\mathbb{R}^N)$ . The observation model is:  $\mathbf{Y} = \Theta + \mathbf{X}$  with  $\Theta \sim \mu$  the signal of interest, assumed to be random of unknown distribution, and  $\mathbf{X} \sim f$  the noise of distribution  $f$ . Thus,  $\mathbb{P}_\mu = f * \mu$ , with  $*$  denoting the convolution. We consider a group of transformation  $G$ .

**Lemma 2.5.1.** *Let  $\mathbf{Y} \sim \mathcal{F}$  and  $G$  be a group of measurable transformations  $g$ . For  $A \in \mathcal{B}(\mathbb{R}^N)$ , the family  $\mathcal{F} = \{\mathbb{P}_\mu : \mu \in \mathcal{D}\}$  defined in Eq. (2.5) is invariant under any  $G$ .*

*Proof.* Proving that the family  $\mathcal{F}$  is invariant under the action of  $G$  means we need to prove that for  $g \in G$ , for all  $\mu \in \mathcal{D}$  there exists  $\mu' \in \mathcal{D}$  for any  $A \in \mathcal{B}(\mathbb{R}^N)$  that satisfies:

$$\mathbb{P}_{\mu'}(A) = \mathbb{P}_\mu(g^{-1}(A)) \quad (2.6)$$

First, we know that there exist  $\Theta$  and  $\mathbf{X}$  independent such that:

$$\mathbb{P}_\mu = \mathbb{P}(\Theta + \mathbf{X})^{-1} = f * \mu$$

with  $\mu = \mathbb{P}\Theta^{-1}$  the probability measure associated to  $\Theta$  and  $f = \mathbb{P}\mathbf{X}^{-1}$  the probability measure associated to  $\mathbf{X}$ . Hence, we have:

$$\mathbb{P}_\mu(g^{-1}(A)) = \mathbb{P}(g\Theta + g\mathbf{X})^{-1}(A)$$

On the one hand we can verify that:

$$\mathbb{P}(g\Theta)^{-1}(A) = \mu(g^{-1}(A))$$

On the other hand, we know that  $g\mathbf{X} \sim f$  and that  $g\Theta$  and  $g\mathbf{X}$  are independent. Consequently:

$$\mathbb{P}(g\Theta + g\mathbf{X})^{-1} = f * \mu g^{-1}$$

and we can write:

$$\mathbb{P}_\mu(g^{-1}(A)) = \mathbb{P}_{\mu g^{-1}}(A).$$

Therefore by choosing  $\mu' = \mu g^{-1} \in \mathcal{D}$ , the equality (2.6) is verified and the family  $\mathcal{F} = \{\mathbb{P}_\mu : \mu \in \mathcal{D}\}$  is invariant.  $\square$

Our invariant family  $\mathcal{F}$  is admittedly not conventional, but it allows to extend the initial definition of invariant family of distributions to signals that are assumed to be random of unknown distribution, and for that Lemma 2.5.1 is our first small contribution in this manuscript. Although we answered positively to the question we raised in the end of the first paragraph of this section thanks to Lemma 2.5.1, a supplementary question is yet to be asked: By considering null and alternative events instead of the usual null and alternative hypotheses is it still possible to construct invariant parameter subsets?

The second condition of definition (2.1.3) is clearly not reachable due to the probabilistic nature of the testing hypotheses. However, we can replace the conventional null hypothesis  $\mathcal{H}_0$  and alternative hypothesis  $\mathcal{H}_1$  by respectively the null event  $\Omega_0 = [\mathcal{M}(\Theta) \in \overline{\mathbf{E}}]$ , and the alternative event:  $\Omega_1 = [\mathcal{M}(\Theta) \in \mathbf{E}]$ . The maximal invariant statistic  $\mathcal{M}(\cdot)$  verifies items of Definition 2.2.1 and  $\mathbf{E}$  is a subspace of  $\mathbb{R}$ . The probabilistic events  $\Omega_0(\Theta)$  and  $\Omega_1(\Theta)$  are invariant if:

$$\begin{aligned} \Omega_0(g\Theta) &= \Omega_0(\Theta) \\ \Omega_1(g\Theta) &= \Omega_1(\Theta). \end{aligned} \tag{2.7}$$

Thanks to the properties of a maximal invariant  $\mathcal{M}(\cdot)$ , we have  $\Omega_0(g\Theta) = [\mathcal{M}(g\Theta) \in \overline{\mathbf{E}}] = [\mathcal{M}(\Theta) \in \overline{\mathbf{E}}]$ , thus  $\Omega_0(g\Theta) = \Omega_0(\Theta)$ . We can obtain similarly  $\Omega_1(g\Theta) = \Omega_1(\Theta)$ . Consequently, we proved what we can call, invariance of events:  $\Omega_1$  and  $\Omega_0$ . Although it is conceptually different from the second condition of Definition (2.1.3) [Borovkov, 1998], that is more of an invariance of the parameter subsets, it can be suitable for a wide range of decision problems that rely on the extended conditional model.



## Chapter 3

# A conditional approach: Random Distortion Testing framework

Random Distortion Testing (RDT) is a framework that was first introduced in 2013 by [Pastor and Nguyen, 2013a]. The spirit is the same as my thesis; trying to solve decision problems that are considered by the standard literature as hard to solve, while guaranteeing a certain optimality. The major reason is that in these decision problems, signals of interest are assumed random with unknown distribution. Such an assumption is not only much more realistic with regard to practical applications, but also much less constraining in terms of requisite knowledge about signals to be detected. The RDT framework introduces a relatively new optimality criteria;  $\gamma$ -MCP (Maximal Constant Power) and  $\gamma$ -MCCP (Maximal Constant Conditional Power) tests. These optimality criteria are non-integral, meaning that, as explained in the introduction, the involved probabilities are conditional, consequently, the RDT framework is part of the conditional model. On the basis of these optimality criteria, the resolution procedure requires two closely related stages in order to obtain the main theoretical result. The first stage involves to formulate the RDT problem under a deterministic assumption and is called Deterministic Distortion Testing (DDT) problem. Through the DDT problem, the optimality criterion  $\gamma$ -MCP is introduced and signals are assumed deterministic. Then, in the second stage, the  $\gamma$ -MCCP criterion is presented to solve the RDT problem with the help of the material from the DDT problem and some results that permits the transition from the DDT problem to the RDT problem. This chapter about the RDT framework is the third and last of the state of the art chapters. I restructured it such that I kept only the principal results required for the reasoning. In doing so, it appeared to me that distinguishing the two problems DDT and RDT, with their respective criterion of optimality, could be beneficial in terms of clarity.

## 3.1 The Deterministic Distortion Testing problem

### 3.1.1 Problem statement

Let  $\mathbf{Y}$  be the observation vector such that  $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{C})$ ,  $\boldsymbol{\theta} \in \mathbb{R}^N$  is the signal of interest and  $\mathbf{C}$  is a positive definite covariance matrix of the noise  $\mathbf{X}$ . The DDT problem addresses the problem of deciding whether  $\nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq \tau$ , the null hypothesis or  $\nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) > \tau$ , the alternative hypothesis, is true. The Mahalanobis norm  $\nu_{\mathbf{C}} : \mathbb{R}^N \rightarrow [0, \infty)$  is defined in  $\mathbb{R}^N$  with respect to  $\mathbf{C}$ , some positive definite  $N \times N$  covariance matrix and assigns to any  $\mathbf{x} \in \mathbb{R}^N$  the positive real number  $\nu_{\mathbf{C}}(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}$ . Such a formulation of the hypotheses is inspired by decision problems that lack robustness toward mismatches. Indeed, uncertainties about the signal can be hard to manage when the model is too strict. Hence, the incorporation of a tolerance  $\tau$ , to model these uncertainties, allows a certain flexibility towards the lack of knowledge about the signal of interest  $\boldsymbol{\theta}$ . Thus, the tolerance  $\tau$  is used as a bound on the insufficient information. The DDT problem is summarised in Eq. (3.1):

$$\text{DDT: } \begin{cases} \text{Observation: } \mathbf{Y} = \boldsymbol{\theta} + \mathbf{X}, \\ \text{Null hypothesis: } \nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq \tau, \\ \text{Alternative hypothesis: } \nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) > \tau. \end{cases} \quad (3.1)$$

The signal of interest  $\boldsymbol{\theta} \in \mathbb{R}^N$  is deterministic and the additive noise is assumed white Gaussian such that  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$ , with  $\mathbf{C} > 0$ . The signal used as a reference to measure the distance to the signal of interest  $\boldsymbol{\theta}$  is  $\boldsymbol{\theta}_0 \in \mathbb{R}^N$  and the tolerance is  $\tau \geq 0$ . The metric that measures up this distance is the Mahalanobis norm;  $\nu_{\mathbf{C}}(\cdot)$ . It is defined in the notation section. The notion of tolerance might seem abstract at first, but in fact the tolerance  $\tau$  is homogeneous to the square root of the whitened difference of energy between the signal  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}_0$ . Moreover assuming that the energy of the signal of interest  $\boldsymbol{\theta}$  is bounded depending on which hypothesis is verified is more flexible than assuming that the signal of interest is equal to a determined fixed model  $\boldsymbol{\theta}_0$  for example. Indeed, classically considering for instance that the hypotheses are  $\mathcal{H}_1: \boldsymbol{\theta} = \boldsymbol{\theta}_0$  and  $\mathcal{H}_0: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$  is not very realistic in practice.

The DDT problem is invariant under the action of a group. Let  $\mathcal{G}$  be the group of transformations that leaves the DDT problem invariant. The orbits (Def. 2.2.2) of  $\mathcal{G}$  can be defined as  $\Upsilon_{\rho} = \{\mathbf{y} \in \mathbb{R}^N : \nu_{\mathbf{C}}(\mathbf{y} - \boldsymbol{\theta}_0) = \rho\}$  with  $\rho \geq 0$ . The family of all the orbits  $\Upsilon_{\rho}$  with  $\rho \geq 0$  are denoted  $\mathfrak{F}$ . The group of transformations  $\mathcal{G}$  that leaves the DDT problem invariant under its action is:

$$\mathcal{G} = \{g: g(\mathbf{y}) = \Phi^{-1} R \Phi(\mathbf{y} - \boldsymbol{\theta}_0)\} \quad (3.2)$$

The matrix  $\Phi$  is defined as  $\Phi = \Delta^{-1/2} U^T$  with  $\Delta^{-1/2} = \text{diag}(\xi_1^{-1/2}, \xi_2^{-1/2}, \dots, \xi_N^{-1/2})$  and  $\xi_1, \xi_2, \dots, \xi_N$  are the eigenvalues of  $\mathbf{C}$ . The matrix  $U \in \mathbb{R}^{N \times N}$  is orthogonal such that  $\mathbf{C} = U \Delta U^T$  of  $\mathbf{C}$  and  $\Delta = \text{diag}(\xi_1, \xi_2, \dots, \xi_N)$ . The  $N \times N$  orthogonal matrix  $R$  corresponds to a rotation matrix.

After a reformulation of the DDT problem in accordance with the standard configuration of hypotheses testing problems, we can say that according to [Lehmann and Romano, 2005b] thresholding tests are Uniformly Most Powerful Invariant (UMPI)

tests for solving this detection problem and guarantee the greatest power among all tests that are invariant towards  $\mathcal{G}$ . In other words thresholding tests are optimal in the sense of Neyman-Pearson inside of the class of test that are  $\mathcal{G}$ -invariant. However, (UMPI) optimality criterion is not the strongest one can target by exploiting the invariances of the problem as we will see in the following section. Before presenting the principal results of the RDT framework for the DDT problem, some preliminary definitions and lemmas will be introduced, namely about this other optimality criteria,  $\gamma$ -MCP, that is shown in [Pastor and Nguyen, 2013a] to be stronger than the classical (UMPI) optimality criterion.

### 3.1.2 Main theoretical results

Let us consider  $\mathcal{T} : \mathbb{R}^N \rightarrow \{0, 1\}$  a non-randomised statistical test. Given any  $\boldsymbol{\theta} \in \mathbb{R}^N$  and any  $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{C})$ , the power function of a test  $\mathcal{T}$  is classically defined as:

$$\beta_{\mathcal{T}}(\boldsymbol{\theta}) \triangleq \mathbb{P}[\mathcal{T}(\mathbf{Y}) = 1] \quad (3.3)$$

On the basis of the definition of the power function of a test, The size of a test  $\mathcal{T}$  for the DDT problem is defined by

$$\alpha_{\mathcal{T}} \triangleq \sup_{\boldsymbol{\theta} \in \mathbb{R}^N : \nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq \tau} \beta_{\mathcal{T}}(\boldsymbol{\theta}) \quad (3.4)$$

As it is commonly admitted, a test  $\mathcal{T}$  is said to have a level  $\gamma \in (0, 1)$  for the DDT problem if  $\alpha_{\mathcal{T}} \leq \gamma$ .

The following definition is essential to define the  $\gamma$ -MCP property of a test.

**Definition 3.1.1.** *A test  $\mathcal{T}$  is said to have constant power function on an orbit  $\Upsilon_{\rho} \in \mathfrak{F}$  if  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) = \beta_{\mathcal{T}}(\boldsymbol{\theta}')$  for any  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Upsilon_{\rho}$ .*

The optimality criterion  $\gamma$ -MCP (Maximal Constant Power) for the DDT problem will be defined next in definition 3.1.2, according to [Pastor and Nguyen, 2013a].

**Definition 3.1.2.** *Given  $\tau \geq 0$  and  $\gamma \in (0, 1)$ , a test  $\mathcal{T}^*$  is said to be  $\gamma$ -MCP, more explicitly to be of level  $\gamma$  and have maximal constant power over the family  $\mathfrak{F}$ , if:*

- **[Level]**  $\mathcal{T}^*$  is of level  $\gamma$ .
- **[MCP]** For every  $\rho > \tau$ , the test  $\mathcal{T}^*$  has constant power function on  $\Upsilon_{\rho}$  and for any  $\boldsymbol{\theta} \in \Upsilon_{\rho}$  and any test  $\mathcal{T}$  of level  $\gamma$  with constant power function on  $\Upsilon_{\rho}$  we have:  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) \leq \beta_{\mathcal{T}^*}(\boldsymbol{\theta})$ .

According to [Pastor and Nguyen, 2013a], if a test is  $\gamma$ -MCP then it is (UMP) among the class of tests invariant towards  $\mathcal{G}$ . That is to say that the (UMPI) property of a test is a necessary condition for a test to be  $\gamma$ -MCP, and that the  $\gamma$ -MCP property is a sufficient condition for a test to be (UMPI). Thus, the  $\gamma$ -MCP criterion of optimality is even stronger than the optimality in the sense of (UMPI) tests.

Another interesting point about  $\gamma$ -MCP tests is that, in fact, they are very similar to the (UBCP): Uniformly Best Constant Power tests of Wald [Wald, 1943]. They both define optimality inside a certain family; the family of orbit  $\mathfrak{F}$  for the RDT framework and a family of surfaces in the space of observation for Wald's UBCP tests. There are also some discrepancies between the two, for instance the null tolerance  $\tau = 0$  for Wald's tests and more fundamentally the fact that the latter is dedicated for decision problems where the mean of a Gaussian distribution is tested, whereas the RDT problem is more general.

**Lemma 3.1.1.**  $\forall \gamma \in (0, 1), \forall \rho \geq 0$ ,  $\lambda_\gamma(\rho)$  is the unique solution in  $x$  of  $1 - \mathcal{R}(\rho, x) = \gamma$ , with  $\mathcal{R}(\rho, \cdot)$  the cumulative distribution function of the square root of any random variable that follows the non-central  $\chi^2$  distribution with  $N$  degrees of freedom and  $\rho^2$  the non-centrality parameter.

*Proof.*  $\mathcal{R}(\rho, \cdot)$  is a one-to-one mapping from  $[0, \infty)$  into  $[0, 1)$ . Thence, the existence and the uniqueness of  $\lambda_\gamma(\rho)$  for  $\gamma \in (0, 1]$ . □

Lemma 3.1.1 shows how to seek the thresholds  $\lambda_\gamma(\rho)$  for  $\rho \geq 0$  in order to lead the way for finding the expression of the threshold  $\lambda_\gamma(\tau)$  and consequently the threshold test  $\mathcal{T}_{\lambda_\gamma(\tau)}$ .

**Definition 3.1.3.** The thresholding test  $\mathcal{T}_{\lambda_\gamma(\tau)}$  is defined such that:

$$\mathcal{T}_{\lambda_\gamma(\tau)}(\mathbf{y}) = \begin{cases} 1 & \text{if } \nu_{\mathbf{C}}(\mathbf{y} - \boldsymbol{\theta}_0) > \lambda_\gamma(\tau) \\ 0 & \text{if } \nu_{\mathbf{C}}(\mathbf{y} - \boldsymbol{\theta}_0) \leq \lambda_\gamma(\tau) \end{cases} \quad (3.5)$$

with  $\lambda_\gamma(\tau)$  the unique solution in  $x$  of  $1 - \mathcal{R}(\tau, x) = \gamma$ .

Theorem 3.1.2 states that the test  $\mathcal{T}_{\lambda_\gamma(\tau)}$  from Definition 3.1.3 is  $\gamma$ -MCP. Hence, it is this section's principal result.

**Theorem 3.1.2.** For any  $\tau \geq 0$  and any  $\rho \in (0, 1)$ , the threshold test  $\mathcal{T}_{\lambda_\gamma(\tau)}$  is  $\gamma$ -MCP with size  $\gamma$  and power  $\beta_{\mathcal{T}_{\lambda_\gamma(\tau)}}(\boldsymbol{\theta}) = 1 - \mathcal{R}(\rho, \lambda_\gamma(\tau))$  for any  $\rho \geq 0$  and any  $\boldsymbol{\theta} \in \Upsilon_\rho$ .

*Proof.* See Appendix D. □

In the next section, the DDT problem will be extended such that the signal of interest  $\boldsymbol{\theta}$  is no longer deterministic, but a random signal  $\boldsymbol{\Theta}$  of unknown distribution is considered instead. The DDT problem is in fact a particular case of the RDT problem where  $\boldsymbol{\Theta} = \boldsymbol{\theta}$  almost surely.

## 3.2 The Random Distortion Testing problem

### 3.2.1 Problem statement

Let  $\mathbf{Y}, \Theta, \mathbf{X}$  be three  $N$ -dimensional real random vectors. The signal of interest  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  is assumed random of unknown distribution. We recall that  $\mathcal{M}(\Omega, \mathbb{R}^N)$  is the set of  $N$ -dimensional real random vectors defined on  $(\Omega, B)$ . The noise  $\mathbf{X}$  is assumed, analogously to the DDT problem, additive white and Gaussian.

The Random Distortion testing problem designated by RDT, is presented, according to [Pastor and Nguyen, 2013a], as:

$$\text{RDT:} \begin{cases} \text{Observation : } \mathbf{Y} = \Theta + \mathbf{X} & \begin{cases} \Theta \text{ and } \mathbf{X} \text{ independent,} \\ \Theta \in \mathcal{M}(\Omega, \mathbb{R}^N), \\ \mathbf{X} \sim \mathcal{N}(0, \mathbf{C}), \end{cases} \\ \text{Null event:} [\nu_{\mathbf{C}}(\Theta - \theta_0) \leq \tau], \\ \text{Alternative event:} [\nu_{\mathbf{C}}(\Theta - \theta_0) > \tau]. \end{cases} \quad (3.6)$$

The signal of interest  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  is assumed, for the RDT problem, random in opposite to the DDT problem of section 3.1. The additive noise is assumed again white Gaussian such that  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$  and  $\mathbf{C}$  is definite positive. The signal of reference  $\theta_0 \in \mathbb{R}^N$  is also kept in the problem statement and the tolerance  $\tau \geq 0$  is still used as a parameter to model the deviation from the reference signal  $\theta_0$ . Contrary to what one might think, the nature of the signal of interest is not the only difference between the DDT and RDT problem. The null event and the alternative events replace the standard null and alternative hypotheses in the RDT problem. Hence, the decision must be done by choosing between two complementary probabilistic events, instead of two hypotheses. However, this is still due to the randomness of the signal.

The RDT problem is also invariant under the action of a group. This group is also  $\mathcal{G} = \{g: g(\mathbf{y}) = \Phi^{-1}R\Phi(\mathbf{y} - \theta_0)\}$  with  $R$  and  $\Phi$  defined as in subsection 3.1.1. The orbits of  $\mathcal{G}$  are defined similarly such that  $\Upsilon_\rho = \{\mathbf{y} \in \mathbb{R}^N : \nu_{\mathbf{C}}(\mathbf{y} - \theta_0) = \rho\}$  with  $\rho \geq 0$ .

In the next subsection, a few definitions will be first introduced, then some supporting lemmas to help present the central theorem of the RDT framework about the  $\gamma$ -MCCP test for the RDT problem.

### 3.2.2 Main results

We consider  $\mathcal{T}: \mathbb{R}^N \rightarrow \{0, 1\}$  a non-randomised test. The [RDT] size of the test  $\mathcal{T}$  for the RDT problem is defined by

$$\alpha_{\mathcal{T}}^{[RDT]} = \sup_{\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N): \mathbb{P}[\nu_{\mathbf{C}}(\Theta - \theta_0) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \nu_{\mathbf{C}}(\Theta - \theta_0) \leq \tau] \quad (3.7)$$



Comparably to the standard terminology, a test  $\mathcal{T}$  is said to have a level  $\gamma$  for the RDT problem if  $\alpha_{\mathcal{T}}^{[RDT]} \leq \gamma$ .

The following definitions introduce the notions of conditional power function and constant conditional power function on an orbit  $\Upsilon_{\rho} \in \mathfrak{F}$ .

**Definition 3.2.1.** • Let  $\rho \geq 0$  and  $\forall \Upsilon_{\rho} \in \mathfrak{F}$ . Let  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  be independent of  $\mathbf{X}$ . The conditional power function of a test  $\mathcal{T}$  is defined as:  $\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}]$ .

- Let  $\rho \geq 0$ . Let  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  be independent of  $\mathbf{X}$ . A test  $\mathcal{T}$  is said to have constant conditional power function given  $\Theta \in \Upsilon_{\rho}$  if:

$$\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}] = \beta_{\mathcal{T}}(\theta), \forall \theta \in \Upsilon_{\rho}. \quad (3.8)$$

The next lemma states that the [RDT] size of a test defined in Eq. (3.7) for the RDT problem is equal to the classical size of a test defined for the DDT in Eq. (3.4).

**Lemma 3.2.1.** For any test  $\mathcal{T}$ , we have:

$$\alpha_{\mathcal{T}}^{[RDT]} = \alpha_{\mathcal{T}} \quad (3.9)$$

*Proof.* See Appendix E. □

Lemma 3.2.1 is not essential for proving the main result about the  $\gamma$ -MCCP property (theorem 3.2.2), but it helps showing the equivalence between the standard notion of size; also called probability of false alarm, and the size in the context of RDT framework.

The following definition relies on definition 3.2.1 to define tests of level  $\gamma$  and that have maximal constant conditional power function on  $\Upsilon_{\rho}$ .

**Definition 3.2.2.** Given  $\tau \geq 0$  and  $\gamma \in (0, 1)$ , a test  $\mathcal{T}^*$  is said to be  $\gamma$ -MCCP if:

- [**Level**]  $\mathcal{T}^*$  is of level  $\gamma$ .
- [**MCCP**] Given any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  and for  $\mathbb{P}_{\nu_{\mathbf{C}}(\Theta - \theta_0)}$ -almost every  $\rho > \tau$ , the test  $\mathcal{T}^*$  has constant conditional power function for any  $\Theta \in \Upsilon_{\rho}$  and for any test  $\mathcal{T}$  of level  $\gamma$  with constant conditional power function on  $\Upsilon_{\rho}$  we have:

$$\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}] \leq \mathbb{P}[\mathcal{T}^*(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}]$$

Theorem 3.2.2 is one of the main results about the RDT framework [Pastor and Nguyen, 2013a]. It states that the test  $\mathcal{T}_{\lambda_{\gamma}(\tau)}$  is  $\gamma$ -MCCP for the RDT problem.

