國 立 交 通 大 學

電子工程學系 電子研究所碩士班

碩 士 論 文

多重假設測試於訊號偵測之應用

Signal Detection Using A Multiple Hypothesis Test

研 究 生：林宗熹

指導教授：鍾佩蓉博士

林大衛博士

中 華 民 國 九 十 五 年 六 月

多重假設測試於訊號處理之應用

Signal Detection Using A Multiple Hypothesis Test

研究生: 林宗熹   Student: John-Si Lin

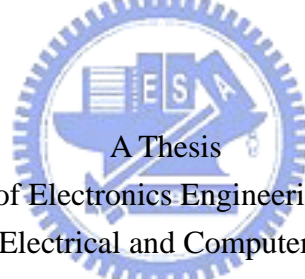指導教授: 鍾佩蓉博士   Advisor: Dr. P. J. Chung

林大衛博士   Dr. David W. Lin

國 立 交 通 大 學

電子工程學系　　電子研究所碩士班

碩士論文

A Thesis

Submitted to Department of Electronics Engineering & Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of Requirements

for the Degree of

Master of Science

in

Electronics Engineering

June 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年六月

多重假設測試於訊號偵測之應用

研究生：林宗熹　　　　　　　　指導教授： 鍾佩蓉 博士

　　　　　　　　　　　　　　　　　　　　　　林大衛 博士

國立交通大學

電子工程學系 電子研究所碩士班

摘要

　　在多樣化的應用裡，訊號源的檢測估計在理論上一直佔有很大的研究空間，隨著不同的應用，陣列感應器訊號處理已發展成一個活躍的研究領域。許多的技術已經發展出針對現實上不同問題的解決之道，例如雷達或利用聲納探測出訊號源的位置。MUSIC 與 ESPRIT 是早期用於偵測訊號來源方向的方法。1970 年代，AIC 與 MDL 準則是為了選擇模型規模而發展出得兩種不同方法。很多方法在過去幾年當中，已經發展出針對不同的情境而有不同的演算法，例如在同步訊號源或感應器數目非常少的情況下。本論文的目的即是藉由結合統計的處理方法，使新的演算法達到更好的成效。

　　本論文的方向始自於多重假設測試在訊號處理上的應用。Bonferroni 程序目的在於控制在即時考量下，同類比較測試時導致第一型錯誤的機率。如此一般的同類型錯誤率得出的結果是趨於保守的。隨著統計處理的發展，多重假設測試的錯誤率可以控制在某一個信心水準之下，尤其在某些特定的程序比較中可以控制的更低。Benjamini-Hochberg 程序展現出一種更加有效用的方法來達到更好的效果，而且保持錯誤率在一定的信心水準之下。本論文將在介紹這一個程序的流程與效用，並且應用於訊號偵測的領域，創新並檢驗是否能達到更好的效用。

Signal Detection Using A Multiple Hypothesis Test

Student: John-Si Lin

Advisor: Dr. P. J. Chung
Dr. David W. Lin

Department of Electronics Engineering & Institute of Electronics
College of Electrical and Computer Engineering
National Chiao Tung University

Abstract

Estimation problems in theoretical have been of great research interest in a great variety of applications. With application variation, sensor array signal processing developed as an active area of research. Many techniques have shown solutions for several real-world problems, such as source localization in radar and sonar. MUSIC and ESPRIT are parametric-based methods for detecting the locations of sources. AIC and MDL criteria are two kinds of approaches for model selection developed around 1970s. Many approaches in different situations, such as coherent signals or small number of sensors, were developed for the past years. Our goal is attempt to improve the performance by combining statistics processing.

Our idea comes from the multiple test which were tested in strong sense. The Bonferroni procedure controls the probability of committing any type I error in families of comparisons under simultaneous consideration. This control of familywise error rate results in conservative effects. With statistics development, the error rate of multiple hypotheses tests can be controlled under a certain level, especially lower in some of the procedure points. The Benjamini-Hochberg procedure shows a more powerful procedure to reach better performance and still maintain error rate under certain significance level. This work will introduce the controlling procedure and show the result with combing source detection problem.