**Theorem 3.2.2.** Given any  $\gamma \in (0, 1)$  and any  $\tau \geq 0$ , the test  $\mathcal{T}_{\lambda_{\gamma}(\tau)}$  is  $\gamma$ -MCCP with  $\mathbb{P}[\mathcal{T}_{\lambda_{\gamma}(\tau)}(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_{\rho}] = 1 - \mathcal{R}(\rho, \lambda_{\gamma}(\tau))$  for any given  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  and  $\mathbb{P}_{\nu_{\mathbf{C}}(\Theta - \theta_0)}$ -almost every  $\rho \geq 0$ .

*Proof.* See Appendix F. □

### 3.3 Conclusions

In conclusion, the RDT framework has proposed relatively new optimal tests;  $\gamma$ -MCP and  $\gamma$ -MCCP for the problem of deciding whether the Mahalanobis norm between a signal of interest and a reference signal is greater or lesser than a tolerance  $\tau$ . The most interesting about this framework is that the signal of interest is assumed unknown, be it for the DDT problem when  $\theta$  is deterministic but unknown, or the RDT problem when  $\Theta$  is random of unknown distribution. Moreover, the RDT framework bring together both the benefits of parametric and non-parametric methods. Indeed, the  $\gamma$ -MCP and  $\gamma$ -MCCP criteria of optimality are analogous to the statistical optimality of Neyman Pearson, which is normally hard to pursue when the signal of interest is unknown. Regarding the robustness of non-parametric approaches, it is attained too thanks to a formulation of the problem that is robust towards the possible deviations of the signal of interest due to the presence of interference for example or the imprecision of measuring devices etc.

With all these advantages and its particular scope of application, the RDT framework constituted the guiding principle for my thesis. Accordingly, and before our endeavour to generalise it to an even more general framework in chapter 5, we started by applying it in the case of a distributed detection in order to test its suitability to such scenarios. This is the subject of the next chapter.



# Chapter 4

## Distributed Random Distortion Testing

This chapter represents the first contribution of the manuscript. Throughout this chapter, we will consider a distributed scenario of detection where observations are collected by a network of  $p$  sensors. Just like chapter 3, the signals we want to detect are assumed of unknown distributions, in presence of additive white Gaussian noise in addition to interference of also unknown distributions. Choosing such assumptions can be very interesting in practice since it is less constraining. Moreover, literature does not provide enough references for such decision problems that can be found in a lot of applications. This chapter is also, somehow, a familiarisation with the RDT framework in order to pursue one of the goals of my thesis; generalising the RDT framework to the cases where the noise is not necessarily white Gaussian.

### 4.1 Problem statement

We consider  $p$  sensors, located in some region of interest (RoI) and equipped with a fusion center (FC). We want the sensors to decide collaboratively if a source is emitting a signal or not. Each sensor collects an observation vector  $\mathbf{Y}_i$  that is, under the alternative hypothesis, the sum of a signal  $\mathbf{S}_i$  emanating from the source, an additive white Gaussian noise  $\mathbf{X}_i$ , and some poorly known interference  $\boldsymbol{\xi}_i$ . Under the null hypothesis, the observation vector  $\mathbf{Y}_i$  of each sensor contains only the additive noise  $\mathbf{X}_i$  and interference  $\boldsymbol{\xi}_i$ . After collecting observation vectors by each sensor and sending it to the (FC) unprocessed or pre-processed, a decision must be made and the fusion center chooses which hypothesis is true or equivalently if a source is emitting or not. The observation vector received by the  $i^{th}$  sensor ( $i=1,2,\dots,p$ ) denoted by  $\mathbf{Y}_i$  is an element of  $\mathcal{M}(\Omega, \mathbb{R}^N)$ . Accordingly, the observation model is:

$$\mathbf{Y}_i = \epsilon \mathbf{S}_i + \boldsymbol{\xi}_i + \mathbf{X}_i \text{ with } \begin{cases} h_0 : \epsilon = 0 \\ h_1 : \epsilon = 1 \end{cases} \quad (4.1)$$

where  $\epsilon$  is a Bernoulli random variable such that  $\epsilon \in \{0, 1\}$ , assumed independent of  $\mathbf{S}_i$ ,  $\boldsymbol{\xi}_i$  and  $\mathbf{X}_i$ . It symbolises the presence or the absence of the source signal. The

signal  $\mathbf{S}_i$  emitted by a source, the interference  $\boldsymbol{\xi}_i$  and the noise  $\mathbf{X}_i$  are all elements of  $\mathcal{M}(\Omega, \mathbb{R}^N)$ ,  $\forall i \in \{1, 2, \dots, p\}$ . We assume that the noise  $\mathbf{X}_i$  and the signal  $\mathbf{S}_i$  are independent and that  $\mathbf{X}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$  where  $\boldsymbol{\Sigma}_i$  is a known  $N \times N$  positive-definite covariance matrix. The signal  $\mathbf{S}_i$  along with the interference  $\boldsymbol{\xi}_i$  are assumed random of unknown distribution.

Sensors are assumed to operate a linear transformation on data for dimensional reduction. Each sensor transformation is modeled by a full rank matrix  $\mathbf{A}_i \in \mathbb{R}^{n \times N}$  with  $n < N$ . In practice,  $\mathbf{A}_i$  can be a compression matrix, a projection matrix for subspace detection etc. We denote by  $\mathbf{Z}_i \in \mathcal{M}(\Omega, \mathbb{R}^n)$  the observation after transformation at each sensor  $i$  and we have:

$$\mathbf{Z}_i = \mathbf{A}_i \mathbf{Y}_i = \epsilon \mathbf{A}_i \mathbf{S}_i + \mathbf{A}_i \boldsymbol{\xi}_i + \mathbf{W}_i \quad (4.2)$$

with  $\mathbf{W}_i \sim \mathcal{N}(0, \mathbf{C}_i)$  and  $\mathbf{C}_i = \mathbf{A}_i \boldsymbol{\Sigma}_i \mathbf{A}_i^T$  the positive definite covariance matrix of the noise after transformation. The detection strategy involves deciding whether a signal of interest  $\mathbf{S}_i$  is present ( $\epsilon = 1$ ) or absent ( $\epsilon = 0$ ), without any assumption about the distribution of the signal  $\mathbf{S}_i$  nor the interference  $\boldsymbol{\xi}_i$ . Nonetheless, we assume the two following properties about the signal and the interference to be almost surely (a.s.) true:

$$\begin{cases} \mathcal{P}_0 : \forall i \quad \nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\xi}_i) \leq \tau \quad \forall i & (a.s.) \\ \mathcal{P}_1 : \forall i \quad \nu_{\mathbf{C}_i}(\mathbf{A}_i \mathbf{S}_i + \mathbf{A}_i \boldsymbol{\xi}_i) > \tau' \geq \tau \quad \forall i & (a.s.) \end{cases} \quad (4.3)$$

meaning that the probability for  $\mathcal{P}_0$  and  $\mathcal{P}_1$  to happen is almost surely equal to 1.  $\tau \in [0, \infty)$  the tolerance is introduced to model the effect of the poorly known interference. Similarly the parameter  $\tau' \in [\tau, \infty)$  is used to model the only available knowledge about the signal of interest  $\mathbf{S}_i$ . The Mahalanobis norm  $\nu_{\mathbf{C}} : \mathbb{R}^n \rightarrow [0, \infty)$  defined in  $\mathbb{R}^n$  with respect to  $\mathbf{C}$ , some positive definite  $n \times n$  covariance matrix, assigns to any  $\mathbf{x} \in \mathbb{R}^n$  the positive real number  $\nu_{\mathbf{C}}(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}$ . The Mahalanobis norm  $\nu_{\mathbf{C}_i}$  in (4.3), allows to measure a distance while taking into account the noise covariance matrix  $\mathbf{C}_i$  in our case. Introducing properties  $\mathcal{P}_0$  and  $\mathcal{P}_1$  is a way of bounding the lack of knowledge about the signal and the interference. Indeed, assuming that we do not know the distributions of the signal  $\mathbf{S}_i$  and the interference  $\boldsymbol{\xi}_i$  might seem very inconvenient and prevents us from applying standard approaches like the the Uniformly Most Powerful tests or any other test with optimality criteria (see chapter 1). In this vein, the two properties in (4.3) introduce just enough information to permit us to reformulate our initial hypothesis testing problem (4.1) into a more solvable problem, at least in the sense of the Random Distortion Testing (RDT) framework (see chapter 3). Considering  $\mathcal{P}_0$  and  $\mathcal{P}_1$  remains less binding in practice than having to know the signal and/or interference subspace or distribution.

## 4.2 Reformulation of the hypothesis testing problem

Before introducing the different distributed configurations we would like to consider and before trying to solve the decision problem for each configuration, we will need to rewrite our initial decision problem (4.1, 4.2). Indeed, by assuming true almost surely the two properties  $\mathcal{P}_0$  and  $\mathcal{P}_1$  about interference and signal energy, we will be able to cast our initial problem in the terms of the RDT framework. We define the signal

$\Theta_i \in \mathcal{M}(\Omega, \mathbb{R}^N)$  as the sum of the interference  $\xi_i$  perceived by the  $i^{\text{th}}$  sensor and a signal of interest  $\mathbf{S}_i \in \mathcal{M}(\Omega, \mathbb{R}^N)$  if present:

$$\Theta_i = \epsilon \mathbf{S}_i + \xi_i \quad (4.4)$$

The following proposition, whose proof comes next and is based on standard arguments in probability theory, restates Problem (4.1) when we assume  $\mathcal{P}_0$  and  $\mathcal{P}_1$ .

**Proposition 4.2.1.** *Assuming  $\mathcal{P}_0$  and  $\mathcal{P}_1$ , problem (4.1) can be equivalently stated as:*

$$\begin{cases} \text{Observation : } \mathbf{Z}_i = \mathbf{A}_i \Theta_i + \mathbf{W}_i \\ \mathcal{H}_0 : \nu_{\mathbf{C}_i}(\mathbf{A}_i \Theta_i) \leq \tau \quad \forall i \\ \mathcal{H}_1 : \nu_{\mathbf{C}_i}(\mathbf{A}_i \Theta_i) > \tau' \geq \tau \quad \forall i \end{cases} \quad (4.5)$$

*Proof.* See Appendix A. □

Consequently, based on Proposition (4.2.1), there is an equivalence between problem (4.1) and (4.5), while assuming  $\mathcal{P}_0$  and  $\mathcal{P}_1$ :

$$[\epsilon = 0] = [\nu_{\mathbf{C}_i}(\mathbf{A}_i \Theta_i) \leq \tau] \quad (\text{a-s}) \quad (4.6)$$

$$[\epsilon = 1] = [\nu_{\mathbf{C}_i}(\mathbf{A}_i \Theta_i) > \tau'] \quad (\text{a-s}) \quad (4.7)$$

## 4.3 Multi-sensors configurations

Sending raw data, i.e the observation as it is received by the sensors, to the fusion center is what we call a centralised configuration [Varshney, 2012]. It is known to be optimal and is usually used to avoid information loss, even if it requires heavy data transmission, in opposition to distributed configurations where the traffic of data is lighter. For distributed configurations, when a fusion center is available, the data is preprocessed at a local scale before being sent to the fusion center where the decision is made. On another note, when no fusion center is used, the sensors take their decision straightforwardly or after consulting each other. As for us, we consider a network of  $p$  sensors with a fusion center. We study the case where all the observations  $\mathbf{Y}_i$  are sent to the fusion center so that the test statistic can be computed without any performance loss. Then, we will examine the distributed configuration to avoid high communication cost, where local test statistics are calculated by each sensor before being sent to the fusion center where the decision is made. Later on, we will compare the two configurations, after solving the detection problems with the help of the RDT framework, in terms of probability of detection.

### 4.3.1 The centralised configuration

In this subsection, we describe the centralised configuration when the sensors observe data according to the model in equations (4.1,4.2). The observations  $\mathbf{Z}_i$  are collected

from the sensors and sent to the fusion center where the test statistic is computed according to chapter 3. At the fusion center, the detection problem can be formulated as:

$$\text{Observation : } \mathbf{Z} = \epsilon \mathbf{A} \mathbf{S} + \mathbf{A} \boldsymbol{\xi} + \mathbf{W} \text{ with } \begin{cases} h_0^{\text{FC}} : \epsilon = 0 \\ h_1^{\text{FC}} : \epsilon = 1 \end{cases} \quad (4.8)$$

with:

$$\begin{cases} \mathbf{Z} = [\mathbf{Z}_1^T, \dots, \mathbf{Z}_p^T]^T, \quad \mathbf{Z} \in \mathbb{R}^{np} \\ \mathbf{W} = [\mathbf{W}_1^T, \dots, \mathbf{W}_p^T]^T, \quad \mathbf{W} \in \mathbb{R}^{np} \\ \mathbf{S} = [\mathbf{S}_1^T, \dots, \mathbf{S}_p^T]^T, \quad \mathbf{S} \in \mathbb{R}^{Np} \\ \boldsymbol{\xi} = [\boldsymbol{\xi}_1^T, \dots, \boldsymbol{\xi}_p^T]^T, \quad \boldsymbol{\xi} \in \mathbb{R}^{Np} \\ \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{A}_p \end{pmatrix}, \quad \mathbf{A} \in \mathbb{R}^{np \times Np} \end{cases} \quad (4.9)$$

The centralised detection problem in (4.8) is a scaled version of the one-sensor case (4.1). Properties  $\mathcal{P}_0$  and  $\mathcal{P}_1$  (4.3) induce that, at the fusion center, we can consider the following properties:

$$\begin{cases} \mathcal{P}_0^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A} \boldsymbol{\xi}) \leq p\tau & (a.s.) \\ \mathcal{P}_1^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A} \mathbf{S} + \mathbf{A} \boldsymbol{\xi}) > p\tau' \geq p\tau & (a.s.) \end{cases} \quad (4.10)$$

Therefore, analogously to the one-sensor based problem, Problem (4.8) can be re-framed as an RDT problem according to Proposition (4.3.1). Proof of Proposition (4.3.1) relies on the same arguments as those used to demonstrate Proposition (4.2.1). We define  $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{Np})$  such that  $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1^T, \dots, \boldsymbol{\Theta}_p^T]^T = \epsilon \mathbf{S} + \boldsymbol{\xi}$  and  $\mathbf{C} = \text{blkdiag}(\mathbf{C}_1, \dots, \mathbf{C}_p) \in \mathbb{R}^{np \times np}$  with  $\mathbf{C}_i = \mathbf{A}_i \boldsymbol{\Sigma}_i \mathbf{A}_i^T$ .

**Proposition 4.3.1.** *Assuming  $\mathcal{P}_0^{\text{FC}}$  and  $\mathcal{P}_1^{\text{FC}}$  of (4.3.1), Problem (4.8) can be equivalently stated as:*

$$\begin{cases} \text{Observation : } \mathbf{Z} = \mathbf{A} \boldsymbol{\Theta} + \mathbf{W} \\ \mathcal{H}_0^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A} \boldsymbol{\Theta}) \leq p\tau \\ \mathcal{H}_1^{\text{FC}} : \nu_{\mathbf{C}}(\mathbf{A} \boldsymbol{\Theta}) > p\tau' \geq p\tau \end{cases} \quad (4.11)$$

*Proof.* See Appendix A. □

According to Proposition (4.3.1), there is an equivalence between Problem (4.8) and (4.11), while assuming  $\mathcal{P}_0^{\text{FC}}$  and  $\mathcal{P}_1^{\text{FC}}$ :

$$[\epsilon = 0] = [\nu_{\mathbf{C}}(\mathbf{A} \boldsymbol{\Theta}) \leq p\tau] \quad (\text{a-s}) \quad (4.12)$$

$$[\epsilon = 1] = [\nu_{\mathbf{C}}(\mathbf{A} \boldsymbol{\Theta}) > p\tau'] \quad (\text{a-s}) \quad (4.13)$$

### 4.3.2 The distributed configuration

The second configuration that we want to study in this chapter is a distributed configuration with a fusion center. We consider a fusion center in communication with  $p$

sensors that collect a  $n$  observation vector  $\mathbf{Y}_i$ . The purpose is to minimise the quantity of information exchanged between the fusion center and the sensors. Thus we intend to send to the fusion center the vital minimum information that can be sent: a scalar. This scalar must contain as much information as possible or at least keep performance losses to a minimum. We keep assuming that the two properties  $\mathcal{P}_0^{\text{FC}}$  and  $\mathcal{P}_1^{\text{FC}}$  are almost surely true. We consider that we want to solve the same detection problem as in (4.8) for the centralised configuration, only, with an additive constraint about the communication burden between the fusion center and the sensors.

The problem, as it is exposed, is quite different from classical approaches, where the signals to be detected are assumed deterministic with unknown parameters that are estimated or assigned priors. Indeed, in this approach, no assumption about the signals nor the interferences distribution is made. This extends the detection problem to the substantial class of all random vectors with unknown distribution and bounded energy.

The next step is to find a test that is optimal according to a certain criterion to be defined. As far as we know, with these assumptions, only the Random Distortion Testing framework (see chapter 3) can provide an optimal test as we will see in Section 4.4.

## 4.4 RDT for multi-sensors configurations

In standard statistical inference, and more particularly in likelihood theory, when the signal distribution is unknown, no Uniformly Most Powerful (UMP) test exists, therefore no optimality in the sense of Neyman-Pearson or Karlin-Rubin is targeted. In this respect, the RDT framework turns out to be useful by proposing an alternative optimality criterion the maximal constant conditional power inside a new class of tests called coherent class of tests. For an appropriate use of the RDT approach to solve our problem (4.11), let us recall some basics of the approach that we will adapt to our context.

### 4.4.1 Theoretical material

We consider the following observation model:

$$\mathbf{\Gamma} = \mathbf{\Xi} + \mathbf{N} \quad (4.14)$$

with  $\mathbf{\Xi}, \mathbf{N}, \mathbf{\Gamma} \in \mathcal{M}(\Omega, \mathbb{R}^D)$ ,  $\mathbf{N} \sim \mathcal{N}(0, \mathbf{J})$  and  $\mathbf{J}$  is a positive definite matrix. The hypotheses are defined as:

$$\begin{cases} \mathcal{H}_0^* : \nu_{\mathbf{K}}(\mathbf{B}\mathbf{\Xi}) \leq \tau \\ \mathcal{H}_1^* : \nu_{\mathbf{K}}(\mathbf{B}\mathbf{\Xi}) > \tau' \geq \tau \end{cases} \quad (4.15)$$

where  $\mathbf{B} \in \mathbb{R}^{d \times D}$  is a full rank matrix representing the linear preprocessing of the observations and  $\mathbf{K} = \mathbf{B}\mathbf{J}\mathbf{B}^T$ .



The problem is to decide whether  $\nu_{\mathbf{K}}(\mathbf{B}\Xi(\omega)) \leq \tau$  or  $\nu_{\mathbf{K}}(\mathbf{B}\Xi(\omega)) > \tau'$  given  $\Gamma(\omega)$  for an unknown  $\omega \in \Omega$ . Note that Problem (4.11) is a particular case of Problem (4.15).

We define the size  $\alpha_{\mathcal{T}}$  of a test  $\mathcal{T} : \mathbb{R}^D \rightarrow \{0, 1\}$  as:

$$\alpha_{\mathcal{T}} = \sup_{\Xi \in \mathcal{M}(\Omega, \mathbb{R}^D) : \mathbb{P}[\nu_{\mathbf{K}}(\mathbf{B}\Xi) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\Gamma) = 1 \mid \nu_{\mathbf{K}}(\mathbf{B}\Xi) \leq \tau] \quad (4.16)$$

We say that  $\mathcal{T}$  has level (resp. size)  $\gamma$  if  $\alpha_{\mathcal{T}} \leq \gamma$  (resp.  $\alpha_{\mathcal{T}} = \gamma$ ). Similarly, we can define the power  $\beta_{\mathcal{T}}(\Xi)$  of  $\mathcal{T}$  for any  $\Xi$  such that  $\mathbb{P}[\nu_{\mathbf{K}}(\mathbf{B}\Xi) > \tau'] \neq 0$  by:

$$\beta_{\mathcal{T}}(\Xi) = \mathbb{P}[\mathcal{T}(\Gamma) = 1 \mid \nu_{\mathbf{K}}(\mathbf{B}\Xi) > \tau'] \quad (4.17)$$

As it is commonly done when no optimal test can be found inside the set of all possible tests, we restrict our attention to a class of tests invariant to the group of transformations that leaves the problem invariant. In this respect, the following definition is provided [Pastor and Socheleau, 2018]:

**Definition 4.4.1** (Coherent test). *A test  $\mathcal{T} : \mathbb{R}^D \rightarrow \{0, 1\}$  is said to be a coherent test if:*

- 1)  $\mathcal{T}$  is invariant: For any  $(\Gamma, \Gamma') \in \mathbb{R}^D \times \mathbb{R}^D$  such that  $\mathbf{B}\Gamma = \mathbf{B}\Gamma'$ ,  $\mathcal{T}(\Gamma) = \mathcal{T}(\Gamma')$ .
- 2)  $\mathcal{T}$  has a constant conditional power: Given  $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^D)$  independent of  $\mathbf{N}$  and for almost every  $\rho > \tau$ ,  $\mathbb{P}[\mathcal{T}(\Gamma) = 1 \mid \nu_{\mathbf{K}}(\mathbf{B}\Xi) = \rho]$  is independent of  $\mathbb{P}_{\nu_{\mathbf{K}}(\mathbf{B}\Xi)}$ .

The class of all coherent tests with level  $\gamma$  is denoted by  $\mathcal{C}_{\gamma}$ . To define optimal tests within this class, we recall the notions of pre-order and maximality from chapter (3):

**Proposition 4.4.1.** *Pre-order: Given two tests  $\mathcal{T}$  and  $\mathcal{T}^*$  of  $\mathcal{C}_{\gamma}$ , we have  $\mathcal{T} \leq \mathcal{T}^*$  if, given any  $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^D)$  and  $\rho \in (\tau, \infty)$ :*

$$\mathbb{P}[\mathcal{T}(\Gamma) = 1 \mid \nu_{\mathbf{K}}(\mathbf{B}\Xi) = \rho] \leq \mathbb{P}[\mathcal{T}^*(\Gamma) = 1 \mid \nu_{\mathbf{K}}(\mathbf{B}\Xi) = \rho]$$

**Proposition 4.4.2.** *Maximality: For any  $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^D)$  and any  $\mathcal{T} \in \mathcal{C}_{\gamma}$ ,  $\exists \mathcal{T}^* \in \mathcal{C}_{\gamma}$  such that:  $\mathcal{T} \leq \mathcal{T}^*$ .*

The existence of a maximal test in  $\mathcal{C}_{\gamma}$  is established for the detection problem (4.15) via the next theorem:

**Theorem 4.4.1.** [Pastor and Nguyen, 2013a; Pastor and Socheleau, 2018] *Given a positive definite covariance matrix  $\mathbf{K}$ , let  $\nu_{\mathbf{K}} : \mathbb{R}^d \rightarrow [0, \infty)$  be the norm defined for each  $\mathbf{V} \in \mathbb{R}^d$  by setting  $\nu_{\mathbf{K}}(\mathbf{V}) = \sqrt{\mathbf{V}^T \mathbf{K}^{-1} \mathbf{V}}$ , with  $\mathbf{K} = \mathbf{B}\mathbf{J}\mathbf{B}^T$ . Given  $\gamma \in (0, 1)$  and  $\tau \geq 0$ , the test  $\mathcal{T}^* : \mathbb{R}^D \rightarrow \{0, 1\}$  defined for each  $\mathbf{x} \in \mathbb{R}^D$  by:*

$$\mathcal{T}^*(\mathbf{x}) = \begin{cases} 1 & \text{if } \nu_{\mathbf{K}}(\mathbf{B}\mathbf{x}) > \lambda(\tau, \gamma) \\ 0 & \text{otherwise} \end{cases} \quad (4.18)$$

*is maximal in  $(\mathcal{C}_{\gamma}, \leq)$  with size  $\gamma$  and for any  $\rho \in (0, \infty)$ :*

$$\mathbb{P}[\mathcal{T}^*(\Gamma) = 1 \mid \nu_{\mathbf{K}}(\mathbf{B}\Xi) = \rho] = Q_{d/2}(\rho, \lambda(\tau, \gamma)). \quad (4.19)$$

Theorem (4.4.1) enables us to design optimal tests in the sense of the RDT framework for problem (4.11) in case of a centralised configuration and a distributed configuration.

#### 4.4.2 Optimality for centralised configuration

By theorem (4.4.1), the optimal test  $\mathcal{T}_{FC} : \mathbb{R}^{np} \rightarrow \{0, 1\}$  inside the class of tests  $\mathcal{C}_\gamma$ , for the centralised configuration and the detection problem (4.8) is defined for every  $\mathbf{z} \in \mathbb{R}^{np}$  by:

$$\mathcal{T}_{FC}(\mathbf{z}) = \begin{cases} 1 & \text{if } \nu_{\mathbf{C}}(\mathbf{z}) > \lambda_{FC} \\ 0 & \text{otherwise} \end{cases} \quad (4.20)$$

The threshold  $\lambda_{FC}$  is the unique solution in  $\eta$  to the equation:  $Q_{np/2}(p\tau, \eta) = \gamma$ , where  $\gamma \in (0, 1)$  is the level of the test.  $Q_{np/2}$  is the generalized Marcum function defined for any pair  $(a, b) \in [0, \infty) \times [0, \infty)$  such that  $Q_{np/2}(a, b) = 1 - \mathbb{F}_{\chi_{np}^2(a^2)}(b^2)$  where  $\mathbb{F}_{\chi_{np}^2(a^2)}$  is the cumulative distribution function of the non-central Chi-square distribution:  $\chi_{np}^2(a^2)$  with  $np$  degree of freedom and  $a^2$  the non-centrality parameter.

Thanks to the equivalence between the initial problem (4.8) and problem (4.11) we can compute the detection probability  $P_{\text{det}}(\mathcal{T}_{FC})$  and the false alarm probability  $P_{\text{FA}}(\mathcal{T}_{FC})$  of the test  $\mathcal{T}_{FC}$ .

**Proposition 4.4.3.** *With the same notation as above, the detection probability and the false alarm probability are bounded such that:*

$$P_{\text{FA}}(\mathcal{T}_{FC}) \leq \gamma \quad (4.21)$$

$$P_{\text{det}}(\mathcal{T}_{FC}) \geq Q_{np/2}(p\tau', \lambda_{FC}) \quad (4.22)$$

*Proof.* The size  $\alpha_{\mathcal{T}_{FC}}$  of the RDT test  $\mathcal{T}_{FC}$  can be defined for problem (4.11) as:

$$\begin{aligned} \alpha_{\mathcal{T}_{FC}} &= \sup_{\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N) : \mathbb{P}[\nu_{\mathbf{C}}(\mathbf{A}\Theta) \leq p\tau] \neq 0} \mathbb{P}[\mathcal{T}_{FC}(\Theta + \mathbf{X}) = 1 \mid \nu_{\mathbf{C}}(\mathbf{A}\Theta) \leq p\tau] \\ &= \gamma \end{aligned} \quad (4.23)$$

Notice that the size of the test  $\mathcal{T}_{FC}$  for problem (4.11) is equal to the level  $\gamma$  according to theorem (4.4.1). With respect to the false alarm probability  $P_{\text{FA}}$  of the initial problem (4.8), we have:

$$\begin{aligned} P_{\text{FA}}(\mathcal{T}_{FC}) &\triangleq \mathbb{P}[\mathcal{T}_{FC}(\Theta + \mathbf{X}) = 1] \text{ under } h_0^{FC} \\ &= \mathbb{P}[\nu_{\mathbf{C}}(\mathbf{A}\xi + \mathbf{W}) > \lambda_{FC}] \end{aligned} \quad (4.24)$$

Now, we want to compare the two analogous quantities  $P_{\text{FA}}$  and  $\alpha_{\mathcal{T}_{FC}}$  knowing that the second can be seen as the worst case of the first.

Let  $\Theta = \epsilon\mathbf{S} + \xi$  such that  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  and  $\mathbb{P}[\nu_{\mathbf{C}}(\mathbf{A}\Theta) \leq p\tau] = \mathbb{P}[\epsilon = 0] \neq 0$ , hence, using Equation (4.23):

$$\begin{aligned} &\mathbb{P}[\nu_{\mathbf{C}}(\mathbf{A}\Theta + \mathbf{W}) > \lambda_{FC} \mid \nu_{\mathbf{C}}(\mathbf{A}\Theta) \leq p\tau] \leq \gamma \\ \Leftrightarrow &\mathbb{P}[\nu_{\mathbf{C}}(\epsilon\mathbf{A}\mathbf{S} + \mathbf{A}\xi + \mathbf{W}) > \lambda_{FC} \mid \epsilon = 0] \leq \gamma \\ \Leftrightarrow &\mathbb{P}[\nu_{\mathbf{C}}(\mathbf{A}\xi + \mathbf{W}) > \lambda_{FC}] \leq \gamma \\ \Rightarrow &P_{\text{FA}}(\mathcal{T}_{FC}) \leq \gamma \end{aligned} \quad (4.25)$$

Therefore, we prove that the false alarm probability of our initial detection problem, in case of a centralized configuration with a fusion center, stays below the level  $\gamma$  which allows the respect of the constraint about the level of the test and guarantees optimality in a sense similar to Neyman Pearson.

Now, let us compare again two other analogous quantities:  $\beta_{\mathcal{T}_{FC}}$  and  $P_{det}$  in case of a centralized configuration with a fusion center.

By using the RDT formalism, we define the power of the test  $\mathcal{T}_{FC}$  for  $p$  sensors at the fusion center as:

$$\beta_{\mathcal{T}_{FC}}(\mathbf{z}) = \mathbb{P}[\mathcal{T}_{FC}(\mathbf{z}) = 1 \mid \nu_C(\mathbf{A}\Theta) > p\tau'] \quad (4.26)$$

Let us note that:

$$\begin{aligned} \beta_{\mathcal{T}_{FC}}(\mathbf{z}) &= \mathbb{P}[\nu_C(\epsilon\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi} + \mathbf{W}) > \lambda_{FC} \mid \nu_C(\mathbf{A}\Theta) > p\tau'] \\ &\stackrel{(a)}{=} \frac{\mathbb{P}([\nu_C(\epsilon\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi} + \mathbf{W}) > \lambda_{FC}] \cap [\nu_C(\mathbf{A}\Theta) > p\tau'])}{\mathbb{P}[\nu_C(\mathbf{A}\Theta) > p\tau']} \\ &\stackrel{(b)}{=} \frac{\mathbb{P}([\nu_C(\epsilon\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi} + \mathbf{W}) > \lambda_{FC}] \cap [\epsilon=1])}{\mathbb{P}[\epsilon=1]} \\ &\stackrel{(c)}{=} \frac{\mathbb{P}[\nu_C(\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi} + \mathbf{W}) > \lambda_{FC}] \mathbb{P}[\epsilon=1]}{\mathbb{P}[\epsilon=1]} \\ &= \mathbb{P}[\nu_C(\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi} + \mathbf{W}) > \lambda_{FC}] \end{aligned} \quad (4.27)$$

In (a) we used the Bayes theorem, (b) follows from Proposition (4.3.1) and (c) holds since  $\epsilon$  is independent of  $\mathbf{S}, \boldsymbol{\xi}$  and  $\mathbf{X}$ .

The power function  $\beta_{\mathcal{T}_{FC}}(\mathbf{z})$  of the detection problem in (4.11) can be lower-bounded according to Equation (4.19):

$$\beta_{\mathcal{T}_{FC}}(\mathbf{z}) \geq Q_{np/2}(p\tau', \lambda_{FC}) \quad (4.28)$$

Regarding the probability of detection of the initial problem presented in (4.8), it can be defined as:

$$\begin{aligned} P_{det}(\mathcal{T}_{FC}) &\triangleq \mathbb{P}[\mathcal{T}_{FC}(\mathbf{z}) = 1] \text{ under } h_1^{FC} \\ &= \mathbb{P}[\nu_C(\mathbf{A}\mathbf{S} + \mathbf{A}\boldsymbol{\xi} + \mathbf{W}) > \lambda_{FC}] \end{aligned} \quad (4.29)$$

Then, it appears that the probability of detection of our initial problem (4.8) is equal to the RDT power function:  $\beta_{\mathcal{T}_{FC}}(\mathbf{z}) = P_{det}(\mathcal{T}_{FC})$ . Consequently  $P_{det}(\mathcal{T}_{FC})$  can be lower-bounded too, i.e.  $P_{det}(\mathcal{T}_{FC}) \geq Q_{np/2}(p\tau', \lambda_{FC})$ , which concludes the demonstration.  $\square$

Next, we present the one-sensor based optimal test, in the sense of the RDT framework, in order to numerically compare it to distributed configurations in terms of detection performance.

Let  $\mathcal{T}_i : \mathbb{R}^n \rightarrow \{0, 1\}$  be the optimal test inside the class of tests  $\mathcal{C}_\gamma$  that the  $i^{\text{th}}$  sensor may use to solve the one-sensor based testing problem (4.1) with guaranteed level of false alarm  $\gamma$  at a local scale. This test is defined according to Theorem 4.4.1 for each  $\mathbf{z}_i \in \mathbb{R}^n$  as:

$$\mathcal{T}_i(\mathbf{z}_i) = \begin{cases} 1 & \text{if } \nu_{C_i}(\mathbf{z}_i) > \lambda_i \\ 0 & \text{otherwise} \end{cases} \quad (4.30)$$

where  $\lambda_i$  is the unique solution in  $\eta$  to:  $Q_{n/2}(\tau, \eta) = \gamma$ . The following corollary is a particular case of Proposition (4.4.3) for  $p = 1$ . Hence, the inequalities about probabilities of false alarm and detection  $P_{FA}(\mathcal{T}_i)$  and  $P_{det}(\mathcal{T}_i)$  of  $\mathcal{T}_i$  are obtained straightforwardly from equations (4.21) and (4.22).

**Corollary 4.4.1.** *With the same notation as above, we have:*

$$P_{FA}(\mathcal{T}_i) \leq \gamma \quad (4.31)$$

$$P_{det}(\mathcal{T}_i) \geq Q_{n/2}(\tau', \lambda_i) \quad (4.32)$$

Note that only lower bounds can be provided in (4.22) and (4.32), due to the poor prior knowledge on  $\Theta$ , whose distribution is unknown.

Finally, proving the equivalence between our initial detection problem (4.8) and the RDT problem (4.11) implies the equality between the detection probability of our initial problem and the power function of the RDT problem. As regards the false alarm probability of our initial problem, it appeared that it stayed below the size of the RDT problem. These results are quite important because optimality reached for the RDT problem are transferred to our initial problem that is supposedly not solvable with guarantee of optimality according to standard approaches.

### 4.4.3 Optimality for the distributed configuration

As described in subsection (4.3.2), the goal of the distributed configuration is to reduce information transmitted from the sensors to the fusion center, while maintaining performance loss to a minimum. We still assume true  $\mathcal{P}_0^{FC}$  and  $\mathcal{P}_1^{FC}$  and address the same detection problem (4.8) and equivalently (4.11). In this respect, we show that sending the local RDT statistic (which is a scalar) to the fusion center not only reduces the communication cost but also guarantees the same performance as the centralised approach (see Proposition (4.4.3)). More specifically, we show that sending the raw observations to a fusion center that derives the overall test statistic is equivalent to sending the local test statistics computed individually by each sensor to the fusion center where only a summation is calculated. This is all formalised by the next proposition.

**Proposition 4.4.4.** *If  $\mathcal{T}_{FCD} : [0, \infty)^p \rightarrow \{0, 1\}$  is defined for any  $p$ -tuple  $(x_1, \dots, x_p)$  of non-negative reals by:*

$$\mathcal{T}_{FCD}(x_1, \dots, x_p) = \begin{cases} 1 & \text{if } \sum_{i=1}^p x_i > \lambda_{FC} \\ 0 & \text{otherwise} \end{cases} \quad (4.33)$$

*then, with the same notation as above:*

$$\mathcal{T}_{FC}(\mathbf{Z}) = \mathcal{T}_{FCD}(\nu_{C_1}(\mathbf{Z}_1), \dots, \nu_{C_p}(\mathbf{Z}_p)) \quad (4.34)$$

*Proof.* Proposition (4.4.4) is an immediate consequence of the stretchability of the RDT's statistic of test.

we remind that:  $\Sigma = \begin{pmatrix} \Sigma_1 & & 0 \\ & \ddots & \\ 0 & & \Sigma_p \end{pmatrix}$  and

$$\mathbf{C} = \begin{pmatrix} \mathbf{A}_1 \boldsymbol{\Sigma}_1 \mathbf{A}_1^T & & 0 \\ & \ddots & \\ 0 & & \mathbf{A}_p \boldsymbol{\Sigma}_p \mathbf{A}_p^T \end{pmatrix}.$$

Since all the matrices  $\mathbf{A}_i$  are full rank, then the matrices  $\mathbf{C}_i = \mathbf{A}_i \boldsymbol{\Sigma}_i \mathbf{A}_i^T$  are invertible and we have:

$$\mathbf{C}^{-1} = \begin{pmatrix} (\mathbf{A}_1 \boldsymbol{\Sigma}_1 \mathbf{A}_1^T)^{-1} & & 0 \\ & \ddots & \\ 0 & & (\mathbf{A}_p \boldsymbol{\Sigma}_p \mathbf{A}_p^T)^{-1} \end{pmatrix}.$$

And finally, since the covariance matrix  $\mathbf{C}^{-1}$  is a block diagonal matrix:

$$\nu_{\mathbf{C}}(\mathbf{y}) = \sum_{i=1}^{i=p} \nu_{\mathbf{C}_i}(\mathbf{Y}_i) \quad (4.35)$$

which implies (4.34) and concludes the proof.  $\square$

Therefore, instead of sending the whole observation,  $\mathbf{Z}$ , to the fusion center, sending only the Mahalanobis norm,  $\nu_{\mathbf{C}_i}(\mathbf{Z}_i)$ , computed at each sensor, enables the fusion center to compute exactly the same test statistic:  $\nu_{\mathbf{C}}(\mathbf{Z})$ . This implies that the communication overhead is reduced by an order equivalent to the dimensions of the observations, in our case by a factor of  $n$ . This is particularly useful in high dimensional settings with limited communication bandwidth as it is not feasible to forward the complete high dimensional signal to the FC in such settings. Moreover, Proposition 4.4.4 implies equivalence between the two configurations in terms of detection performance also. Thus, the distributed configuration guarantees detection performance along with a lighter data transmission. It is worth emphasising that these results merely follow from the properties of the Mahalanobis norm.

## 4.5 Numerical results

### 4.5.1 Lower bounds on detection probabilities

In this subsection, we compare the lower bounds on detection probabilities derived in equation (4.22) for the centralised and the distributed configurations, i.e.  $p > 1$  sensors, with lower bounds on detection probabilities derived in equation (4.32) for the one-sensor based configuration. Note that the performance bounds of what we called the centralised and the distributed configuration are the same according to Proposition 4.4.4. Let us recall also that the criteria of optimality we used from the the RDT framework is the maximal constant conditional power of a test inside the class of coherent tests of level  $\gamma$ :  $\mathcal{C}_\gamma$ . We will vary some parameters that may impact the performance lower-bound and help bring out the effect of multi-sensors networks in detection. Specifically, we plot the detection performance as a function of the level  $\gamma$  and the SNR  $\tau'$ . Hence, the simulations highlight the gains achieved by deploying multiple sensors compared to a single sensor system. For both figures, we set the tolerance to  $\tau = 0$  dB and  $n = 64$ . Fig. 4.1 plots the lower-bounds on the detection probabilities of the tests  $\mathcal{T}_{FC}$  (for  $p = 2, p = 4, p = 8$ ) and  $\mathcal{T}_i$  (for  $p = 1$ ) as a function of  $\gamma$ . The parameter  $\tau'$  is set to 6 dB. Increasing the number of sensors,  $p$ ,

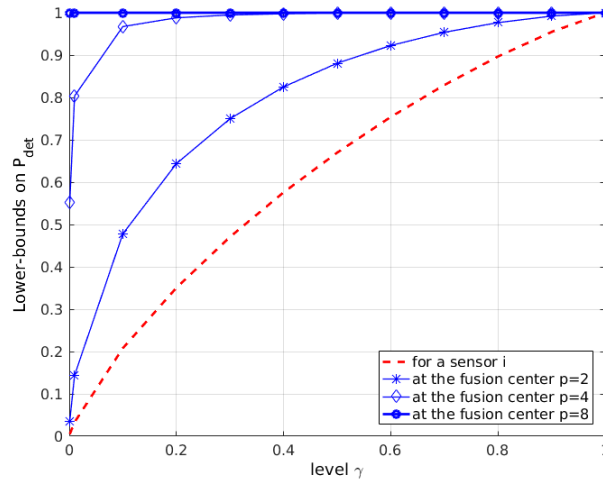


Figure 4.1: Lower bounds on the detection probabilities as a function of  $\gamma$ .  $n = 64$ ,  $\tau = 0$  dB,  $\tau' = 6$  dB.

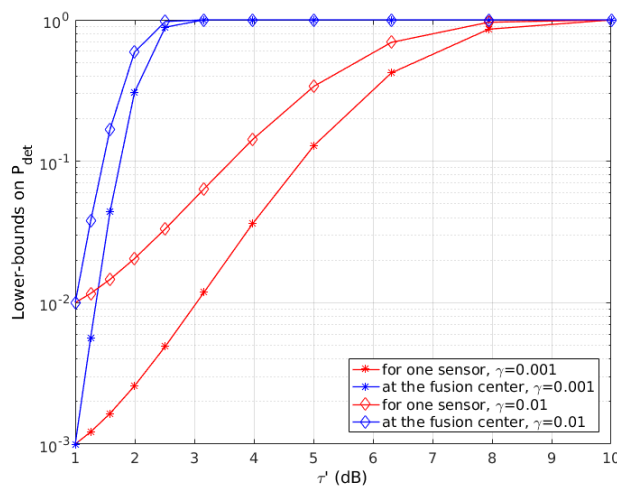


Figure 4.2: Lower bounds on the detection probabilities as a function of  $\tau'$ .  $p = 5$ ,  $n = 64$ ,  $\tau = 0$  dB.

improves the detection performance. For instance, for a level  $\gamma = 10^{-1}$ , as the number of sensors increases from 2 to 4, the lower bound on the detection probability doubles. Therefore, the detection performance is expected to increase when multiple sensors are deployed. The impact of  $\tau'$  is illustrated in Figure 4.2. As expected, the more it departs from the value of the tolerance  $\tau$ , the more the detection performance increases for both values of  $\gamma$ . This behaviour conveys a certain conservatism in the detection strategy and shows that depending on the application, the parameter  $\tau'$  can be chosen to trigger more or less false alarms depending on the nature of the event to be detected and the degree of caution targeted. Again, the detection performance improves with the number of sensors. Indeed, it is obvious in Figure 4.2 that the rise of the two blue curves ( $\mathcal{T}_{FC}$  and  $\mathcal{T}_{FCD}$ ) is steeper and faster than the two red curves ( $\mathcal{T}_i$ ).

### 4.5.2 The constraint of level for the false alarm probability

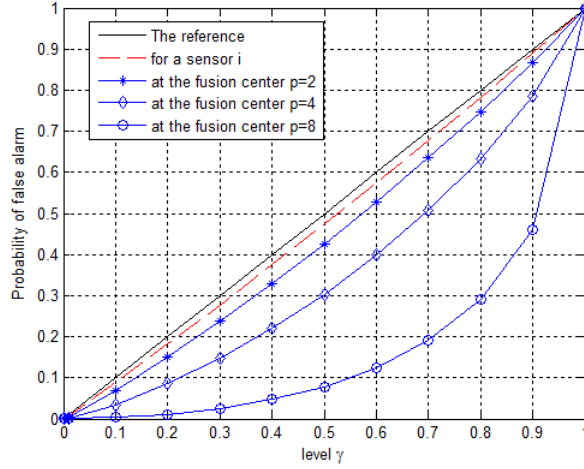


Figure 4.3: Upper bounds on the false alarm probabilities as a function of  $\gamma$ .  $n = 64$ ,  $\tau = 0$  dB.

In this subsection we consider that  $\epsilon = 0$  i.e.  $\Theta = \xi$ , in other words that the null hypothesis;  $h_0^{FC}$  and equivalently  $\mathcal{H}_0^{FC}$ , is true. Therefore we assume that  $\frac{1}{p}\nu_{\mathcal{C}}(\mathbf{A}\Theta) = \frac{1}{p}\nu_{\mathcal{C}}(\mathbf{A}\xi) = \rho_0 \leq \tau$  is verified for each realisation. The parameter  $\rho_0$  can be seen as an average of the energy of all local interferences after transformation collected by the sensors:  $\rho_0 = \frac{1}{p}\sum_{i=1}^{i=p}\rho_{0i}$  with  $\rho_{0i} = \nu_{\mathcal{C}_i}(\mathbf{A}_i\xi_i)$ . We fix  $\rho_0 = \rho_{0i} = -3$ dB. The tolerance is kept to  $\tau = 0$  dB and  $n = 64$ .

Figure 4.3 supports the theoretical results from equations 4.21 and 4.31 about the respect of constraint of conditional probability of false alarm when calculating the thresholds according to the RDT formalism, be it for the test  $\mathcal{T}_{FC}$  computed at the fusion center:  $\alpha_{\mathcal{T}_{FC}} < \gamma$  or the test computed at a local scale  $\mathcal{T}_i$  for each sensor:  $\alpha_{\mathcal{T}_i} < \gamma$ . Indeed, no matter the value of  $p$ , the false alarm probability stays below the level  $\gamma$  which indicates that the optimality criterion is guaranteed. Nevertheless,  $\alpha_{\mathcal{T}_{FC}}$  can be even more reduced by raising the number of sensors of the network, naturally.

## 4.6 Conclusion

Across this chapter, we wanted to widen the scope of the RDT framework and extend it to distributed configuration to test its limits. It appeared that the RDT framework naturally fits the distributed context and gives the opportunity of lessening communication costs between sensors and the fusion center without any performance loss. Moreover, using the RDT framework opens up the doors of classically unsolvable detection problems with the standard approach like the Neyman-Pearson approach, the Bayesian approach, etc.. Thereupon, even though optimality for our initial detection problem 4.8 was not conceivable at first because of certain assumptions about the signal, but a reformulation under the RDT framework allowed us to obtain performance bounds somehow analogous to standard optimality criterion. Besides solving a problem that is, according to standard approaches, unsolvable while guaranteeing a certain

optimality, we wanted to exploit this problem in order to get more familiar with the RDT framework. Indeed, in the next chapter we will propose a generalisation of the RDT framework, hence the current chapter can be considered as a familiarisation step with the RDT framework that led to a contribution.





# Chapter 5

## General Random distortion Testing

In this chapter, a generalisation of the Random Distortion Testing (RDT) framework [Pastor and Nguyen, 2013a] (chapter 3) is presented. The RDT framework is surely already a general framework that can fit a wide range of detection scenarios due to the assumption about the unknown signal's distribution. Nonetheless, the General Random Distortion Testing (GRDT) problem we consider in this chapter contains even less constraints, namely about the noise that is assumed of any distribution, and not necessarily Gaussian. Moreover, the Mahalanobis norm is replaced by any maximal invariant  $\mathcal{M}(\cdot)$  associated to the group of transformation  $\mathcal{G}$  that leaves the problem invariant. Consequently, the expression of the null and alternative hypotheses, more rigorously the null and alternative events, are affected and become more general. Optimality criteria from the RDT framework are redefined to fit the GRDT problem, then theoretical results are provided about  $\gamma$ -MCCP (Maximal Constant Conditional Power) tests for the corresponding newly defined decision problem. The main theorem displays the conditions for a test to be  $\gamma$ -MCCP within the framework of the GRDT problem.

### 5.1 Problem statement

Let  $\mathbf{Y}$ ,  $\Theta$ ,  $\mathbf{X}$  be three  $N$ -dimensional real random vectors. We assume that the signal  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  is random of unknown distribution. The noise  $\mathbf{X} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  admits a probability density function  $f_{\mathbf{X}}$ , and the observation  $\mathbf{Y} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  is such that  $\mathbf{Y} = \Theta + \mathbf{X}$ . The rigorous expression would be  $\mathbf{Y}(\omega) = \Theta(\omega) + \mathbf{X}(\omega)$  for  $\omega \in \Omega$  but for simplification purposes  $\omega$  is omitted. We denote  $\mathcal{M}(\cdot)$ , a maximal invariant (see chapter 2, definition 2.2.1) related to a group of transformation  $\mathcal{G}$  that leaves the problem invariant such that  $\theta \in \mathbb{R}^N \mapsto \mathcal{M}(\theta) \in \mathbb{R}$ . A parameter  $\tau \geq 0$  is called tolerance, analogously to the RDT framework. the GRDT problem consists of deciding whether the event  $[\mathcal{M}(\Theta) \leq \tau]$  or  $[\mathcal{M}(\Theta) > \tau]$  is verified. The null event corresponds to  $[\mathcal{M}(\Theta) \leq \tau]$ , and the alternative event is represented by  $[\mathcal{M}(\Theta) > \tau]$ , and we want to decide which event is true. The GRDT problem can be presented as follows:

$$\text{GRDT: } \begin{cases} \text{Observation: } \mathbf{Y} = \Theta + \mathbf{X} & \begin{cases} \Theta \text{ and } \mathbf{X} \text{ independent,} \\ \Theta \in \mathcal{M}(\Omega, \mathbb{R}^N), \\ \mathbf{X} \sim f_{\mathbf{X}}, \end{cases} \\ \text{Null event: } [\mathcal{M}(\Theta) \leq \tau], \\ \text{Alternative event: } [\mathcal{M}(\Theta) > \tau]. \end{cases} \quad (5.1)$$

The shape of the problem, and more specifically the presence of a maximal invariant statistic, naturally implies, as already said previously, that the GRDT is invariant with respect to some group of transformations  $\mathcal{G}$ . It means that deciding on the basis of  $g(\Theta + \mathbf{X})$  or  $\Theta + \mathbf{X}$  for any  $g \in \mathcal{G}$ , has to lead to the same result. Therefore, these invariances can be exploited while designing tests and decision rules. Notice that the Mahalanobis norm used in the definition of the events/hypotheses in the RDT framework is a particular case of the maximal invariant in 5.1. By considering a maximal invariant, the notion of orbits can be introduced inherently. The orbits of  $\mathcal{G}$  can be defined as  $\Upsilon_\rho = \{\mathbf{y} \in \mathbb{R}^N : \mathcal{M}(\mathbf{y}) = \rho\}$  with  $\rho \in \mathbb{R}$ . The family of all the orbits  $\Upsilon_\rho$  with  $\rho \in \mathbb{R}$  are denoted  $\mathfrak{F}$ .

In the same vein as the RDT framework, we want to use the invariances of the GRDT problem in order to define an appropriate optimal test. However, due to the unknown PDF of the signal of interest  $\Theta$ , these invariances cannot be exploited in the standard way. Indeed, the very notion of invariance changes for the RDT and GRDT problems due to the randomness of the signal. The invariant family of distributions, according to the classical principle of invariance, considers a family  $\{\mathbb{P}_\theta\}$  parametrized by a real scalar  $\theta$ , unlike our new concept of invariance where the family of distribution we examine is parametrized by the whole signal  $\Theta$  such that the family  $\{\mathbb{P}_\Theta\}$  can be seen as the set of all probability density functions (see chapter 2).

Concerning the optimality criterion we intend to use to characterise optimal tests suitable for the GRDT problem, it will be analogous to the  $\gamma$ -MCCP criterion, adapted to the presence of a maximal invariant  $\mathcal{M}(\cdot)$ .

## 5.2 Preliminary definitions

### 5.2.1 Power function and size

Let us consider  $\mathcal{T} : \mathbb{R}^N \rightarrow \{0, 1\}$  a non-randomised statistical test. Given any  $\boldsymbol{\theta} \in \mathbb{R}^N$ , the power function of a test  $\mathcal{T}$  is defined as:

$$\beta_{\mathcal{T}}(\boldsymbol{\theta}) = \mathbb{P}[\mathcal{T}(\boldsymbol{\theta} + \mathbf{X}) = 1]. \quad (5.2)$$

Hence the size of a test  $\mathcal{T}$  is defined as:

$$\alpha_{\mathcal{T}} = \sup_{\boldsymbol{\theta} \in \mathbb{R}^N : \mathcal{M}(\boldsymbol{\theta}) \leq \tau} \beta_{\mathcal{T}}(\boldsymbol{\theta}). \quad (5.3)$$

It is useful for the sequel to define the constant power function.

**Definition 5.2.1.** A test  $\mathcal{T}$  is said to have constant power function on a given orbit  $\Upsilon_\rho \in \mathfrak{F}$  for  $\rho \in \mathbb{R}$  if for any  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Upsilon_\rho$ ,  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) = \beta_{\mathcal{T}}(\boldsymbol{\theta}')$ .

### 5.2.2 Conditional power function

Let  $\rho \in \mathbb{R}$ , and  $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  be independent of  $\mathbf{X}$ . For any  $\Upsilon_\rho \in \mathfrak{F}$ , the conditional power function of a test  $\mathcal{T}$  can be defined as:  $\mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \mathcal{M}(\boldsymbol{\Theta}) = \rho]$ .

On the basis of the notion of the conditional power function of a test, we can define a GRDT size of a test  $\mathcal{T}$  as follows.

**Definition 5.2.2.** The GRDT size of a test  $\mathcal{T}$  for the GRDT problem is defined by

$$\alpha_{\mathcal{T}}^{[GRDT]} = \sup_{\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N): \mathbb{P}[\mathcal{M}(\boldsymbol{\Theta}) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \mathcal{M}(\boldsymbol{\Theta}) \leq \tau] \quad (5.4)$$

The definition of a GRDT size is more appropriate for our problem 5.1. It can be proved that the GRDT size is equal to the standard size of a test through Lemma 5.3.1 in the next section.

Aside from the GRDT size, we also need to define the constant conditional power function that is useful for the theoretical results.

**Definition 5.2.3.** Let  $\rho \in \mathbb{R}$ . Let  $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  be independent of  $\mathbf{X}$ . A test  $\mathcal{T}$  is said to have constant conditional power function given  $\boldsymbol{\Theta} \in \Upsilon_\rho$  if:

$$\mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \mathcal{M}(\boldsymbol{\Theta}) = \rho] = \beta_{\mathcal{T}}(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \Upsilon_\rho. \quad (5.5)$$

## 5.3 Theoretical results

In this section, first, some preliminary material will be presented about the size of a test for the GRDT problem and about the conditional power function of a test for the specific assumptions of the problem. Then, a sufficient condition for a test to be  $\gamma$ -MCCP for the GRDT problem will be stated. This is the most significant result of the chapter and is established in Theorem 5.3.4.

The initial purpose we had when we wanted to generalise the RDT framework, was to find the exact expression of the  $\gamma$ -MCCP test for the GRDT problem. The methodology employed is to generalise one by one each and every key result of the RDT framework along with the definitions in order to conform to the specificities of the GRDT problem. In doing so, we end up forging the same reasoning of the RDT framework while taking into consideration the different model and assumptions of the GRDT problem. This task turned out to be tricky and hard to achieve, as we will explain more in details later on.

### 5.3.1 Preliminary results

The following lemma makes the link between the standard notion of size  $\alpha_{\mathcal{T}}$ , defined in Eq. (5.3), and the GRDT size of a test  $\alpha_{\mathcal{T}}^{[GRDT]}$  defined in Eq. (5.4).

**Lemma 5.3.1.** *For any test  $\mathcal{T}: \mathbb{R}^N \rightarrow \{0, 1\}$ , we have:*

$$\alpha_{\mathcal{T}}^{[GRDT]} = \alpha_{\mathcal{T}} \quad (5.6)$$

*Proof.* See Appendix G. □

Lemma 5.3.1 allows to establish an equivalence between the classical size of a test and the GRDT size of a test (Eq. (5.4)). This result is the first needed for the demonstration of the main result of this chapter. The next two lemmas are also key elements in the demonstration of that result (see Appendix H). They provide some implications between the constant power function and the constant conditional power function.

**Lemma 5.3.2.** *A test  $\mathcal{T}$  has a constant power function on  $\Upsilon_{\rho}$  with  $\rho \in (0, \infty)$  if and only if:  $\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1] = \beta_{\mathcal{T}}(\theta)$  for any  $\theta \in \Upsilon_{\rho}$  and any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  independent of  $\mathbf{X} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  and such that  $\Theta \in \Upsilon_{\rho}$  (a.s).*

*Proof.* This lemma is a direct consequence of the fact that given any test  $\mathcal{T}$  and any  $\rho \in (0, \infty)$ , we have:

$$\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1] = \int_{\Upsilon_{\rho}} \beta_{\mathcal{T}}(\theta) \mathbb{P}_{\Theta}(d\theta) \quad (5.7)$$

for all  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$ , independent of  $\mathbf{X}$  and such that  $\Theta \in \Upsilon_{\rho}$  (a.s). □

**Lemma 5.3.3.** *Let  $\rho \in \mathbb{R}$  and  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$ . If a test  $\mathcal{T}$  has a constant conditional power function given  $\Theta \in \Upsilon_{\rho}$ , then  $\mathcal{T}$  has a constant power function on  $\Upsilon_{\rho}$ .*

*Proof.* It can be demonstrated straightforwardly from the definition of the constant conditional power function (definition 5.2.3). □

### 5.3.2 An optimal test

When we have a decision problem in hand, targeting optimal tests is the first step in the resolution process. Unfortunately, because of the lack of information about the signal's distribution, standard optimality criterion like the MPE, MP and UMP tests, etc. are not conceivable. Even the holy trinity tests<sup>1</sup> are not practicable for our GRDT problem 5.1 since the distribution of the signal of interest is needed to derive them. On the other hand, assuming that  $\mathcal{M}(\Theta) \leq \tau$  or  $\mathcal{M}(\Theta) > \tau$  is verified

---

<sup>1</sup>Generalized Likelihood Ratio Test, Rao Test and Wald Test.

depending on the event that is true, can seem at first sight intangible, but by using the RDT framework as a reference, it turns out to be very concrete. Indeed,  $\mathcal{M}(\Theta)$  can represent for example the SNR,  $\mathcal{M}(\Theta) = \frac{\|\Theta\|^2}{\|\mathbf{X}\|^2}$ , the energy of the signal of interest  $\mathcal{M}(\Theta) = \|\Theta\|^2$ , its projection in some subspace, etc. The use of the tolerance  $\tau$  helps modelling the lack of knowledge by bounding uncertainties. Herewith, we will use the  $\gamma$ -MCCP (Maximal Constant Conditional Power) criterion of optimality introduced by the RDT framework, since it is very suitable in the case of random signals with unknown distribution. However, we will adapt it to fit our more general framework 5.1 given any noise  $\mathbf{X}$  and an associated maximal invariant  $\mathcal{M}(\Theta)$ .

**Definition 5.3.1.** *Given  $\tau \geq 0$  and  $\gamma \in (0, 1)$ , a test  $\mathcal{T}^*$  is said to be  $\gamma$ -MCCP if:*

- **[Level]**  $\mathcal{T}^*$  is of level  $\gamma$ .
- **[MCCP]** *Given any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$  and for  $\mathbb{P}_{\mathcal{M}(\Theta)}$ -almost every  $\rho > \tau$ , the test  $\mathcal{T}^*$  has a constant conditional power function for any  $\Theta \in \Upsilon_\rho$  and for any test  $\mathcal{T}$  of level  $\gamma$  with constant conditional power function on  $\Upsilon_\rho$  we have:*

$$\mathbb{P}[\mathcal{T}(\Theta + \mathbf{X}) = 1 \mid \mathcal{M}(\Theta) = \rho] \leq \mathbb{P}[\mathcal{T}^*(\Theta + \mathbf{X}) = 1 \mid \mathcal{M}(\Theta) = \rho]$$

**Theorem 5.3.4.** *Let  $\mathcal{T}^*$  be a test with level  $\gamma$  for the GRDT problem such that for any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$ ,  $\mathcal{T}^*$  has a constant conditional power function given  $\Theta \in \Upsilon_\rho$  for  $\mathbb{P}_{\mathcal{M}(\Theta)}$ -almost every  $\rho > \tau$ . If, given any  $\rho > \tau$ , there exists two random variables  $\Theta_0$  and  $\Theta_1$ , with  $\Theta_0 \in \Upsilon_\tau$  and  $\Theta_1 \in \Upsilon_\rho$  almost surely, such that the test  $\mathcal{T}^*$  is most powerful with level  $\gamma$  for testing:*

$$\begin{cases} \mathcal{H}_0: & \mathbf{Y} = \Theta_0 + \mathbf{X} \\ \mathcal{H}_1: & \mathbf{Y} = \Theta_1 + \mathbf{X} \end{cases} \quad (5.8)$$

*then the test  $\mathcal{T}^*$  is  $\gamma$ -MCCP for the GRDT problem.*

*Proof.* See Appendix H. □

Theorem 5.3.4 can be considered as one of the main contributions of the manuscript. It is analogous to Lemma D.0.4, that is, the key ingredient to demonstrate the main result of the RDT framework: Theorem 3.2.2. Lemma D.0.4 uses a simple hypothesis testing problem with the help of Neyman-Pearson Theorem to conclude on the optimality of a much more general problem; the RDT problem. Theorem 5.3.4 does the same by considering the existence of a most powerful test for the simple Problem 5.8 to conclude on the existence of a  $\gamma$ -MCCP test for the GRDT problem. However, since the GRDT framework does not consider a specific type of noise  $\mathbf{X}$ , nor a specific maximal invariant statistic, the explicit expression of the optimal test *i.e.* the  $\gamma$ -MCCP test  $\mathcal{T}^*$  can not be presented. In other words the expression of such a test is problem related.

This generalisation has been worked on collaboratively with another PhD student from the SC department of IMT Atlantique Brest: Guillaume Ansel.

## 5.4 Conclusion

This result led us to a very important question: is there other pairs of noise and maximal invariant, aside from the Gaussian noise and the Mahalanobis norm, for which the  $\gamma$ -MCCP test exists. Unfortunately, we still do not have the answer to this question. Nonetheless, it put us on the interesting track of the next chapter where we thought of the asymptotic aspects of optimality when the  $\gamma$ -MCCP criteria of optimality is not possible to pursue. The asymptotic approach we consider is based on the UMP criterion of optimality. As we will see in the sequel, Asymptotically Uniformly Most Powerful tests will be introduced and applied to a problem of deciding whether a deterministic signal is present in a subspace cone.

# Chapter 6

## An asymptotic approach: Asymptotic Karlin-Rubin's Theorem

In this chapter, our purpose was, initially, to generalise the RDT framework, presented in chapter (3), in an asymptotic outlook. However, at the present time, considering an asymptotic standpoint in the generalisation of the RDT framework has become a perspective due to the lack of time. We ended up focusing only on an asymptotic formulation of the Karlin-Rubin theorem for the detection of deterministic signals. Then we apply it to a specific decision problem.

It is worth noting that this whole asymptotic view has emerged while working on the GRDT problem (Chapter 5). Specifically, at a certain point, when we attained the limits of what we could do concerning the generalisation of the RDT due to the general form of the GRDT problem that does not allow a precise specification of the optimal test in the sense of  $\gamma$ -MCCP. That is how we took the turn towards an asymptotic approach.

Thus, in this chapter we will first present an asymptotic formulation of the Neyman-Pearson theorem, that is a key step in the demonstration of the asymptotic formulation of Karlin-Rubin's theorem that comes next. Then a detection problem will be presented in order to show the effectiveness of our approach.

### 6.1 An asymptotic formulation of the Neyman-Pearson theorem

The Neyman-Pearson theorem presented in chapter (1) concerns only detection problems where signals to be detected are deterministic and where the hypotheses are simple. The asymptotic Neyman-Pearson theorem concerns also deterministic signals and simple hypotheses testing problems but in an asymptotic scope that involves lim sups of the power function. The asymptotic Neyman-Pearson theorem can be considered as the stepping stone for our final goal that is presenting an asymptotic formulation of the Karlin-Rubin theorem, since composite hypotheses testing problems are more common.



A few definitions that will prove useful in the sequel will be presented next, before the asymptotic Neyman-Pearson theorem.

**Definition 6.1.1.** Let  $\mathbf{Y}$  be a random vector. A set  $\mathcal{B}$  is called a continuity set if

$$\mathbb{P}[\mathbf{Y} \in \partial\mathcal{B}] = 0 \quad (6.1)$$

where  $\partial\mathcal{B}$  is the border of  $\mathcal{B}$ .

**Definition 6.1.2.** The family of tests  $\Psi_{\mathbf{Y}}$  whose critical regions  $\mathcal{R}^{\mathcal{T}}$  are  $\mathbb{P}$ -continuity sets of the random variable  $\mathbf{Y}$ , are defined as:

$$\Psi_{\mathbf{Y}} \triangleq \{\mathcal{T} : \mathbb{P}[\mathbf{Y} \in \partial\mathcal{R}^{\mathcal{T}}] = 0\} \quad (6.2)$$

**Definition 6.1.3.** Let  $\mathbf{Y}$  be a random vector :  $\Omega \mapsto \mathbb{R}^N$  with a probability density function  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$ , where  $\theta \in \Theta$  and  $\Theta$  the set of parameters of  $f_{\mathbf{Y}}^{\theta}$ . For a test  $\mathcal{T}$ , we note

$$\mathbb{E}_{\theta}[\mathcal{T}(\mathbf{Y})] \triangleq \int_{\mathbb{R}^N} \mathcal{T}(\mathbf{y}) f_{\mathbf{Y}}^{\theta}(\mathbf{y}) d\mathbf{y} = \mathbb{P}[\mathbf{Y} \in \mathcal{R}^{\mathcal{T}}] \quad (6.3)$$

with  $\mathcal{R}^{\mathcal{T}}$  the critical region of the test  $\mathcal{T}$

**Definition 6.1.4.** Let  $(\mathbf{Y}_n)_{n \in \mathbb{N}^*}$  be a sequence of random vectors of probability density function  $f_{\mathbf{Y}_n}^{\theta}(\mathbf{y})$  with  $\theta \in \Theta \subset \mathbb{R}$ . We consider the following detection problem

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1, \quad \theta_0, \theta_1 \in \Theta \end{cases} \quad (6.4)$$

A test  $\mathcal{T}$  is called AMP (Asymptotically Most Powerful) of asymptotic level  $\gamma$  if

$$\limsup_n \mathbb{E}_{\theta_0}[\mathcal{T}(\mathbf{Y}_n)] \leq \gamma \quad (6.5)$$

and if for any other test  $\mathcal{T}'$  satisfying

$$\limsup_n \mathbb{E}_{\theta_0}[\mathcal{T}'(\mathbf{Y}_n)] \leq \gamma$$

We have

$$\limsup_n (\mathbb{E}_{\theta_1}[\mathcal{T}'(\mathbf{Y}_n)] - \mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y}_n)]) \leq 0. \quad (6.6)$$

**Theorem 6.1.1.** Let  $(\mathbf{Y}_n)_{n \in \mathbb{N}^*}$  be a sequence of random vectors, where every  $\mathbf{Y}_n$  has a pdf  $f_{\mathbf{Y}_n}^{\theta}(\mathbf{y})$ ,  $\theta \in \Theta \subset \mathbb{R}$ . Let  $\mathbf{Y}$  be a random vector absolutely continuous of pdf  $f_{\mathbf{Y}}^{\theta}(\mathbf{y})$ ,  $\theta \in \Theta$ , and let  $\mathcal{T} \in \Psi_{\mathbf{Y}}$  such that

$$\mathcal{T}(\mathbf{y}) = \begin{cases} 1 & \text{if } f_{\mathbf{Y}}^{\theta_1}(\mathbf{y}) \geq \eta f_{\mathbf{Y}}^{\theta_0}(\mathbf{y}) \\ 0 & \text{otherwise,} \end{cases} \quad (6.7)$$

where  $\eta \geq 0$  and  $\mathbb{E}_{\theta_0}\{\mathcal{T}(\mathbf{Y})\} = \gamma$ .

If  $\mathbf{Y}_n \xrightarrow{\mathcal{L}} \mathbf{Y}$ , then  $\mathcal{T}$  is Asymptotically Most Powerful (AMP) of asymptotic level  $\gamma$  for testing

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1 \end{cases} \quad (6.8)$$

among all tests belonging to  $\Psi_{\mathbf{Y}}$ .

*Proof.* See Appendix B □

Note that the existence of the test  $\forall \gamma \in [0, 1]$  is a consequence of  $\mathbf{Y}$  being absolutely continuous.

Theorem 6.1.1 is an extension of the classical Neyman-Pearson Theorem 1.2.1 where the observation vector is replaced by a sequence of random vectors. When this sequence of random vectors converges in law towards some random vector whose pdf is known, a test is derived, analogously to Neyman-Pearson's test and is proven to be Asymptotically Most Powerful according to Definition 6.1.4.

**Proposition 6.1.1.** *The test  $\mathcal{T}$  is asymptotically unbiased, i.e.  $\mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y})] \geq \mathbb{E}_{\theta_0}[\mathcal{T}(\mathbf{Y})] = \gamma$ .*

*Proof.* Similarly to the previous proof, we use the definition of  $\mathcal{T}$  to get  $\forall \mathbf{y} \in \mathbb{R}^N$  and  $\forall \gamma' \in [0, 1]$

$$\begin{aligned} (\gamma' - \mathcal{T}(\mathbf{y}))(f_{\mathbf{Y}}^{\theta_1}(\mathbf{y}) - \eta f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})) &\leq 0 \\ \Rightarrow \gamma' - \mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y})] &\leq \eta(\gamma' - \gamma) \end{aligned} \tag{6.9}$$

If we put:  $\gamma' = \gamma$  we can get  $\mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y})] \geq \gamma$ . □

The unbiasedness of a test  $\mathcal{T}$  means that its probability of detection is always higher than its probability of false alarm, which is the least that can be expected from any supposedly optimal test. Indeed, not guaranteeing that the power of a test is not greater than its size put us in the strange position of having to better flip a coin and decide which hypothesis is true based on that. At least in that case we can certify that the detection probability is equal to the probability of false alarm, which are both equal to 1/2. Therefore, the most basic property of an optimal test is to be unbiased, otherwise it will no be reasonable to define or use it. Consequently, a Most Powerful (MP) test and Uniformly Most Powerful (UMP) test is necessarily unbiased. For this reason, we needed to check this property for our Asymptotically Most Powerful criterion of optimality.

We will next define what an Asymptotically Uniformly Most Powerful test is, for a composite hypothesis testing problem, then proof the Asymptotic formulation of the Karlin Rubin theorem. Note that the Asymptotically Most Powerful (AMP) and Asymptotically Uniformly Most Powerful (AUMP) criteria of optimality are very comparable. The only difference concerns the nature of the hypotheses testing problem; for the first criterion involves simple hypotheses testing problems, while the second one covers the composite hypotheses testing problems.

## 6.2 An asymptotic formulation of the Karlin-Rubin theorem

**Definition 6.2.1.** Let  $(\mathbf{Y}_n)_{n \in \mathbb{N}^*}$  be a sequence of random vectors of probability density function  $f_{\mathbf{Y}_n}^\theta(\mathbf{y})$  with  $\theta \in \Theta$ . We consider the following detection problem

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1, \quad \Theta_0 \cup \Theta_1 = \Theta, \quad \Theta_0 \cap \Theta_1 = \emptyset \end{cases} \quad (6.10)$$

A test  $\mathcal{T}$  is called AUMP (Asymptotically Uniformly Most Powerful) of asymptotic level  $\gamma$  if

$$\limsup_n \sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\mathcal{T}(\mathbf{Y}_n)] \leq \gamma \quad (6.11)$$

and if for any other test  $\mathcal{T}'$  satisfying

$$\limsup_n \sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] \leq \gamma$$

We have

$$\limsup_n \sup_{\theta \in \Theta_1} (\mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] - \mathbb{E}_\theta[\mathcal{T}(\mathbf{Y}_n)]) \leq 0. \quad (6.12)$$

**Theorem 6.2.1.** Let  $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^N$  be an absolutely continuous random vector of pdf  $f_{\mathbf{Y}}^\theta(\mathbf{y})$ , where  $\theta \in \Theta \subset \mathbb{R}$ , and let  $\Theta_1$  and  $\Theta_0$  be two subsets of  $\Theta$  such that:  $\Theta = \Theta_0 \cup \Theta_1$  and  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Let:

$$\Lambda_{\theta_0, \theta_1}(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\theta_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})} \mathbb{1}_{[f_{\mathbf{Y}}^{\theta_0} f_{\mathbf{Y}}^{\theta_1} \neq 0]} + \infty \mathbb{1}_{[f_{\mathbf{Y}}^{\theta_0} = 0] \cap [f_{\mathbf{Y}}^{\theta_1} \neq 0]} \quad (6.13)$$

be the likelihood ratio such that:  $\Lambda_{\theta_0, \theta_1} = h_{\theta_0, \theta_1}(V)$ , with  $V : \mathbb{R}^N \mapsto \mathbb{R}$ .

Let  $(\mathbf{Y}_n)_{n \in \mathbb{N}^*}$  be a sequence of random vectors where every  $\mathbf{Y}_n$  has a pdf  $f_{\mathbf{Y}_n}^\theta(\mathbf{y})$ .

If for any  $\theta_0 < \theta_1$ ,  $h_{\theta_0, \theta_1}$  is strictly increasing in  $V$  and  $\mathbf{Y}_n \xrightarrow{\mathcal{L}} \mathbf{Y}$ , then the test

$$\mathcal{T}'(\mathbf{y}) = \begin{cases} 1 & \text{if } V(\mathbf{y}) \geq \lambda \\ 0 & \text{otherwise,} \end{cases} \quad (6.14)$$

is AUMP among all tests belonging to  $\Psi_{\mathbf{Y}}$  for testing

$$\begin{cases} \mathcal{H}'_0 : \theta \leq \theta_0 \\ \mathcal{H}'_1 : \theta > \theta_0 \end{cases} \quad (6.15)$$

*Proof.* First we need to prove that the test  $\mathcal{T}'$  is AUMP for the detection problem (6.7) of theorem (6.1.1). Then, we will demonstrate that the test  $\mathcal{T}'$  is also AUMP when the alternative hypothesis  $\mathcal{H}_1$  is composite. Last when the null hypothesis  $\mathcal{H}_0$  is composite. See Appendix C for a full demonstration.  $\square$

In order to show the practical interest of such an approach, we will consider in the next section a detection problem where the signal of interest is assumed deterministic and present inside a subspace cone. We will see that for our decision problem, standard methods can not be used due the shape of the problem and lack of knowledge about some parameters of the problem.

## 6.3 Testing the presence of a deterministic signal in a subspace cone

### 6.3.1 Problem statement

Let  $\boldsymbol{\theta} \in \mathbb{R}^N$  denote an unknown deterministic signal observed in white Gaussian noise  $\mathbf{X} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$  with unknown variance  $\sigma^2$ . The observation vector is modeled as

$$\mathbf{Y} = \boldsymbol{\theta} + \mathbf{X}. \quad (6.16)$$

The signal-to-noise ratio (SNR), defined as  $\zeta = \frac{\|\boldsymbol{\theta}\|^2}{N\sigma^2}$ , is assumed unknown but bounded with known bounds so that  $\zeta_{\min} \leq \zeta \leq \zeta_{\max}$ . Let  $\mathbf{H}$  denote an  $N \times n$  matrix that spans a rank- $n$  subspace  $\langle \mathbf{H} \rangle$  with  $n \leq N$  and  $\mathbf{P}_H$  its corresponding projection matrix, i.e.,  $\mathbf{P}_H = \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ .

Given a tolerance  $0 \leq \tau \leq 1$ , let  $\mathcal{C}_H$  denote the subspace cone defined as  $\mathcal{C}_H = \left\{ \mathbf{y} \in \mathbb{R}^N : \rho = \frac{\|\mathbf{P}_H \mathbf{y}\|^2}{\|\mathbf{y}\|^2} > \tau \right\}$ . We address the problem of testing whether  $\boldsymbol{\theta}$  belongs to the cone  $\mathcal{C}_H$  or not, when we are given  $\mathbf{y}$ . The hypothesis testing problem is therefore

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{\theta} \notin \mathcal{C}_H \\ \mathcal{H}_1 : \boldsymbol{\theta} \in \mathcal{C}_H. \end{cases} \quad (6.17)$$

Problem (6.17) amounts to testing whether a sufficient proportion of the signal energy lies in the subspace  $\langle \mathbf{H} \rangle$ . Such a problem formulation can be relevant when the signal to be detected obeys the linear subspace model and when it is corrupted by some poorly known interference  $\boldsymbol{\xi}$ , i.e.,  $\boldsymbol{\theta} = \mu \mathbf{H} \boldsymbol{\phi} + \boldsymbol{\xi}$ , where  $\mu$  is a random variable valued in  $\{0, 1\}$  modeling the possible presence or absence of the signal, and where  $\boldsymbol{\phi} \in \mathbb{R}^n$  is the unknown coordinates of this signal in the subspace  $\langle \mathbf{H} \rangle$ . For instance, in some applications, the interfering signals can be so heterogeneous that it cannot be represented by one single parametric model and that the detector cannot learn the interferences' shared features through a training data-set. In such cases,  $\boldsymbol{\xi}$  is unknown, non-orthogonal to the signal of interest, and as opposed to [Socheleau and Samaran, 2017; Socheleau et al., 2015], it does not obey a known subspace model. Given the paucity of knowledge available on the interference,  $\boldsymbol{\xi}$  can only be loosely defined as a signal having "not much in common" with the signal of interest. More formally,  $\boldsymbol{\xi}$  may be defined as a vector whose energy lies mostly outside the subspace in which the signal of interest resides, i.e.,  $\|\mathbf{P}_H \boldsymbol{\xi}\|^2 \leq \tau \|\boldsymbol{\xi}\|^2$ . Assuming that  $\|\mathbf{H} \boldsymbol{\phi} + \mathbf{P}_H \boldsymbol{\xi}\|^2 > \tau \|\mathbf{H} \boldsymbol{\phi} + \boldsymbol{\xi}\|^2$ , (6.17) is then equivalent to the problem of testing whether  $\mu = 0$  or  $\mu = 1$ . A similar model is used and applied to real data in [Socheleau and Samaran, 2017; Socheleau et al., 2015], which attests of its practical relevance. The introduction of such an assumption about the interference is in line with the scope of our previous work in chapter 4. Indeed, the model considers a tolerance  $\tau$  too, but in opposition to the Distributed Random Distortion Testing (DRDT) problem its value cannot exceed 1 which is much easier to determine in practice. Another point of divergence is the supposed known noise variance, whilst in this framework we will end up assuming that the SNR  $\zeta$  is only bounded. Also, note that as opposed to other subspace cone detectors such as [Besson, 2006; Ramprasad et al., 1996], the null hypothesis in (6.17) is not restricted to the observation of Gaussian noise only. Problem (6.17) can also

be relevant in machine learning-like applications where one wants to check on a given data-set whether the assumed linear model matches the analysed data with a sufficient accuracy  $\tau$ .

### 6.3.2 Invariance under group action

The most desirable way to solve the problem would be to find a statistical test that is optimal according to a certain criterion. UMP tests could be the answer we are looking for. Indeed, ensuring that a test has the greatest power among the class of tests with the same level is a strong optimality criterion. However, since  $\boldsymbol{\theta}$  is not known perfectly no such UMP test exists. As Problem (6.17) is invariant according to [Lehmann and Scheffé, 1950, Ch. 3], our attention is restricted to tests invariant to sets of transformations for which the problem is itself invariant. This invariance can be formalised through the group of transforms in  $\mathbb{R}^N$

$$\mathcal{G} = \{g : g(\mathbf{y}) = \kappa(\mathbf{U}_H \mathbf{Q} \mathbf{U}_H^T + \mathbf{U}_{H^\perp} \mathbf{R} \mathbf{U}_{H^\perp}^T) \mathbf{y}\}, \quad (6.18)$$

where  $\kappa \in \mathbb{R}$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  are  $n \times n$  and  $(N-n) \times (N-n)$  orthogonal matrices, respectively.  $\mathbf{U}_H$  is defined such that  $\mathbf{P}_H = \mathbf{U}_H \mathbf{U}_H^T$  and  $\mathbf{U}_{H^\perp}$  such that  $\mathbf{P}_H = \mathbf{I}_N - \mathbf{U}_{H^\perp} \mathbf{U}_{H^\perp}^T$ .

Given the invariance of our problem, it is desirable to find a test  $\mathcal{T}$  such that  $\mathcal{T}(g(\mathbf{y})) = \mathcal{T}(\mathbf{y})$  for all  $\mathbf{y} \in \mathbb{R}^N$  and all  $g \in \mathcal{G}$ . It can easily be checked that

$$\mathbf{y} \in \mathbb{R}^N \mapsto M(\mathbf{y}) = \frac{N-n}{n} \times \frac{\|\mathbf{P}_H \mathbf{y}\|^2}{\|(\mathbf{I}_N - \mathbf{P}_H) \mathbf{y}\|^2} \in [0, \infty) \quad (6.19)$$

is a maximal invariant of  $\mathcal{G}$ , that is  $M(\mathbf{y}) = M(g(\mathbf{y}))$  for all  $g \in \mathcal{G}$  and  $M(\mathbf{y}_1) = M(\mathbf{y}_2) \Rightarrow \mathbf{y}_2 = g(\mathbf{y}_1)$  for some  $g \in \mathcal{G}$ . Therefore, it follows from [Lehmann and Romano, 2005b, Theorem 6.2.1] that any  $\mathcal{G}$ -invariant test is a function of this maximal invariant.

### 6.3.3 Uniformly Most Powerful Invariant test

If the UMPI test exists, its statistic can be obtained by computing the likelihood ratio of the maximal invariant  $M(\mathbf{y})$ . Since  $M(\mathbf{y})$  can be expressed as the ratio of two (scaled) non-central chi-square random variables, it is distributed according to a doubly non-central  $\mathcal{F}$ -distribution [Johnson et al., 1995, Ch. 30] so that the likelihood ratio is expressed as

$$L_M(\mathbf{y}) = \frac{\mathcal{F}(n, N-n, N\zeta\rho_1, N\zeta(1-\rho_1), M(\mathbf{y}))}{\mathcal{F}(n, N-n, N\zeta\rho_0, N\zeta(1-\rho_0), M(\mathbf{y}))} \quad (6.20)$$

where  $\rho_0 \leq \tau$  (resp.  $\rho_1 > \tau$ ) denotes the actual percentage of the energy of  $\boldsymbol{\theta}$  in the subspace  $\langle \mathbf{H} \rangle$  under hypothesis  $\mathcal{H}_0$  (resp.  $\mathcal{H}_1$ ). Since the hypotheses of Problem (6.17) are composite,  $\rho_0$  and  $\rho_1$  are not provided in the problem statement so that the UMPI test cannot be implemented using  $L_M(\mathbf{y})$ . For composite hypothesis testing, the UMPI test can sometimes be derived by invoking Karlin Rubin's theorem that states that the comparison of a scalar maximal invariant statistic to a threshold leads

to a UMPI test if this statistic has monotone nondecreasing likelihood ratio [Scharf, 1991]. Unfortunately,  $L_M(\mathbf{y})$  is not a nondecreasing function of  $M(\mathbf{y})$  for every pair  $(\rho_1 > \rho_0, \rho_0)$  and any set of parameters  $(n, N, \zeta)$ . Therefore, the UMPI test does not exist. In this kind of scenarios, the “suboptimal” Generalised Likelihood Ratio test (GLRT) becomes very often the chosen option after the search for an optimal test has failed. Although the GLRT is always invariant with respect to transformations for which the problem itself is invariant [Levy, 2008, Ch. 5], it does not always guarantee optimality. In contrast, we can choose to derive a test for large  $N$  that is asymptotically UMPI using theorem 6.2.1 exposed in the previous section 6.2.

### 6.3.4 Asymptotically Uniformly Most Powerful Invariant test

Given  $\alpha \in (0, 1)$ , let the test  $\mathcal{T}_A$  be defined, for any  $\mathbf{y} \in \mathbb{R}^N$ , as

$$\mathcal{T}_A(\mathbf{y}) = \begin{cases} 1 & \text{if } M(\mathbf{y}) \geq \lambda_A \\ 0 & \text{otherwise,} \end{cases} \quad (6.21)$$

with  $\lambda_A$  such that

$$F(n, N - n, N\zeta\tau, N\zeta(1 - \tau), \lambda_A) = 1 - \alpha. \quad (6.22)$$

The power function of this test satisfies  $\mathbb{P}[\mathcal{T}_A(\boldsymbol{\theta} + \mathbf{X}) = 1] = 1 - F(n, N - n, N\zeta\rho, N\zeta(1 - \rho), \lambda_A)$  where  $\rho = \|\mathbf{P}_H\boldsymbol{\theta}\|^2/\|\boldsymbol{\theta}\|^2$ .

**Proposition 6.3.1.**  $\mathcal{T}_A$  is AUMP with size  $\alpha$  among  $\mathcal{G}$ -invariant tests for Problem (6.17).

*Proof.*  $M(\mathbf{y}) = (N - n)Z_1/(nZ_2)$ , where  $Z_1 \sim \chi_n^2(N\zeta\rho)$  and  $Z_2 \sim \chi_{(N-n)}^2(N\zeta(1 - \rho))$ . For  $n$  fixed, as  $N \rightarrow \infty$ ,  $Z_2/(N - n) \xrightarrow{\mathbb{P}} 1$ . Applying Slutsky’s theorem, for  $N \rightarrow \infty$  we get  $M(\mathbf{y}) \xrightarrow{\mathcal{L}} Z_1/n$ . Since the non-central chi-squared distribution is known to have a monotone increasing likelihood ratio [Eaton, 1983, pp. 469], Prop. 6.2.1 applies with  $\Theta_0 = [0, \tau]$  and  $\Theta_1 = (\tau, 1]$ . Thence the result.  $\square$

Interestingly, the maximal invariant statistic  $M(\mathbf{y})$  is the same as the one used in other detection problems such as [Scharf and Friedlander, 1994, Ch. 4.12] and [Desai and Mangoubi, 2003, App. A]. However, note that the test itself is different since the power function as well as the optimality properties are problem-dependent.

### 6.3.5 Unknown SNR

Knowledge of the SNR  $\zeta$  is required in (6.22) to compute the detection threshold  $\lambda_A$  so as to satisfy the constraint on the size  $\alpha$ . In practice,  $\zeta$  is rarely known perfectly so that  $\mathcal{T}_A$  cannot be strictly applied. However,  $\mathcal{T}_A$  can provide guidelines to design

relevant ad-hoc tests when no optimal test exists. For instance, a robust test  $\mathcal{T}_B$  can be defined given  $\alpha \in (0, 1)$  and for any  $\mathbf{y} \in \mathbb{R}^N$ , as

$$\mathcal{T}_B(\mathbf{y}) = \begin{cases} 1 & \text{if } M(\mathbf{y}) \geq \lambda_B \\ 0 & \text{otherwise,} \end{cases} \quad (6.23)$$

the threshold  $\lambda_B$  is determined such that:

$$\sup_{\zeta \in \mathcal{L}} F(n, N - n, N\zeta\tau, N\zeta(1 - \tau), \lambda_B) = 1 - \alpha, \quad (6.24)$$

where, (i)  $\mathcal{L} = \mathbb{R}$  when nothing is known about  $\zeta$ , (ii)  $\mathcal{L} = [\zeta_{\min}, \zeta_{\max}]$  when it is easy to specify SNR bounds within which the observation lies with a high probability, or (iii)  $\mathcal{L} = \{\zeta_0\}$  when (asymptotically) optimal properties are required for a specific SNR  $\zeta_0$  and not necessarily for other values. Other approaches, such as marginalization, are possible when prior information are available on the SNR distribution.

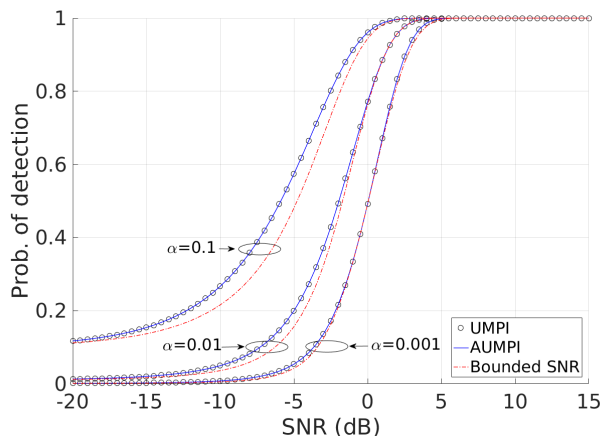


Figure 6.1: Probability of detection versus SNR,  $n = 5$ ,  $N = 50$ ,  $\tau = 0.25$ ,  $\rho_0 = \tau$ ,  $\rho_1 = 0.75$ ,  $\mathcal{L} = [\zeta - 3, \zeta + 3]$  dB.

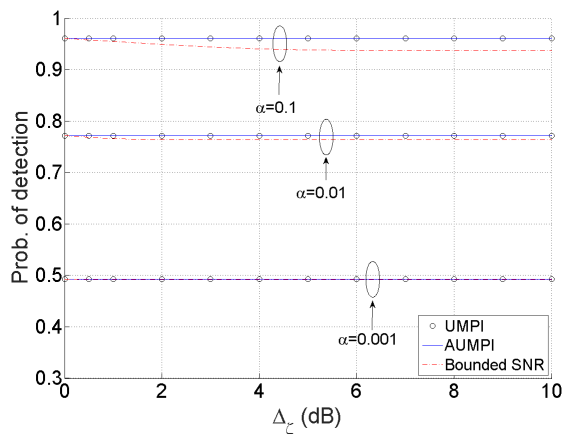


Figure 6.2: Probability of detection versus the SNR gap,  $n = 5$ ,  $N = 50$ ,  $\tau = 0.25$ ,  $\rho_0 = \tau$ ,  $\rho_1 = 0.75$ ,  $\zeta = 0$  dB.

### 6.3.6 Numerical results

Numerical simulations are presented to illustrate the analytic results found previously in subsection (6.3.3), (6.3.4) and (6.3.5). Three tests are compared : the AUMPI test  $\mathcal{T}_A$ , the test  $\mathcal{T}_B$  with bounded SNR where  $\mathcal{L} = [\zeta_{\min}, \zeta_{\max}]$ , and the UMPI test that compares the likelihood ratio (6.20) to a threshold such that the size  $\alpha$  is satisfied. This UMPI test does not exist when  $\rho_0$  and  $\rho_1$  are not provided and is therefore used as a performance bound. For all the simulations,  $n = 5$  and the tolerance is set to  $\tau = 0.25$ . The percentage of the energy of  $\boldsymbol{\theta}$  in the subspace  $\langle \mathbf{H} \rangle$  under hypothesis  $\mathcal{H}_1$  is set to  $\rho_1 = 0.75$ . For the UMPI bound,  $\rho_0$  is set to  $\rho_0 = \tau$ .

Fig. 6.1 shows the probability of detection versus SNR for three different values of level  $\gamma$ , with  $N = 50$ ,  $\mathcal{L} = [\zeta - 3, \zeta + 3]$  dB. It can be seen that the performance gap between the AUMPI test  $\mathcal{T}_A$  and the UMPI bound is insignificant for this example. The loss due to unknown SNR is mostly visible for low SNR signals and for a high value of  $\gamma$ . Such a loss is explained by the conservative behavior of the test  $\mathcal{T}_B$ .

Fig. 6.2 illustrates the performance loss induced by a lack of knowledge on the true SNR value. The probability of detection is shown as a function of the SNR gap  $\Delta_\zeta$ , defined in dB as  $\Delta_\zeta = \zeta - \zeta_{\min} = \zeta_{\max} - \zeta$ . As expected, the performance deteriorates with increasing uncertainty on the SNR. This is mostly detrimental for high sizes  $\gamma$ . It can be noticed that for a gap  $\Delta_\zeta$  greater than 6 dB, there is no additional loss due to additional SNR uncertainty. Such a figure can be relevant for engineers to know how much effort is needed to design an efficient SNR estimator, possibly used before detection.

### 6.3.7 Asymptotic optimality of the GLRT

The Generalised Likelihood Ratio Test is a well-known method normally used in composite hypothesis testing problems when no optimal test, like UMP tests, can be derived. Its advantage is that, by replacing unknown parameters by their maximum likelihood estimators, the (GLRT) acquire a property of invariance towards the inherent invariances of the problem. The limitations of the (GLRT) is that it provides no guarantee of optimality. Nonetheless, the local and asymptotic optimality of (GLRT) has been widely investigated in the literature [Levy, 2008], [Zeitouni et al., 1992], [Kay, 1998], for different scenarios of detection, due to its practical facility. Hence we will explore this state of the art method under the assumption that the SNR  $\zeta$  is unknown but bounded and compare it to our test  $\mathcal{T}_A$  that is considered an upper bound of performance available when the value of the SNR  $\zeta$  is known.

First let us derive the GLRT. We define  $\boldsymbol{\Pi}_0$  and  $\boldsymbol{\Pi}_1$  as

$$\boldsymbol{\Pi}_0 = \left\{ (\boldsymbol{\theta}, \sigma^2) : \boldsymbol{\theta} \notin \mathcal{C}_H, \zeta_{\min} \leq \frac{\|\boldsymbol{\theta}\|^2}{N\sigma^2} \leq \zeta_{\max} \right\} \quad (6.25)$$

$$\boldsymbol{\Pi}_1 = \left\{ (\boldsymbol{\theta}, \sigma^2) : \boldsymbol{\theta} \in \mathcal{C}_H, \zeta_{\min} \leq \frac{\|\boldsymbol{\theta}\|^2}{N\sigma^2} \leq \zeta_{\max} \right\}. \quad (6.26)$$



Problem (6.17) can be rewritten as

$$\begin{cases} \text{Observation : } \mathbf{Y} \sim f_{\mathbf{Y}}^{\boldsymbol{\pi}} \text{ with } \boldsymbol{\pi} \in \boldsymbol{\Pi}_0 \cup \boldsymbol{\Pi}_1 \\ \mathcal{H}_0 : \boldsymbol{\pi} \in \boldsymbol{\Pi}_0 \\ \mathcal{H}_1 : \boldsymbol{\pi} \in \boldsymbol{\Pi}_1. \end{cases} \quad (6.27)$$

The GLR statistic associated is

$$L_G(\mathbf{y}) = \frac{f_{\mathbf{Y}}^{\hat{\boldsymbol{\pi}}_1}(\mathbf{y})}{f_{\mathbf{Y}}^{\hat{\boldsymbol{\pi}}_0}(\mathbf{y})} \quad \text{with} \quad \hat{\boldsymbol{\pi}}_i = \left( \hat{\boldsymbol{\theta}}_i, \hat{\sigma}_i^2 \right) = \underset{\boldsymbol{\pi}_i \in \boldsymbol{\Pi}_i}{\operatorname{argsup}} f_{\mathbf{Y}}^{\boldsymbol{\pi}_i}(\mathbf{y}) \quad (6.28)$$

are the maximum likelihood estimates (MLE) of the signal and the noise variance. The argsup is duly used since the set  $\boldsymbol{\Pi}_i$  is semi open. Injecting (6.16) in (6.28), we get

$$\log L_G(\mathbf{y}) = \frac{N}{2} \log \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \right) + \frac{1}{2} \left( \frac{\|\mathbf{y} - \hat{\boldsymbol{\theta}}_0\|^2}{\hat{\sigma}_0^2} - \frac{\|\mathbf{y} - \hat{\boldsymbol{\theta}}_1\|^2}{\hat{\sigma}_1^2} \right). \quad (6.29)$$

To derive the GLRT, the MLE  $\left( \hat{\boldsymbol{\theta}}_i, \hat{\sigma}_i^2 \right)$ ,  $i = 0, 1$ , have to be found. Two cases must be considered:  $\mathbf{y} \in \mathcal{C}_H$  and  $\mathbf{y} \notin \mathcal{C}_H$ .

First, when the observation  $\mathbf{y}$  belongs to the cone, it can be checked that  $\mathbf{y}$  is the MLE of  $\boldsymbol{\theta}$  under  $\mathcal{H}_1$  and that  $\hat{\sigma}_1^2$  takes the smallest possible value satisfying the constraints defined in (6.26). Thus, when  $\mathbf{y} \in \mathcal{C}_H$ ,

$$\hat{\boldsymbol{\pi}}_1 = \left( \mathbf{y}, \|\mathbf{y}\|^2 / (N\zeta_{\max}) \right). \quad (6.30)$$

Finding  $\hat{\boldsymbol{\pi}}_0$  when  $\mathbf{y} \in \mathcal{C}_H$  is more involved as it cannot be expressed in closed-form. The problem to solve is the following

$$\begin{cases} \max_{\boldsymbol{\pi}_0 = (\boldsymbol{\theta}_0, \sigma_0^2)} -N \log(2\pi\sigma_0^2) - \frac{\|\mathbf{y} - \boldsymbol{\theta}_0\|^2}{\sigma_0^2} \\ \text{subject to} \begin{cases} \|\mathbf{P}_H \boldsymbol{\theta}_0\|^2 - \tau \|\boldsymbol{\theta}_0\|^2 \leq 0, \\ \|\boldsymbol{\theta}_0\|^2 - \zeta_{\max} N \sigma_0^2 \leq 0, \\ \zeta_{\min} N \sigma_0^2 - \|\boldsymbol{\theta}_0\|^2 \leq 0. \end{cases} \end{cases} \quad (6.31)$$

Using Lagrange multipliers and the Karush-Kuhn-Tucker conditions [Nocedal and Wright, 1999, Ch. 12], it can be found that the MLE  $\hat{\boldsymbol{\theta}}_0$  always lies on the boundary of the cone and that its norm as well as the value of  $\hat{\sigma}_0^2$  depend on the SNR constraints. More formally, let  $\mathbf{y}_{\mathcal{C}_H}$  be defined as the orthogonal projection of  $\mathbf{y}$  onto the cone  $\mathcal{C}_H$  so that [Besson, 2006, Sec. II]

$$\mathbf{y}_{\mathcal{C}_H} = (\mathbf{I}_N + \mu_0 (\mathbf{P}_H - \tau \mathbf{I}_N))^{-1} \mathbf{y}, \quad (6.32)$$

where  $\mu_0 = \operatorname{argmin}_{\mu} \mathbf{y}^T (\mathbf{I}_N + \mu (\mathbf{P}_H - \tau \mathbf{I}_N))^{-1} \mathbf{y}$ . When  $\mathbf{y} \in \mathcal{C}_H$ ,  $\hat{\boldsymbol{\pi}}_0$  is expressed as

**Case 1:** if  $\zeta_{\min} \leq \frac{\|\mathbf{y}_{\mathcal{C}_H}\|^2}{\|\mathbf{y} - \mathbf{y}_{\mathcal{C}_H}\|^2} \leq \zeta_{\max}$  then

$$\hat{\boldsymbol{\pi}}_0 = \left( \mathbf{y}_{\mathcal{C}_H}, \|\mathbf{y} - \mathbf{y}_{\mathcal{C}_H}\|^2 / N \right). \quad (6.33)$$

**Case 2:** if  $\|\mathbf{y}_{\mathcal{C}_H}\|^2 \leq \zeta_{\min}\|\mathbf{y} - \mathbf{y}_{\mathcal{C}_H}\|^2$  then

$$\hat{\boldsymbol{\pi}}_0 = (a_{\min}\mathbf{y}_{\mathcal{C}_H}, \|a_{\min}\mathbf{y}_{\mathcal{C}_H}\|^2/(N\zeta_{\min})), \quad (6.34)$$

with  $a_{\min} = \zeta_{\min} (\sqrt{1 + 4\|\mathbf{y}\|^2/(\zeta_{\min}\|\mathbf{y}_{\mathcal{C}_H}\|^2)} - 1) / 2$ .

**Case 3:** if  $\|\mathbf{y}_{\mathcal{C}_H}\|^2 \geq \zeta_{\max}\|\mathbf{y} - \mathbf{y}_{\mathcal{C}_H}\|^2$  then

$$\hat{\boldsymbol{\pi}}_0 = (a_{\max}\mathbf{y}_{\mathcal{C}_H}, \|a_{\max}\mathbf{y}_{\mathcal{C}_H}\|^2/(N\zeta_{\max})), \quad (6.35)$$

with  $a_{\max} = \zeta_{\max} (\sqrt{1 + 4\|\mathbf{y}\|^2/(\zeta_{\max}\|\mathbf{y}_{\mathcal{C}_H}\|^2)} - 1) / 2$ .

When the SNR constraints are not saturated (case 1),  $\hat{\boldsymbol{\theta}}_0$  is simply the orthogonal projection of the observation onto the cone and  $\hat{\sigma}_0^2$  is the mean square error between  $\hat{\boldsymbol{\theta}}_0$  and the observation. When the SNR constraints are saturated (case 2 and 3),  $\hat{\boldsymbol{\theta}}_0$  still lies on the boundary of the cone but it is scaled by a factor of  $a_{\min/\max}$  so as to meet the SNR constraints.

When the observation is outside the cone,  $\mathbf{y} \notin \mathcal{C}_H$ , the same approach can be applied by inverting the role of  $\hat{\boldsymbol{\pi}}_0$  and  $\hat{\boldsymbol{\pi}}_1$ . In that case,  $\hat{\boldsymbol{\pi}}_0 = (\mathbf{y}, \|\mathbf{y}\|^2/(N\zeta_{\max}))$  and  $\hat{\boldsymbol{\pi}}_1$  depends on the orthogonal projection onto the cone, which is now given by

$$\mathbf{y}_{\mathcal{C}_H} = (\mathbf{I}_N + \mu_1(\tau\mathbf{I}_N - \mathbf{P}_H))^{-1}\mathbf{y}, \quad (6.36)$$

where  $\mu_1 = \operatorname{argmin}_{\mu} \mathbf{y}^T (\mathbf{I}_N + \mu(\tau\mathbf{I}_N - \mathbf{P}_H))^{-1}\mathbf{y}$ .

After injecting (6.30)-(6.36) in (6.29), the GLR statistic can be greatly simplified by noticing that  $\log L_G(\mathbf{y})$  is monotonically increasing with increasing ratio  $\|\mathbf{y} - \mathbf{y}_{\mathcal{C}_H}\|^2/\|\mathbf{y}\|^2$  when  $\mathbf{y} \in \mathcal{C}_H$ , and that it is monotonically decreasing with this ratio when  $\mathbf{y} \notin \mathcal{C}_H$ . According to (6.32) and (6.36),  $\mathbf{y}$  is a linear combination of  $\mathbf{y}_{\mathcal{C}_H}$  and  $\mathbf{P}_H\mathbf{y}$ . A simple geometrical analysis based on the angles involved in the expressions  $\log L_G(\mathbf{y})$  and  $M(\mathbf{y})$  shows that  $\log L_G(\mathbf{y})$  is monotonically increasing with increasing maximal invariant  $M(\mathbf{y})$ . Therefore, the GLRT at level  $\gamma$  is the test  $\mathcal{T}_G$  defined, for any  $\mathbf{y} \in \mathbb{R}^N$ , as

$$\mathcal{T}_G(\mathbf{y}) = \begin{cases} 1 & \text{if } M(\mathbf{y}) \geq \lambda_G \\ 0 & \text{otherwise,} \end{cases} \quad (6.37)$$

with

$$\sup_{\boldsymbol{\pi} \in \Pi_0} \beta_{\mathcal{T}_G}(\boldsymbol{\theta}) \leq \gamma, \quad (6.38)$$

where  $\beta_{\mathcal{T}_G}(\boldsymbol{\theta}) = \mathbb{P}[\mathcal{T}_G(\boldsymbol{\theta} + \mathbf{X}) = 1]$  is the power function of  $\mathcal{T}_G$ . This power function satisfies  $\beta_{\mathcal{T}_G}(\boldsymbol{\theta}) = 1 - F(n, N - n, N\zeta\rho, N\zeta(1 - \rho), \lambda_G)$ , where  $F$  is the cumulative distribution function of a doubly non-central  $\mathcal{F}$ -distribution [Johnson et al., 1995, Ch. 30] and  $\rho = \|\mathbf{P}_H\boldsymbol{\theta}\|^2/\|\boldsymbol{\theta}\|^2$ .

It can be noticed that when SNR is known the (GLRT) is equivalent to the test  $\mathcal{T}_A$  so that the power of  $\mathcal{T}_A$  is always greater than or equal to the power of  $\mathcal{T}_G$ . Therefore, according to Proposition 6.3.1 the test  $\mathcal{T}_G$  is AUMPI if there is no uncertainty on the SNR, i.e., if  $\zeta_{\min} = \zeta_{\max}$ . When  $\zeta_{\min} < \zeta_{\max}$ , the test  $\mathcal{T}_G$  is equivalent to the test  $\mathcal{T}_B$ . Hence,  $\mathcal{T}_A$  can be used as an upper performance bound also for the test our GLRT test  $\mathcal{T}_G$  and can help quantifying the performance loss of the GLRT due to unknown SNR.

## **6.4 Conclusions**

Although this chapter's approach has arisen strongly, from the obstacles we found on our way in the generalisation of the RDT framework, it ended up leading us to some very interesting asymptotic results in the case of unconditional models. Indeed, the field of possibilities for the use of the Asymptotic Karlin-Rubin's theorem is wide, and a lot of applications might benefit from it, namely detection in the marine environment [Socheleau et al., 2015], [Leroy et al., 2016], but surely not only.

# Conclusions and perspectives

During these last three years of thesis, we tried to dig further in some directions. We also discovered new directions we did not have in mind in the beginning of this journey. The subject was about the detection of signals when their distribution is unknown by exploiting the inherent invariances of the problem and while targeting optimality. The first half of this manuscript recalled all the theoretical material we would need about decision theory in addition to the state of the art approaches for both conditional and unconditional models for detection. We also made a focus on the notion of invariance, which was central for the direction of research we took.

## 6.5 Conclusions about our contributions

The second half of this manuscript i.e. Chapter 4, 5 and 6 contain the main contributions of this manuscript:

- Chapter 4 reflects upon the suitability of the RDT framework for a distributed context, i.e. when the observation vectors are not collected by one sensor but a network of sensors. A stochastic signal of unknown distribution is in presence of an unknown interference and an additive white Gaussian noise and we want to decide, on the basis of several observation vectors, if the signal of interest is present. Such a problem with this little information about the signal of interest and the interference is unsolvable according to the standard approaches when optimality is aspired to. However, by casting this decision problem into the RDT framework, an optimal test in the sense of  $\gamma$ -MCCP criterion is obtained. Indeed, by bounding the lack of knowledge about both the signal of interest and the interference, optimality can be reached. The most interesting result about this chapter is the fact that equivalent detection performance can be guaranteed when the chosen configuration is centralised or distributed. Moreover, thanks to its natural scalability, using the RDT statistic to derive the test, allows to lessen the communication burden for the same level of performance when the decision is distributed instead of centralised. This work has been done in collaboration with Prashant Khanduri and Pramod K. Varshney from the department of electrical engineering and computer Science of Syracuse University, New York.
- In chapter 5 our purpose was to generalise the RDT framework to problems where the noise is not necessary white Gaussian and where the associated maximal in-

variant is not the Mahalanobis norm either. The signal of interest is still assumed with unknown distribution. To the best of our knowledge, such decision problems have not been treated in standard literature. We considered the same optimality criteria as in the RDT framework: the  $\gamma$ -MCCP criteria. After generalising every element of the RDT reasoning in accordance to the assumption about the noise and the maximal invariant, we proved a result that allows to find an optimal test in the sense of  $\gamma$ -MCCP, for the GRDT problem. This approach is in line with the spirit of the conditional model where, just like the RDT framework, the unknown distribution of the signal is not sought nor estimated. Which leads to optimality criteria that are non-integral as mentioned in the introduction. We worked on this chapter collaboratively with Guillaume Ansel from the signal and communication department of IMT Atlantique Brest.

- Chapter 6 discusses the notion of optimality in an asymptotic context, for the particular case where signals are assumed to be deterministic. AUMP tests are defined, then an asymptotic formulation of the Karlin-Rubin theorem is proven with the help of an asymptotic formulation of the Neyman-Pearson theorem. Thanks to the obtained theorem, AUMP and AUMPI tests can be derived for composite hypothesis testing problems. Finally, the asymptotic Karlin-Rubin theorem is used for a specific application where we want to test the presence of a deterministic signal in a subspace cone. Similarly to the RDT approach, the decision problem was defined on the basis of a tolerance to restrict uncertainties about the signal. This restriction, for the considered application, transpire through the assumption that the signal's proportion of energy in some known subspace cone is in fact bounded depending on the hypothesis that is true. This application brought to light the usefulness of the asymptotic approach for certain applications. It also allowed us to find that the famous state of the art test: the GLRT has in fact an AUMPI property of optimality for this specific problem.

Although our main contributions are contained in these three chapters, a small contribution about the notion of invariance was presented in the second chapter. In the end of chapter 2, we redefined the concept of invariance for problems where the signal is assumed of unknown distribution hence no family of distributions can be defined. Since the family of distributions is a key element according to the the standard definition of invariance, we needed to establish a more general formulation of invariance to correspond to RDT-like problems like for example the GRDT problem or any problem that considers the signal of interest random with unknown distribution.

## 6.6 Perspectives

Through my three-year period of research, we cleared away a few trails that seemed promising in the beginning but turned out to get one nowhere. Fortunately, it also opened our eyes concerning new directions of research we did not necessary foresee in the beginning of the thesis.

## The straightforward perspectives

- The first direct perspective from our works concerns Chapter 5. In this chapter, Theorem 5.3.4 gives the conditions for a test to be optimal for the GRDT problem. The question that would be interesting to investigate is what type of pairs: noise/maximal invariant can fall within the scope of the GRDT framework and verify the conditions of Theorem 5.3.4 in order to be possible to find optimal tests in the sense of  $\gamma$ -MCCP tests. As we stated in Chapter 5, the Gaussian noise and the Mahalanobis norm are appropriated for the RDT problem. More generally Gaussian noises and the Euclidean norms are a pair that we can associate one to another that can verify the GRDT framework. However we are interested in knowing if there are other shapes of noises from which can be defined appropriate maximal invariants, and vice versa, that can validate the generality of the GRDT framework.
- The second straightforward perspective revolves around the asymptotic aspect of optimality of the RDT. Since an exhaustive search for problems to ascertain the GRDT framework is hard to conduct, we can study an asymptotic formulation of the  $\gamma$ -MCCP criteria of optimality. We would consider scenarios where the signal of interest is random of unknown distribution, and where the noise is possibly asymptotically Gaussian. The maximal invariant statistic would be asymptotically the Mahalanobis norm. And analogously to the asymptotic approach we worked on, in Chapter 6, the RDT main result about  $\gamma$ -MCCP tests can be casted in an asymptotic perspective.

## The indirect perspectives

- A perspective that can be examined within the RDT framework is the sequential aspect for decision problems. Such an axis of research has already been investigated in [Khanduri et al., 2019] and [Khanduri et al., 2018]. The authors want to detect a random signal of unknown distribution while minimising the false alarm probability and the miss detection probability and diminishing the necessary number of samples to make a decision. However, in these papers, optimality is not reached in the sense of the classical RDT framework. Thus, a possible perspective would be to study the optimality aspect when the decision is sequential.
- Another and last perspective point would be to apply the RDT framework for cybersecurity-oriented applications. In this field, the intrusions are normally handled thanks to machine learning tools. The types of intrusions that might be used to infiltrate into the system are learned first in order to be detected afterwards. However, assuming that the malicious agent knows that, and will try to create an intrusion that is shaped so as to lure the system, machine-learning methods can fall into the trap. For such problematic, it is plausible that the RDT framework can provide appropriate solutions thanks to its robustness against uncertainties and the small number of parameters needed to configurate the decision process.



# Appendix A

Proof of equivalence between the initial problems (4.1) and (4.8), and the RDT formalised problems (4.5) and (4.11).

## A.1 Proof of Proposition 4.2.1

The initial detection problem for one sensor to solve is:

$$\begin{cases} h_0 = [\epsilon = 0] \\ h_1 = [\epsilon = 1] \end{cases} \quad (\text{A.1})$$

We assume that the two following properties are almost surely true:

$$\begin{cases} \mathcal{P}_0 : \forall i \quad \nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\xi}_i) \leq \tau & (a.s.) \\ \mathcal{P}_1 : \forall i \quad \nu_{\mathbf{C}_i}(\mathbf{A}_i \mathbf{S}_i + \mathbf{A}_i \boldsymbol{\xi}_i) > \tau' \geq \tau & (a.s.) \end{cases} \quad (\text{A.2})$$

And herewith we define the events  $\Omega_0 = [\nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\xi}_i) \leq \tau]$  and  $\Omega_1 = [\nu_{\mathbf{C}_i}(\mathbf{A}_i \mathbf{S}_i + \mathbf{A}_i \boldsymbol{\xi}_i) > \tau' \geq \tau]$ . We want to prove that the detection problem presented in equ. (A.1) is equivalent to the following detection problem (A.3) when the properties  $\mathcal{P}_0$  and  $\mathcal{P}_1$  are almost surely verified:

$$\begin{cases} \mathcal{H}_0 = [\nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\Theta}_i) \leq \tau] \\ \mathcal{H}_1 = [\nu_{\mathbf{C}_i}(\mathbf{A}_i \boldsymbol{\Theta}_i) > \tau'] \end{cases} \quad (\text{A.3})$$

with  $\boldsymbol{\Theta}_i = \epsilon \mathbf{S}_i + \boldsymbol{\xi}_i$ .

In other words, we want to know whether  $\mathcal{H}_1 = h_1$  (*a.s.*) and  $\mathcal{H}_0 = h_0$  (*a.s.*), when assuming  $\mathcal{P}_0$  (*a.s.*) and  $\mathcal{P}_1$  (*a.s.*).

We know that:  $\Omega = h_1 \uplus h_0$  and that  $\mathcal{P}_0$  (*a.s.*)  $\iff \mathbb{P}(\Omega_0^C) = 0$  and  $\mathcal{P}_1$  (*a.s.*)  $\iff \mathbb{P}(\Omega_1^C) = 0$ .

Proving  $h_1 = \mathcal{H}_1$  (*a.s.*) is equivalent to proving that  $\mathbb{P}[h_1 \Delta \mathcal{H}_1] = 0$  with



$$h_1 \triangle \mathcal{H}_1 = (h_1 \setminus \mathcal{H}_1) \cup (\mathcal{H}_1 \setminus h_1).$$

$$\begin{aligned} h_1 \setminus \mathcal{H}_1 &= h_1 \cap \mathcal{H}_1^C = h_1 \cap \mathcal{H}_1^C \cap (\Omega_1 \uplus \Omega_1^C) \\ &= (h_1 \cap \mathcal{H}_1^C \cap \Omega_1) \uplus (h_1 \cap \mathcal{H}_1^C \cap \Omega_1^C) \\ &= h_1 \cap \mathcal{H}_1^C \cap \Omega_1^C \quad (\text{as } h_1 \cap \mathcal{H}_1^C \cap \Omega_1 = \emptyset) \end{aligned}$$

Since  $(h_1 \setminus \mathcal{H}_1) \subset \Omega_1^C$  then:

$$\mathbb{P}[h_1 \setminus \mathcal{H}_1] = 0 \tag{A.4}$$

Now let us prove that  $\mathbb{P}[\mathcal{H}_1 \setminus h_1] = 0$ .

We can write:

$$\begin{aligned} \mathcal{H}_1 &= \mathcal{H}_1 \cap \Omega = \mathcal{H}_1 \cap (h_1 \uplus h_0) \\ &= (\mathcal{H}_1 \cap h_1) \uplus (\mathcal{H}_1 \cap h_0) \\ &= (\mathcal{H}_1 \cap h_1) \uplus (\mathcal{H}_1 \cap h_0 \cap \Omega) \\ &= (\mathcal{H}_1 \cap h_1) \uplus (\mathcal{H}_1 \cap h_0 \cap (\Omega_0 \uplus \Omega_0^C)) \\ &= (\mathcal{H}_1 \cap h_1) \uplus (\mathcal{H}_1 \cap h_0 \cap \Omega_0) \uplus (\mathcal{H}_1 \cap h_0 \cap \Omega_0^C) \\ &= (\mathcal{H}_1 \cap h_1) \uplus (\mathcal{H}_1 \cap h_0 \cap \Omega_0^C) \\ &\quad (\text{as: } \mathcal{H}_1 \cap h_0 \cap \Omega_0 = \emptyset) \end{aligned}$$

Since  $\mathcal{H}_1 \subset (h_1 \cup \Omega_0^C)$ , then:

$$\mathcal{H}_1 = \mathcal{H}_1 \cap (h_1 \cup \Omega_0^C) \tag{A.5}$$

Then, using Equ. (A.5), we can have:

$$\begin{aligned} \mathcal{H}_1 \setminus h_1 &= \mathcal{H}_1 \cap h_1^C \\ &= \mathcal{H}_1 \cap (h_1 \cup \Omega_0^C) \cap h_1^C \\ &= (\mathcal{H}_1 \cap h_1 \cap h_1^C) \cup (\mathcal{H}_1 \cap \Omega_0^C \cap h_1^C) \\ &= \mathcal{H}_1 \cap \Omega_0^C \cap h_1^C \\ &\quad (\text{as: } \mathcal{H}_1 \cap h_1 \cap h_1^C = \emptyset) \end{aligned}$$

Since  $(\mathcal{H}_1 \setminus h_1) \subset \Omega_0^C$ , then:

$$\mathbb{P}[\mathcal{H}_1 \setminus h_1] = 0 \tag{A.6}$$

Using Equ. (A.4) and (A.6) we finally prove that  $\mathbb{P}[h_1 \triangle \mathcal{H}_1] = 0$ , which is equivalent to:  $\mathcal{H}_1 = h_1$  (a.s.).

Similarly, we can demonstrate that  $\mathcal{H}_0 = h_0$  (a.s.) by proving that  $\mathbb{P}[h_0 \triangle \mathcal{H}_0] = 0$ , with  $h_0 \triangle \mathcal{H}_0 = (h_0 \setminus \mathcal{H}_0) \cup (\mathcal{H}_0 \setminus h_0)$ .

$$\begin{aligned} h_0 \setminus \mathcal{H}_0 &= h_0 \cap \mathcal{H}_0^C = h_0 \cap \mathcal{H}_0^C \cap (\Omega_0 \uplus \Omega_0^C) \\ &= (h_0 \cap \mathcal{H}_0^C \cap \Omega_0) \uplus (h_0 \cap \mathcal{H}_0^C \cap \Omega_0^C) \\ &= h_0 \cap \mathcal{H}_0^C \cap \Omega_0^C \quad (\text{as } h_0 \cap \mathcal{H}_0^C \cap \Omega_0 = \emptyset) \end{aligned}$$

Since  $(h_0 \setminus \mathcal{H}_0) \subset \Omega_0^C$  then:

$$\mathbb{P}[h_0 \setminus \mathcal{H}_0] = 0 \tag{A.7}$$

Now let us prove that  $\mathbb{P}[\mathcal{H}_0 \setminus h_0] = 0$ .

The following results will prove useful in the sequel.

$$\begin{aligned}
\mathcal{H}_0 \cap h_1 &= \mathcal{H}_0 \cap h_1 \cap \Omega \\
&= \mathcal{H}_0 \cap h_1 \cap (\Omega_1 \uplus \Omega_1^C) \\
&= (\mathcal{H}_0 \cap h_1 \cap \Omega_1) \uplus (\mathcal{H}_0 \cap h_1 \cap \Omega_1^C) \\
&= \mathcal{H}_0 \cap h_1 \cap \Omega_1^C \\
&\quad (\text{as: } \mathcal{H}_0 \cap h_1 \cap \Omega_1 = \emptyset)
\end{aligned} \tag{A.8}$$

We will also need:

$$\begin{aligned}
\mathcal{H}_0 &= \mathcal{H}_0 \cap \Omega \\
&= \mathcal{H}_0 \cap (h_1 \uplus h_0) \\
&= (\mathcal{H}_0 \cap h_1) \uplus (\mathcal{H}_0 \cap h_0) \\
&= (\mathcal{H}_0 \cap h_1 \cap \Omega_1^C) \uplus (\mathcal{H}_0 \cap h_0) \\
&\quad (\text{according to the final expression of A.8}) \\
&\Rightarrow \mathcal{H}_0 \subset h_0 \cup \Omega_1^C \\
&\Rightarrow \mathcal{H}_0 = \mathcal{H}_0 \cap (h_0 \cup \Omega_1^C)
\end{aligned} \tag{A.9}$$

And finally we can have:

$$\begin{aligned}
\mathcal{H}_0 \setminus h_0 &= \mathcal{H}_0 \cap h_0^C \\
&= H_0 \cap (h_0 \cup \Omega_1^C) \cap h_0^C \\
&\quad (\text{using the last expression of equation A.9}) \\
&= (H_0 \cap h_0 \cap h_0^C) \uplus (H_0 \cap \Omega_1^C \cap h_0^C) \\
&= H_0 \cap \Omega_1^C \cap h_0^C \quad (\text{as } H_0 \cap h_0 \cap h_0^C = \emptyset) \\
&\Rightarrow (\mathcal{H}_0 \setminus h_0) \subset \Omega_1^C \\
&\quad \text{and since } \mathbb{P}(\Omega_1^C) = 0
\end{aligned}$$

We finally prove that:

$$\mathbb{P}[\mathcal{H}_0 \setminus h_0] = 0 \tag{A.10}$$

According to (A.10) and (A.7) we prove that  $\mathbb{P}[\mathcal{H}_0 \triangle h_0] = 0$ , which is equivalent to:  $\mathcal{H}_0 = h_0$  (a.s.).

## A.2 Proof of Proposition 4.3.1

In order to prove that  $h_0^{\text{FC}} = \mathcal{H}_0^{\text{FC}}$  and  $h_1^{\text{FC}} = \mathcal{H}_1^{\text{FC}}$ , we only need to replace the properties  $\mathcal{P}_0$  and  $\mathcal{P}_1$  (A.2) by  $\mathcal{P}_0^{\text{FC}}$  and  $\mathcal{P}_1^{\text{FC}}$  (4.10), the definition of the events  $h_0$  and  $h_1$  in (A.1) by:  $h_0^{\text{FC}} = [\epsilon = 0]$  and  $h_1^{\text{FC}} = [\epsilon = 1]$  and finally the definition of the events  $\mathcal{H}_1$  and  $\mathcal{H}_0$  (A.3) by  $\mathcal{H}_0^{\text{FC}} = [\nu_{\mathcal{C}}(\mathbf{A}\Theta) \leq p\tau]$  and  $\mathcal{H}_1^{\text{FC}} = [\nu_{\mathcal{C}}(\mathbf{A}\Theta) > p\tau']$ . The rest of the demonstration can be held similarly to the proof of Proposition (4.2.1).



# Appendix B

## Proof of the Asymptotic Neyman Pearson Theorem 6.1.1

First, according to the Portmanteau Theorem we have

$$\forall \varphi \in \Psi_n \quad \limsup_n \mathbb{E}_\theta[\varphi(\mathbf{Y}_n)] = \lim_n \mathbb{E}_\theta[\varphi(\mathbf{Y}_n)] = \mathbb{E}_\theta[\varphi(\mathbf{Y})] \quad (\text{B.1})$$

Let us consider  $\tilde{\mathcal{T}} \in \Psi_{\mathbf{Y}}$  such that  $\mathbb{E}_{\theta_0}[\mathcal{T}(\mathbf{Y})] \leq \gamma$ . By using the definitions of  $\mathcal{T}$  and  $\tilde{\mathcal{T}}$ , we have the following inequality  $\forall \mathbf{y} \in \mathbb{R}^N$

$$\begin{aligned} & (\tilde{\mathcal{T}}(\mathbf{y}) - \mathcal{T}(\mathbf{y}))(f_{\mathbf{Y}}^{\theta_1}(\mathbf{y}) - \eta f_{\mathbf{Y}}^{\theta_0}(\mathbf{y})) \leq 0 \\ \Rightarrow & \int_{\mathbb{R}^N} \tilde{\mathcal{T}}(\mathbf{y}) f_{\mathbf{Y}}^{\theta_1}(\mathbf{y}) d\mathbf{y} - \int_{\mathbb{R}^N} \mathcal{T}(\mathbf{y}) f_{\mathbf{Y}}^{\theta_1}(\mathbf{y}) d\mathbf{y} \leq \eta \left( \int_{\mathbb{R}^N} \tilde{\mathcal{T}}(\mathbf{y}) f_{\mathbf{Y}}^{\theta_0}(\mathbf{y}) d\mathbf{y} - \int_{\mathbb{R}^N} \mathcal{T}(\mathbf{y}) f_{\mathbf{Y}}^{\theta_0}(\mathbf{y}) d\mathbf{y} \right) \\ \Rightarrow & \mathbb{E}_{\theta_1}[\tilde{\mathcal{T}}(\mathbf{Y})] - \mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y})] \leq \eta (\mathbb{E}_{\theta_0}[\tilde{\mathcal{T}}(\mathbf{Y})] - \mathbb{E}_{\theta_0}[\mathcal{T}(\mathbf{Y})]) \end{aligned} \quad (\text{B.2})$$

As the right term of the previous inequality is negative and by applying the Portmanteau Theorem we can deduce that

$$\begin{aligned} & \mathbb{E}_{\theta_1}[\tilde{\mathcal{T}}(\mathbf{Y})] - \mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y})] \\ &= \lim_n \mathbb{E}_{\theta_1}[\tilde{\mathcal{T}}(\mathbf{Y}_n)] - \lim_n \mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y}_n)] \\ &= \lim_n \{ \mathbb{E}_{\theta_1}[\tilde{\mathcal{T}}(\mathbf{Y}_n)] - \mathbb{E}_{\theta_1}[\mathcal{T}(\mathbf{Y}_n)] \} \\ &\leq 0 \end{aligned} \quad (\text{B.3})$$

Therefore the test  $\mathcal{T}$  is Asymptotically Most Powerful (AMP) among all tests belonging to  $\Psi_{\mathbf{Y}}$ .



# Appendix C

## Proof of the Asymptotic Karlin Rubin Theorem 6.2.1

We can set apart the proof into three major steps. The first step is to demonstrate that the test  $\mathcal{T}'$  is AUMP for the detection problem (6.7) of theorem (6.1.1). The second is to prove that the test  $\mathcal{T}'$  is also AUMP when the alternative hypothesis  $\mathcal{H}_1$  is composite, and last when the null hypothesis  $\mathcal{H}_0$  is composite.

**Lemma C.0.1.** *For the detection problem and the hypotheses of test of theorem (6.1.1), if the likelihood ratio  $\Lambda_{\theta_0, \theta_1}(\mathbf{y})$  is a strictly increasing function of  $V(\mathbf{y})$ , then the test  $\mathcal{T}'$  is AUMP among all the tests belonging to the family of tests  $\Psi_{\mathbf{Y}}$ .*

*Proof.*

$$\mathcal{T}(\mathbf{y}) = \begin{cases} 1 & \text{if } \Lambda_{\theta_0, \theta_1}(\mathbf{y}) \geq \eta \\ 0 & \text{otherwise,} \end{cases} \quad (\text{C.1})$$

If we put  $\Lambda_{\theta_0, \theta_1}(\mathbf{y}) = h_{\theta_0, \theta_1}(V(\mathbf{y}))$ , then  $\mathcal{T}(\mathbf{y}) = \begin{cases} 1 & \text{if } h_{\theta_0, \theta_1}(V(\mathbf{y})) \geq \eta \\ 0 & \text{otherwise} \end{cases}$

If  $h_{\theta_0, \theta_1}(V(\mathbf{y}))$  is strictly increasing in  $V(\mathbf{y})$  then  $\mathcal{T}(\mathbf{y}) = \mathcal{T}'(\mathbf{y}) = \begin{cases} 1 & \text{if } V(\mathbf{y}) \geq \lambda \\ 0 & \text{otherwise} \end{cases}$

where  $\lambda = h_{\theta_0, \theta_1}^{-1}(\eta)$ . □

**Lemma C.0.2.** *The test  $\mathcal{T}'$  is AUMP among all the tests belonging to  $\Psi_{\mathbf{Y}}$  for testing*

$$\begin{cases} \mathcal{H}_0'' : \theta = \theta_0 \\ \mathcal{H}_1'' : \theta > \theta_0 \end{cases} \quad (\text{C.2})$$

*Proof.* The test  $\mathcal{T}'$  is independent of  $\theta_1$  and only depends on the two parameters  $\theta_0$  and  $\gamma$ . Moreover  $h_{\theta_0, \theta_1}$  is strictly increasing for  $\theta_0 < \theta_1$ , consequently the test is AUMP among the tests belonging to  $\Psi_{\mathbf{Y}}$  for any  $\theta_1$  and  $\theta_0$  satisfying  $\theta_0 < \theta_1$ . □

In order to demonstrate the third step, we must prove that

$$\limsup_n \sup_{\theta \leq \theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] \leq \lim_n \mathbb{E}_{\theta_0}[\mathcal{T}'(\mathbf{Y}_n)] = \gamma \quad (\text{C.3})$$

(i.e. that the constraint about the size is asymptotically verified for any  $\theta \leq \theta_0$ ) We proceed in two stages.

**Lemma C.0.3.**

$$\limsup_n \sup_{\theta \leq \theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] = \sup_{\theta \leq \theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \quad (\text{C.4})$$

*Proof.* According to Portmanteau theorem [Billingsley, 1995, Th. 2.1] we have:  $\lim_n \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] = \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})]$

For any  $n \in \mathbb{N}$  and any  $\theta \leq \theta_0$  we define  $M$  and  $M_n$  such that:

$$\begin{aligned} M &= \sup_{\theta \leq \theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \\ M_n &= \sup_{\theta \leq \theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] \end{aligned}$$

For  $\epsilon > 0$ , and  $n \geq N_\epsilon$  with  $N_\epsilon \in \mathbb{N}$ , we want to prove that we have:  $\lim_n M_n = M$ , i.e.  $M + \epsilon \geq M_n$  and  $M_n + \epsilon \geq M$ . Let us prove these two inequalities by contradiction. First, let us prove that  $M + \epsilon \geq M_n$ . We assume that

$$M + \epsilon < M_n$$

Then, there exists  $\theta \leq \theta_0$  such that

$$M + \epsilon < \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] \leq M_n$$

Consequently:  $M + \epsilon \leq \lim_n \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)]$ . According to Portmanteau theorem we have  $M + \epsilon \leq \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})]$ . But since  $\mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \leq M$ , we obtain the absurd inequality

$$M + \epsilon \leq M$$

Thus:

$$M + \epsilon \geq M_n$$

Now, let us prove that  $M_n + \epsilon \geq M$ . We assume that

$$M_n + \epsilon < M$$

There exists  $\theta \leq \theta_0$  such that

$$M_n + \epsilon < \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \leq M$$

Then  $\mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] + \epsilon < \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})]$  and  $\lim_n \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] \leq \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] - \epsilon$ . Finally, according to Portmanteau theorem, we obtain the inequality:

$$\mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] \leq \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y})] - \epsilon$$

This last inequality implies that  $\epsilon \leq 0$  which is absurd. Hence we prove that:  $M_n + \epsilon \geq M$ , which concludes the demonstration. □

**Lemma C.0.4.**

$$\sup_{\theta \leq \theta_0} \mathbb{E}_\theta[\mathcal{T}'(\mathbf{Y}_n)] = \mathbb{E}_{\theta_0}[\mathcal{T}'(\mathbf{Y})] \quad (\text{C.5})$$

---

*Proof.* We consider the test  $\mathcal{T}'$  used for testing  $\begin{cases} \mathcal{H}_0'' : \theta = \theta_2 \\ \mathcal{H}_1'' : \theta > \theta_2 \end{cases}$ , with  $\theta_2 < \theta_0$

The test is of asymptotic size  $\gamma''$ , as  $\mathcal{T}' \equiv \mathcal{T}$  and the test  $\mathcal{T}$  is unbiased, we can deduce that  $\mathbb{E}_{\theta_2}[\mathcal{T}'(\mathbf{Y})] \leq \mathbb{E}_{\theta_0}[\mathcal{T}'(\mathbf{Y})]$ ,  $\forall \theta_2 < \theta_0$ . Hence  $\sup_{\theta \leq \theta_0} \mathbb{E}_{\theta}[\mathcal{T}'(\mathbf{Y})] = \mathbb{E}_{\theta_0}[\mathcal{T}'(\mathbf{Y})]$ .  $\square$

From lemma (C.0.3) and (C.0.4) we can write  $\lim_n \sup_{\theta \leq \theta_0} \mathbb{E}_{\theta}[\mathcal{T}'(\mathbf{Y}_n)] = \mathbb{E}_{\theta_0}[\mathcal{T}'(\mathbf{Y})] = \gamma$ .

Thereby we conclude the proof of theorem (6.2.1).





# Appendix D

## Proof of Theorem 3.1.2

The following Lemmas will prove useful in the sequel:

**Lemma D.0.1.** *Given any  $\eta \in (0, \infty)$ ,  $\mathcal{R}(\cdot, \eta)$  is strictly decreasing.*

*Proof.* Let  $\rho$  and  $\rho'$  be two real numbers such that  $0 \leq \rho < \rho' < \infty$ . Let  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}'$  be two collinear vectors of  $\mathbb{R}^N$  such that  $\nu_{\mathcal{C}}(\boldsymbol{\theta}) = \rho$  and  $\|\boldsymbol{\theta}'\| = \rho'$ . According to the equation:  $\mathbb{P}[\nu_{\mathcal{C}}(Z) \leq \eta] = \mathcal{R}(\nu_{\mathcal{C}}(\boldsymbol{\theta}), \eta)$ ,  $\mathcal{R}(\rho, \eta) = \int_{B(\boldsymbol{\theta}, \eta)} f(x) dx$  and  $\mathcal{R}(\rho', \eta) = \int_{B(\boldsymbol{\theta}', \eta)} f(x) dx$  where  $f$  is the probability density function of  $X$  and  $B(\boldsymbol{\theta}, \eta)$  (resp.  $B(\boldsymbol{\theta}', \eta)$ ) is the closed ball, in  $\mathbb{R}^n$ , centered at  $\boldsymbol{\theta}$  (resp.  $\boldsymbol{\theta}'$ ) with radius  $\eta$ . We have  $\mathcal{R}(\rho, \eta) - \mathcal{R}(\rho', \eta) = \int_{B(\boldsymbol{\theta}, \eta) \setminus B(\boldsymbol{\theta}', \eta)} (f(x) - f(\boldsymbol{\theta} + \boldsymbol{\theta}' - x)) dx$ . Let  $(e_1, e_2, \dots, e_n)$  be an orthonormal basis of  $\mathbb{R}^n$  such that  $\boldsymbol{\theta} = \rho e_1$  and  $\boldsymbol{\theta}' = \rho' e_1$ . We have  $\|\boldsymbol{\theta} + \boldsymbol{\theta}' - x\|^2 - \|x\|^2 = (\rho + \rho')(\rho + \rho' - 2x_1)$  for any  $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ . If  $x \in B(\boldsymbol{\theta}, \eta) \setminus B(\boldsymbol{\theta}', \eta)$ , then  $\|x - \boldsymbol{\theta}'\| > \|x - \boldsymbol{\theta}\|$ , which implies that  $(\rho' - \rho)(\rho + \rho' - 2x_1) > 0$  and, thus, that  $\rho + \rho' - 2x_1 > 0$  since  $\rho' > \rho$ . Therefore,  $\|\boldsymbol{\theta} + \boldsymbol{\theta}' - x\| > \|x\|$ . Since  $f$  decreases strictly with the norm of its argument, it follows that  $f(x) - f(\boldsymbol{\theta} + \boldsymbol{\theta}' - x) > 0$  so that  $\mathcal{R}(\rho, \eta) > \mathcal{R}(\rho', \eta)$  and the proof is complete.  $\square$

**Lemma D.0.2.**  $\forall \gamma \in (0, \infty)$ ,  $\lambda_\gamma$  is strictly increasing and continuous in  $(0, \infty)$ .

*Proof.* [ $\lambda_\tau$  is strictly increasing] : Let  $\rho$  and  $\rho'$  be two non-negative real number such that  $\rho < \rho'$ . According to Lemma D.0.1,  $\mathcal{R}(\rho', \lambda_\tau(\rho)) < \mathcal{R}(\rho, \lambda_\tau(\rho))$ . The right hand side (rhs) in this inequality equals  $1 - \gamma$  and, thus,  $\mathcal{R}(\rho', \lambda_\tau(\rho'))$ . The result then follows from the fact that  $\mathcal{R}(\rho', \cdot)$  is strictly increasing.

[Continuity of  $\lambda_\tau$ ] : Given  $\rho_0 \in [0, \infty)$ ,  $\lambda_\tau$  is strictly increasing. Therefore, there exist a limit  $\lambda_\tau(\rho_0^-) \in [0, \infty)$  when  $\rho$  tends to  $\rho_0$  from below and a limit  $\lambda_\tau(\rho_0^+) \in [0, \infty)$  when  $\rho$  tends to  $\rho_0$  from above. Since  $\mathcal{R}$  is continuous in the plane and  $\mathcal{R}(\rho, \lambda_\tau(\rho)) = 1 - \gamma$  for every  $\rho \geq 0$ ,  $\mathcal{R}(\rho_0, \lambda_\tau(\rho_0^-)) = \mathcal{R}(\rho_0, \lambda_\tau(\rho_0^+)) = 1 - \gamma$ . Since  $\mathcal{R}(\rho_0, \cdot)$  is one-to-one,  $\lambda_\tau(\rho_0^-) = \lambda_\tau(\rho_0^+) = \lambda_\tau(\rho_0)$  and  $\lambda_\tau$  is continuous.  $\square$

**Lemma D.0.3.** *A test  $\mathcal{T}$  has constant power function on  $\Upsilon_\rho$  with  $\rho \in (0, \infty)$  if and only if:  $\mathbb{P}[\mathcal{T}(\boldsymbol{\Xi} + \mathbf{X}) = 1] = \beta_{\mathcal{T}}(\boldsymbol{\theta})$  for any  $\boldsymbol{\theta} \in \Upsilon_\rho$  and any  $\boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^d)$*

independent of  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$  and such that  $\boldsymbol{\Xi} \in \Upsilon_\rho$  (a.s).

*Proof.* The lemma can be proven thanks to the fact that given any test  $\mathcal{T}$  and any  $\rho \in (0, \infty)$ , we have:

$$\mathbb{P}[\mathcal{T}(\boldsymbol{\Xi} + \mathbf{X}) = 1] = \int_{\Upsilon_\rho} \beta_{\mathcal{T}}(\boldsymbol{\theta}) \mathbb{P}_{\boldsymbol{\Xi}}(d\boldsymbol{\theta}) \quad (\text{D.1})$$

for all  $\boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^N)$ , independent of  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$  and such that  $\boldsymbol{\Xi} \in \Upsilon_\rho$  (a-s).  $\square$

**Lemma D.0.4.** *Let matrix  $\Phi = \Delta^{-1/2}U^T$  defined as in the notation block. Given  $\rho_0 < \rho_1$ , consider any two random vectors  $\boldsymbol{\Xi}_0$  and  $\boldsymbol{\Xi}_1$  such that  $\Phi(\boldsymbol{\Xi}_0 - \boldsymbol{\theta}_0)$  and  $\Phi(\boldsymbol{\Xi}_1 - \boldsymbol{\theta}_0)$  are uniformly distributed on  $\rho_0 S^{d-1}$  and  $\rho_1 S^{d-1}$  respectively. Given any  $\gamma \in (0, 1)$ , the threshold test  $\mathcal{T}_{\lambda_\gamma(\rho_0)}$  is most powerful with size  $\gamma$  for testing  $\mathcal{H}_0 : \mathbf{Y} = \boldsymbol{\Xi}_0 + \mathbf{X}$  against  $\mathcal{H}_1 : \mathbf{Y} = \boldsymbol{\Xi}_1 + \mathbf{X}$  with  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$ . The power of this test is  $1 - \mathcal{R}(\rho_1, \lambda_\gamma(\rho_0))$ .*

*Proof.* Given any  $\mathbf{y} \in \mathbb{R}^N$ ,  $\mathcal{L}(\mathbf{y}) = f_{\boldsymbol{\Xi}_1 + \mathbf{X}}(\mathbf{y})/f_{\boldsymbol{\Xi}_0 + \mathbf{X}}(\mathbf{y})$  is the likelihood ratio for testing  $\mathcal{H}_0$  against  $\mathcal{H}_1$ , where  $f_{\boldsymbol{\Xi}_i + \mathbf{X}}$  is the probability density function (pdf) of  $\boldsymbol{\Xi}_i + \mathbf{X}$ ,  $i = 0, 1$ . The Neyman-Pearson Lemma [Lehmann and Romano, 2005a, Theorem 3.2.1, Sec. 3.2, p. 60] implies the existence of a most powerful test with size  $\gamma$  for testing  $\mathcal{H}_0$  against  $\mathcal{H}_1$ . This test accepts (resp. rejects)  $\mathcal{H}_0$  if  $\mathcal{L}(\mathbf{y}) < \lambda$  (resp.  $\mathcal{L}(\mathbf{y}) > \lambda$ ), where  $\lambda$  is such that  $\mathbb{P}[\mathcal{L}(\boldsymbol{\Xi}_0 + \mathbf{X}) > \lambda] = \gamma$ . Under each hypothesis  $\mathcal{H}_i$ ,  $i = 0, 1$ ,  $\boldsymbol{\Xi}_i + \mathbf{X} = \Phi^{-1}(\boldsymbol{\Lambda}_i + \mathbf{Z}) + \boldsymbol{\theta}_0$  where  $\boldsymbol{\Lambda}_i = \Phi(\boldsymbol{\Xi}_i - \boldsymbol{\theta}_0)$  is uniformly distributed on  $\rho_i S^{n-1}$  and  $\mathbf{Z} = \Phi\mathbf{X} \sim \mathcal{N}(0, \mathbf{I}_N)$ . The pdf of  $\boldsymbol{\Lambda}_i + \mathbf{Z}$  for  $i \in \{0, 1\}$  is given by [Pastor et al., 2002, Proposition V.1, p. 232]. Therefore, the pdf's  $f_{\boldsymbol{\Xi}_0 + \mathbf{X}}$  and  $f_{\boldsymbol{\Xi}_1 + \mathbf{X}}$  can then be calculated by using [Muirhead, 1982, Theorem 2.1.4, p. 57] and the standard change of variable [Billingsley, 1995, Theorem 17.2, p. 225]. By taking into account the equation:  $\nu_{\mathbf{C}}(\mathbf{y}) = \|\Phi\mathbf{y}\|_2$ , the reader will then find that  $\mathcal{L}(\mathbf{y}) = f_{\chi_N^2(\rho_1^2)}(\nu_{\mathbf{C}}(\mathbf{y} - \boldsymbol{\theta}_0)^2)/f_{\chi_N^2(\rho_0^2)}(\nu_{\mathbf{C}}(\mathbf{y} - \boldsymbol{\theta}_0)^2)$ , where, for  $i \in \{0, 1\}$ ,  $f_{\chi_N^2(\rho_i^2)}$  is the pdf of the non-central  $\chi^2$  distribution with  $N$  degrees of freedom and non-central parameter  $\rho_i^2$ . The family of non-central  $\chi^2$  distributions with  $N$  degrees of freedom has monotone likelihood ratio with its non-central parameter [Eaton, 1983, Example A.1, pp. 468 – 469]. The most powerful test for testing  $\mathcal{H}_0$  against  $\mathcal{H}_1$  is therefore the threshold test  $\mathcal{T}_{\lambda'}$  from above  $\lambda'$  on the Mahalanobis distance to  $\boldsymbol{\theta}_0$ , where  $\lambda'$  is such that  $\mathbb{P}[\mathcal{T}_{\lambda'}(\boldsymbol{\Xi}_0 + \mathbf{X}) = 1] = \gamma$ . From  $\nu_{\mathbf{C}}(\mathbf{y}) = \|\Phi\mathbf{y}\|_2$  and [Pastor et al., 2002, Proposition V.1 & Eq. (19), p. 232], we derive that  $\mathbb{P}[\mathcal{T}_{\lambda'}(\boldsymbol{\Xi}_i + \mathbf{X}) = 1] = 1 - \mathcal{R}(\rho_i, \lambda')$  for  $i = 0, 1$ . Thence, the value of  $\lambda'$  and the power of  $\mathcal{T}_{\lambda'}$ .  $\square$

**Lemma D.0.5.**  $\mathcal{T}_\eta$  with  $\eta \geq 0$  has constant power function:  $\beta_{\mathcal{T}_\eta}(\boldsymbol{\theta}) = 1 - \mathcal{R}(\rho, \eta)$ ,  $\boldsymbol{\theta} \in \Upsilon_\rho$ .

We now establish the proof of Theorem 3.1.2. To begin with, Lemma D.0.5 straightforwardly implies that  $\mathcal{T}_{\lambda_\gamma(\tau)}$  has constant power function on every  $\Upsilon_\rho$ , with  $\beta_{\mathcal{T}_{\lambda_\gamma(\tau)}}(\boldsymbol{\theta}) = 1 - \mathcal{R}(\rho, \lambda_\gamma(\tau))$  for any  $\rho \geq 0$  and any  $\boldsymbol{\theta} \in \Upsilon_\rho$ . It now follows from

---

Lemma D.0.1 that  $\beta_{\mathcal{T}_{\lambda_\gamma(\tau)}}(\boldsymbol{\theta}) \leq 1 - \mathcal{R}(\tau, \lambda_\gamma(\tau))$  for any  $\boldsymbol{\theta} \in \mathbb{R}^n$  with  $\nu_C(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in [0, \tau]$ . Lemma D.0.2 then implies that  $\mathcal{T}_{\lambda_\gamma(\tau)}$  has size  $\gamma$ . Now, we consider two real values  $\rho_0$  and  $\rho_1$  such that  $\rho_0 < \tau < \rho_1$  and assume that  $\mathcal{T}$  is of level  $\gamma$  and has constant power function on  $\Upsilon_{\rho_1}$ . Let  $\boldsymbol{\Xi}_0$  and  $\boldsymbol{\Xi}_1$  be any two elements of  $\mathcal{M}(\Omega, \mathbb{R}^N)$ , independent of  $\mathbf{X}$  and such that  $\boldsymbol{\Xi}_i \in \Upsilon_{\rho_i}$  (a-s) for  $i \in \{0, 1\}$ . On the one hand, it follows from the equation:  $\mathbb{P}[\mathcal{T}(\boldsymbol{\Xi} + \mathbf{X}) = 1] = \int_{\Upsilon_\rho} \beta_{\mathcal{T}}(\boldsymbol{\theta}) P_{\boldsymbol{\Xi}}(d\boldsymbol{\theta})$  that  $\mathbb{P}[\mathcal{T}(\boldsymbol{\Xi}_0 + \mathbf{X}) = 1] \leq \gamma$ . On the other hand, given any  $\boldsymbol{\theta} \in \Upsilon_{\rho_1}$ , Lemma D.0.3 implies that  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) = \mathbb{P}[\mathcal{T}(\boldsymbol{\Xi}_1 + \mathbf{X}) = 1]$ . Therefore,  $\mathcal{T}$  has level  $\gamma$  and power equal to  $\beta_{\mathcal{T}}(\boldsymbol{\theta})$  for testing  $\mathcal{H}_0 : \mathbf{Y} = \boldsymbol{\Xi}_0 + \mathbf{X}$  against  $\mathcal{H}_1 : \mathbf{Y} = \boldsymbol{\Xi}_1 + \mathbf{X}$ . This holding true for any  $\boldsymbol{\Xi}_i \in \Upsilon_{\rho_i}$  (a-s) with  $i = 0, 1$ , we choose  $\boldsymbol{\Xi}_0$  and  $\boldsymbol{\Xi}_1$  so that  $\Phi(\boldsymbol{\Xi}_0 - \boldsymbol{\theta}_0)$  and  $\Phi(\boldsymbol{\Xi}_1 - \boldsymbol{\theta}_0)$  are uniformly distributed on  $\rho_0 S^{n-1}$  and  $\rho_1 S^{n-1}$ , respectively. According to Lemma D.0.4,  $\mathcal{T}_{\lambda_\gamma(\rho_0)}$  is most powerful with size  $\gamma$  and power equal to  $1 - \mathcal{R}(\rho_1, \lambda_\gamma(\rho_0))$  for testing  $\mathcal{H}_0$  against  $\mathcal{H}_1$ . Therefore,  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) \leq 1 - \mathcal{R}(\rho_1, \lambda_\gamma(\rho_0))$ . The rhs in this inequality tends to  $1 - \mathcal{R}(\rho_1, \lambda_\gamma(\tau))$  when  $\rho_0$  tends to  $\tau$  because of the continuity of  $\mathcal{R}(\rho_1, \cdot)$  and  $\lambda_\gamma$ . We have  $1 - \mathcal{R}(\rho_1, \lambda_\gamma(\tau)) = \beta_{\mathcal{T}_{\lambda_\gamma(\tau)}}(\boldsymbol{\theta})$  because  $\boldsymbol{\theta} \in \Upsilon_{\rho_1}$ . Therefore,  $\beta_{\mathcal{T}}(\boldsymbol{\theta}) \leq \beta_{\mathcal{T}_{\lambda_\gamma(\tau)}}(\boldsymbol{\theta})$ .



# Appendix E

## Proof of lemma 3.2.1

Let  $\mathcal{T}$  be some test. Let  $\boldsymbol{\theta} \in \mathbb{R}^N$  such that  $\nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq \tau$ . For any  $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  such that  $\boldsymbol{\Theta} = \boldsymbol{\theta}$  (a-s),  $\mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau] = \beta_{\mathcal{T}}(\boldsymbol{\theta})$ . Therefore,  $\alpha_{\mathcal{T}} \leq \sup_{\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N): \mathbb{P}[\nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau]$ , since  $\boldsymbol{\theta}$  has been chosen arbitrarily.

Conversely, for any  $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N)$  independent of  $X \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$ , any  $\boldsymbol{\theta} \in \mathbb{R}^N$  and any borel set  $B$  of  $\mathbb{R}^N$ , we have:

$$\mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1, \nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \in B \mid \boldsymbol{\Theta} = \boldsymbol{\theta}] = \mathbf{I}_B(\nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0))\beta_{\mathcal{T}}(\boldsymbol{\theta}), \quad (\text{E.1})$$

where  $\mathbf{I}_B$  is the indicator function of  $B$ :  $\mathbf{I}_B(x) = 1$  if  $x \in B$  and  $\mathbf{I}_B(x) = 0$ , otherwise. In particular, if  $\boldsymbol{\Theta}$  is such that  $\mathbb{P}[\nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau] \neq 0$  and is independent of  $X \sim \mathcal{N}(0, \mathbf{C})$ , it follows from the definition of a conditional probability and Eq. (E.1) with  $B = [0, \tau]$  that

$$\mathbb{P}([\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1] \cap [\nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau]) = \int_{\{\boldsymbol{\theta} \in \mathbb{R}^N: \nu_{\mathbf{C}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq \tau\}} \beta_{\mathcal{T}}(\boldsymbol{\theta}) \mathbb{P}_{\boldsymbol{\Theta}}(d\boldsymbol{\theta})$$

The rhs in this equality is less than or equal to  $\alpha_{\mathcal{T}} \mathbb{P}[\nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau]$ . From Bayes's rule, we get that  $\sup_{\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^N): \mathbb{P}[\nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \nu_{\mathbf{C}}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) \leq \tau] \leq \alpha_{\mathcal{T}}$ , which completes the proof.



# Appendix F

## Proof of Theorem 3.2.2

We need the following results to prove Theorem 3.2.2.

**Lemma F.0.1.** *A  $\gamma$ -MCP test with constant power function on every  $\Upsilon_\rho \in \mathfrak{F}$  is  $\gamma$ -MCCP.*

*Proof.* See [Pastor and Nguyen, 2013a, App. E]. □

**Lemma F.0.2.** *A test  $\mathcal{T}$  has constant power function on every  $\Upsilon_\rho \in \mathfrak{F}$  if and only if, for any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$ ,  $\mathcal{T}$  has constant conditional power function given  $\Theta \in \Upsilon_\rho$  for  $\mathbb{P}_{\nu_{\mathcal{C}}(\Theta - \theta_0)}$  - almost every  $\rho \geq 0$ .*

*Proof.* See [Pastor and Nguyen, 2013a, App. E]. □

**Lemma F.0.3.** *Threshold tests  $\mathcal{T}_\lambda$  with  $\lambda \geq 0$  on the Mahalanobis distance to  $\theta_0$  has constant power function on every  $\Upsilon_\rho \in \mathfrak{F}$  with  $\beta_{\mathcal{T}_\lambda}(\theta) = 1 - \mathcal{R}(\rho, \lambda)$ .*

According to lemma F.0.1, if a test  $\mathcal{T}$  is  $\gamma$ -MCP and has constant power function on every  $\Upsilon_\rho \in \mathfrak{F}$  then  $\mathcal{T}$  is  $\gamma$ -MCCP. From theorem 3.1.2 the test  $\mathcal{T}_{\lambda_\gamma(\tau)}$  is  $\gamma$ -MCP, and from lemma F.0.3 the test  $\mathcal{T}_{\lambda_\gamma(\tau)}$  has constant power function on every  $\Upsilon_\rho \in \mathfrak{F}$ . Consequently the test  $\mathcal{T}_{\lambda_\gamma(\tau)}$  is  $\gamma$ -MCCP. Concerning the second part of theorem 3.2.2, it is proven by the following theorem that derives from lemma F.0.2 and F.0.3.

**Lemma F.0.4.** *Given the threshold test  $\mathcal{T}_\lambda$  with  $\lambda \geq 0$  and any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)$ ,  $\mathbb{P}[\mathcal{T}_\lambda(\Theta + \mathbf{X}) = 1 \mid \Theta \in \Upsilon_\rho] = 1 - \mathcal{R}(\rho, \lambda)$  for  $\mathbb{P}_{\nu_{\mathcal{C}}(\Theta - \theta_0)}$  - almost every  $\rho \in [0, \infty)$ .*





# Appendix G

## Proof of lemma 5.3.1

Let  $\mathcal{T}$  be some test. Let  $\theta \in \mathbb{R}^N$  such that  $\mathcal{M}(\theta) \leq \tau$ . For any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)\mathbb{R}^N$  such that  $\Theta = \theta$  almost surely, we have  $\mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \mathcal{M}(\Theta) \leq \tau] = \beta_{\mathcal{T}}(\theta)$ . Since this is true for any  $\theta$  such that  $\mathcal{M}(\theta) \leq \tau$ , we have:

$$\alpha_{\mathcal{T}} \leq \sup_{\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)(\Omega, \mathbb{R}^N): \mathbb{P}[\mathcal{M}(\Theta) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \mathcal{M}(\Theta) \leq \tau]$$

Conversely, for any  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)\mathbb{R}^N$  independent of  $\mathbf{X}$ , any  $\theta \in \mathbb{R}^N$ , and any borel set  $B$  of  $\mathbb{R}$ , we have:

$$\mathbb{P}\left([\mathcal{T}(\Theta + X) = 1] \cap [\mathcal{M}(\Theta) \in B] \mid \Theta = \theta\right) = \mathbf{I}_B(\mathcal{M}(\theta))\beta_{\mathcal{T}}(\theta) \quad (\text{G.1})$$

where  $\mathbf{I}_B$  is the indicator function of  $B$ :  $\mathbf{I}_B(x) = 1$  if  $x \in B$  and  $\mathbf{I}_B(x) = 0$  otherwise. In particular, if  $\Theta$  is such that  $\mathbb{P}[\mathcal{M}(\Theta) \leq \tau] \neq 0$  and is independent of  $\mathbf{X}$ , it follows from the definition of a conditional probability and G.1 with  $B = ]-\infty, \tau]$  that:

$$\mathbb{P}\left([\mathcal{T}(\Theta + X) = 1] \cap [\mathcal{M}(\Theta) \leq \tau]\right) = \int_{\{\theta \in \mathbb{R}^N, \mathcal{M}(\theta) \leq \tau\}} \beta_{\mathcal{T}}(\theta)\mathbb{P}_{\Theta}(d\theta)$$

The right hand side of this equality is less than or equal to  $\alpha_{\mathcal{T}}\mathbb{P}[\mathcal{M}(\Theta) \leq \tau]$ . From Bayes' rule, we get that:

$$\sup_{\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N)(\Omega, \mathbb{R}^N): \mathbb{P}[\mathcal{M}(\Theta) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \mathcal{M}(\Theta) \leq \tau] \leq \alpha_{\mathcal{T}}$$

which completes the proof.



# Appendix H

## Proof of Theorem 5.3.4

Assume that for any  $\rho > \tau$  there exists two random variables  $\Theta_0$  and  $\Theta_1$  with  $\Theta_0 \in \Upsilon_\tau$  and  $\Theta_1 \in \Upsilon_\rho$  almost surely, such that the test  $\mathcal{T}^*$  is most powerful with level  $\gamma$  for testing 5.8. Let  $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N) \mathbb{R}^d$ . Let  $\rho > \tau$  and let  $\mathcal{T}$  be a test with level  $\gamma$  for the RDT problem and constant conditional power function given  $\Theta \in \Upsilon_\rho$ . By hypothesis, there exists  $\Theta_0$  and  $\Theta_1$  such that  $\mathcal{T}^*$  is most powerful with level  $\gamma$  for testing 5.8. Since  $\Theta_0 \in \Upsilon_\tau$  almost surely, we have:

$$\begin{aligned} \mathbb{P}[\mathcal{T}(\Theta_0 + X) = 1] &= \mathbb{P}[\mathcal{T}(\Theta_0 + X) = 1 \mid \Theta_0 \in \Upsilon_\tau] \\ &\leq \sup_{\Theta \in \mathcal{M}(\Omega, \mathbb{R}^N): \mathbb{P}[\mathcal{M}(\Theta) \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \mathcal{M}(\Theta) \leq \tau] \\ &\leq \alpha_{\mathcal{T}} && \text{from 5.3.1} \\ &\leq \gamma && \text{because } \mathcal{T} \text{ has level } \gamma \text{ for RDT} \end{aligned}$$

Therefore  $\mathcal{T}$  has level  $\gamma$  for 5.8. Since  $\mathcal{T}^*$  is most powerful with level  $\gamma$  for testing 5.8, we have:

$$\mathbb{P}[\mathcal{T}^*(\Theta_1 + X) = 1] \geq \mathbb{P}[\mathcal{T}(\Theta_1 + X) = 1]$$

Since  $\mathcal{T}$  and  $\mathcal{T}^*$  both have constant conditional power function given  $\Theta \in \Upsilon_\rho$ , we have:

$$\begin{aligned} \forall \theta \in \Upsilon_\rho, \mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \Theta \in \Upsilon_\rho] &= \beta_{\mathcal{T}}(\theta) \\ \text{and } \forall \theta \in \Upsilon_\rho, \mathbb{P}[\mathcal{T}^*(\Theta + X) = 1 \mid \Theta \in \Upsilon_\rho] &= \beta_{\mathcal{T}^*}(\theta) \end{aligned}$$

From lemma 5.3.3, since  $\mathcal{T}$  and  $\mathcal{T}^*$  both have constant conditional power function given  $\Theta \in \Upsilon_\rho$ , they both have constant power function on  $\Upsilon_\rho$ . From lemma 5.3.2, since  $\Theta_1 \in \Upsilon_\rho$  almost surely, we also have:

$$\begin{aligned} \forall \theta \in \Upsilon_\rho, \mathbb{P}[\mathcal{T}(\Theta_1 + X) = 1] &= \beta_{\mathcal{T}}(\theta) \\ \text{and } \forall \theta \in \Upsilon_\rho, \mathbb{P}[\mathcal{T}^*(\Theta_1 + X) = 1] &= \beta_{\mathcal{T}^*}(\theta) \end{aligned}$$

Hence:

$$\begin{aligned} \mathbb{P}[\mathcal{T}(\Theta_1 + X) = 1] &= \mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \Theta \in \Upsilon_\rho] \\ \text{and } \mathbb{P}[\mathcal{T}^*(\Theta_1 + X) = 1] &= \mathbb{P}[\mathcal{T}^*(\Theta + X) = 1 \mid \Theta \in \Upsilon_\rho] \end{aligned}$$

Therefore:

$$\mathbb{P}[\mathcal{T}^*(\Theta + X) = 1 \mid \Theta \in \Upsilon_\rho] \geq \mathbb{P}[\mathcal{T}(\Theta + X) = 1 \mid \Theta \in \Upsilon_\rho]$$

This proves that  $\mathcal{T}^*$  is  $\gamma$ -MCCP.



## Bibliography

- José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.
- O. Besson. Detection of a signal in linear subspace with bounded mismatch. *IEEE Transactions on Aerospace and Electronic Systems*, 42(3), 2006.
- P. Billingsley. *Probability and Measure, 3rd edition*. Wiley, 1995.
- A. A. Borovkov. *Mathematical statistics*. Gordon and Breach Science Publishers, 1998.
- Bradley P Carlin and Thomas A Louis. *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC, 2010.
- Tapas K Chandra and SN Joshi. Comparison of the likelihood ratio, rao's and wald's tests and a conjecture of cr rao. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 226–246, 1983.
- T.K. Chandra and R.M. Sankhya. Comparison of the Likelihood Ratio, Wald's and Rao's tests. *The Indian Journal of Statistics, Series A*, 47(2):271 – 284, June 1985.
- M. N. Desai and R. S. Mangoubi. Robust gaussian and non-gaussian matched subspace detection. *IEEE Transactions on Signal Processing*, 51(12):3115–3127, 2003.
- M. L. Eaton. *Multivariate statistics. A vector space approach*. Wiley, 1983.
- Morris L Eaton. Group invariance applications in statistics. In *Regional conference series in Probability and Statistics*, pages i–133. JSTOR, 1989.
- Joseph R Gabriel and Steven M Kay. Use of wijsman's theorem for the ratio of maximal invariant densities in signal detection applications. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, volume 1, pages 756–762. IEEE, 2002.
- Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- P. J. Huber. *Robust statistics*. Wiley, 1981.
- N. L. Johnson, S. Kotz, and N Balakrishnan. *Continuous Univariate distributions, 2nd edition*, volume 2. John Wiley & Sons, Inc, 1995.
- T. Kariya and B. K. Sinha. *Robustness of statistical tests*. Academic Press, 1989.
- S. Karlin and H. Rubin. Distributions possessing a monotone likelihood ratio. *J. Amer. Statist. Assoc.*, 51:637 – 643, Feb. 1956.
- S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume II, Detection Theory, 14th printing*. Prentice Hall, 2009.
- Steven M Kay. *Fundamentals of signal processing: Detection theory*, 1998.

- Prashant Khanduri, Dominique Pastor, Vinod Sharma, and Pramod K Varshney. On random distortion testing based sequential non-parametric hypothesis testing. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 328–334. IEEE, 2018.
- Prashant Khanduri, Dominique Pastor, Vinod Sharma, and Pramod K Varshney. Sequential random distortion testing of non-stationary processes. *IEEE Transactions on Signal Processing*, 67(21):5450–5462, 2019.
- E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses, 3rd edition*. Springer, 2005a.
- E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer, 3rd edition, 2005b.
- E.L. Lehmann and H. Scheffé. Completeness, similar regions, and unbiased estimation: Part I. *The Indian Journal of Statistics (1933-1960)*, 10(4):305 – 340, Nov. 1950.
- Emmanuelle C Leroy, Flore Samaran, Julien Bonnel, and Jean-Yves Royer. Seasonal and diel vocalization patterns of antarctic blue whale (*balaenoptera musculus intermedia*) in the southern indian ocean: A multi-year and multi-site study. *PloS one*, 11(11):e0163587, 2016.
- B. C. Levy. *Principles of signal detection and parameter estimation*. Springer Science & Business Media, 2008.
- Jean-Michel Marin and Christian Robert. *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Science & Business Media, 2007.
- R. J. Muirhead. *Aspects of multivariate statistical theory*. Wiley, 1982.
- J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transaction of the Royal Society of London. Series A.*, 231:289 – 337, 1933.
- J. Nocedal and S. J. Wright. *Numerical optimization*. Springer-Verlag, New York, 1999.
- D. Pastor and Q.-T. Nguyen. Random Distortion Testing and Optimality of Thresholding Tests. *IEEE Trans. Signal Process.*, 61(16):4161 – 4171, Aug. 2013a.
- D. Pastor, R. Gay, and A. Gronenboom. A sharp Upper Bound for the probability of error of Likelihood Ratio test for Detecting Signals in White Gaussian noise. *IEEE Trans. Inf. Theory*, 48(1):228 – 238, Jan. 2002.
- Dominique Pastor and Quang-Thang Nguyen. Random distortion testing and applications. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6347–6351. IEEE, 2013b.
- Dominique Pastor and Francois-Xavier Socheleau. Random distortion testing with linear measurements. *Signal Processing*, 145:116–126, 2018.
- S. Ramprasad, T. W Parks, and R. Shenoy. Signal modeling and detection using cone classes. *IEEE transactions on signal processing*, 44(2):329–338, 1996.

- L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.
- L. L. Scharf and B. Friedlander. Matched Subspace Detectors. *IEEE Trans. Signal Process.*, 42(8):2146 – 2157, 1994.
- F.-X. Socheleau and F. Samaran. Detection of Mysticete Calls: a Sparse Representation-Based Approach, Research report RR-2017-04-SC, Oct. 2017. <https://hal.archives-ouvertes.fr/hal-01736178/document>.
- F.-X. Socheleau, E. Leroy, A. Carvalho Pecci, F. Samaran, J. Bonnel, and J.-Y. Royer. Automated detection of antarctic blue whale calls. *J. Acoust. Soc. Am.*, 138(5): 3105–3117, 2015.
- Charles Stein. Some problems in multivariate analysis, part i. Technical report, STANFORD UNIV CA, 1956.
- Pramod K Varshney. *Distributed detection and data fusion*. Springer Science & Business Media, 2012.
- A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54 (3):426 – 482, Nov. 1943.
- Robert A Wijsman. Cross-sections of orbits and their application to densities of maximal invariants. In *Proc. Fifth Berkeley Symp. on Math. Stat. and Prob*, volume 1, pages 389–400, 1967.
- Ofer Zeitouni, Jacob Ziv, and Neri Merhav. When is the generalized likelihood ratio test optimal? *IEEE Transactions on Information Theory*, 38(5):1597–1602, 1992.





---

**Titre :** Décision binaire pour observations de distribution inconnue: Une approche optimale basée sur l'invariance.

**Mots clés :** Décision binaire, Invariance, Optimalité, Signaux aléatoires, Distribution inconnue.

**Résumé :** A travers ma thèse, nous nous sommes intéressés à la décision binaire quand le signal d'intérêt est supposé aléatoire de distribution inconnue. Dans la littérature classique, ce genre de scénario de détection n'offre pas beaucoup de possibilités de résolution qui garantissent une certaine optimalité. À notre connaissance, seule l'approche RDT traite de ce genre de problèmes en assurant une certaine optimalité. C'est ainsi que, durant ma thèse, nous avons tout d'abord étendu l'approche RDT dans le cadre d'une configuration distribuée, toujours en considérant le signal aléatoire de distribution inconnue. Ensuite, nous avons généralisé le RDT pour le cas d'un bruit non-Gaussien (GRDT). Enfin nous nous sommes orienté vers le cadre asymptotique dans l'espoir de dépasser les limites du GRDT. Pour ce faire, nous avons considéré un modèle simple de signaux déterministes et nous avons démontré une formulation asymptotique du théorème de Karlin-Rubin. Tous nos travaux se sont basés sur la notion clé d'invariance que nous avons redéfinie pour convenir à une classe de problèmes de décision plus générale, i.e. quand la distribution du signal d'intérêt n'est pas connue. Outre la notion d'invariance, l'optimalité a aussi eu une place importante dans les directions que nous avons considérées étant donné l'esprit de nos travaux.

---

**Title:** Binary decision for observations with unknown distribution: An optimal and invariance-based framework.

**Keywords :** Binary decision, Invariance, Optimality, Random signals, Unknown distributions.

**Abstract:** During my thesis, we took interest in decision problems where signals are assumed to be stochastic with unknown distributions. In standard literature, such an assumption does not allow to seek solutions that guarantees a certain optimality. At least, aside from the RDT framework developed a few years ago in our laboratory. Hence, we took interest in the philosophy behind the RDT framework, and we follow the same guidelines concerning the unknown distribution of the signal. Apart from our optimality purposes, we also have an invariance based perspective in how we intend to solve this type of decision problems. Indeed, when there are uncertainties about the signal of interest, we can try to derive solutions that are invariant towards them. These are the two key notions we consider throughout our investigations.

In this manuscript, first, we apply the RDT framework for a distributed decision to test its suitability to such decision scenarios where the signal of interest is random of unknown distribution and where the observations are collected by a network of sensors instead of just one sensor. Then, we generalise the theoretical material of the RDT framework to when the noise is not necessary Gaussian while still considering the signal of interest random of unknown distribution. Finally, we adopt an asymptotic outlook to circumvent the limitations of the RDT and the developed GRDT approach. Although the considered decision scenarios concern unconditional models in the simple case of deterministic signals, it allows to think ahead of the eventual upcoming generalisations in the asymptotic scope.