誌謝

　　這篇論文能夠順利完成，最要感謝的人是我的指導教授 鍾佩蓉 博士。在這二年的研究生涯中，不論是學業上或生活上，處處感受到老師的用心，讓我能在遇到瓶頸時，有勇氣與毅力繼續前進。感謝另一位共同指導老師 林大衛 博士，除了豐富的學識和研究，老師親切、認真的待人處事態度，也是我景仰、學習的目標。

　　另外要感謝的，是楊華龍學長。感謝學長撥冗提供我了許多研究方面相關的疑問以及研究進行的方向。

　　感謝通訊電子與訊號處理實驗室(commlab)，提供了充足的軟硬體資源，讓我在研究中不虞匱乏。感謝實驗室的同學，平日一起唸書，一起討論，也一起分享喜悅，讓我的研究生涯充滿歡樂又有所成長。

　　感謝我所有的親愛的同學與朋友，蕙君、俊甫、元銘、耕毅、國和、士浤、紹茗、信豪、軒榮，在我口試前夕生病住院那幾天前來照顧與打氣，以及陪我度過二年的研究所生活。

　　最後，要感謝的是我的家人，他們的支持讓我能夠心無旁騖的從事研究工作。

　　謝謝所有幫助過我、陪我走過這一段歲月的師長、同儕與家人。謝謝！

<div align="right">

誌於 2006.6 風城交大

宗熹

</div>

Contents

List of Figures

Chapter 1

Introduction

Estimation problems in theoretical have been of great research interest in a great variety of applications. With application variation, sensor array signal processing developed as an active area of research. Many techniques have shown solutions for several real-world problems, such as source localization in radar and sonar. Subspace-Based methods, such as MUltiple SIgnal Classification (MUSIC), was indeed originally presented as a DOA estimator. Estimation of Signal Parameters Via Rotational Invariance Techniques (ESPRIT) is an algorithm that dramatically reduces theses computation and storage cost. It is more robust with respect to array imperfections than MUSIC; however, both of them are not the techniques we focus on in this work because they do not involve the concepts of multiple hypotheses test.

The parametric methods of signal processing require often not only the estimation of a vector of real-valued parameters but also the selection of one ore several integer-valued parameters that are equally important for the specification of a data model. For example, orders of an autoregressive moving average model, number of sinusoidal components in a sinusoids-in-noise signal, number of source signals impinging on a sensor array are theses kind of cases. The integer-valued parameters determine the dimension of the parameter vector of the data model, and thy must be estimated from the data. The task of selecting the above mentioned integer parameters is known as model order selection. Akaike Information Criterion (AIC)

and Minimum Description Length (MDL) are two kind of information Theoretic approaches developed around 1970s for such a problem. The approach MDL is consistent, whereas the AIC is not. The definition of consistency means that when the sample size increases to infinity, estimated probability goes to one. The advantage of information theoretic criterion is that it does not need any subjective judgment to detect the number of sources. It determined as the value for which the AIC or MDL criterion is minimized. However, the drawback of these methods is failure in the presence of coherent signals. Many approaches in different situations, such as coherent signals or small number of sensors, were developed for the past years. Our goal is attempt to improve the performance by combining statistics processing.

Our idea comes from the multiple test which were tested in strong sense. The Bonferroni procedure [1] controls the probability of committing any type I error in families of comparisons under simultaneous consideration. This control of familywise error rate (FWER, we will introduce in the following chapter) results in conservative effects. With statistics development, the error rate of multiple hypotheses tests can be controlled under a certain level, especially lower in some of the procedure points. The Benjamini-Hochberg procedure [2] shows a more powerful procedure to reach better performance and still maintain error rate under certain significance level. The following chapters will introduce the controlling procedure and show the result with combing source detection problem.

The organization of this thesis is:

- Chapter 2 introduce the background knowledge and the signal model we use.

- Chapter 3 analyzes the detection problem and algorithm of our approach.

- Chapter 4 will show the results of simulation in different scenarios.

- Chapter 5 gives conclusion and future work which could be done further.

- Appendix A briefly introduces the Neyman-Pearson theory and appendix B talks about the concept of p value used in statistics problem, which is a parameter for hypotheses test.

Chapter 2

Background and Signal Model

In this chapter some background that is used in analysis of signal model is introduced. In array signal processing, the signals propagated by waves from sources to the array are functions of position as well as time. They obey the laws of physics, especially the wave equation. For in-depth understanding and development of array signal processing algorithms, knowledge about spatiotemporal signal propagation and noise characteristics is required.

2.1 System and Propagating Waves

Let a three-dimensional Cartesian grid, such as $p(x, y, z)$, represent space. Considering time as the fourth dimension, a space-time signal can be written as $s(x, y, z, t)$, as shown in Fig. 2.1. A point is described by its distance r from the origin, azimuth θ within an equatorial plane containing the origin, and its angle ϕ down from the vertical axis. Simple trigonometric formulas connect spherical coordinates and Cartesian coordinates. The notation $\vec{x}$ denotes the position vector (x, y, z), where the x, y, and z are the spatial variables. We have
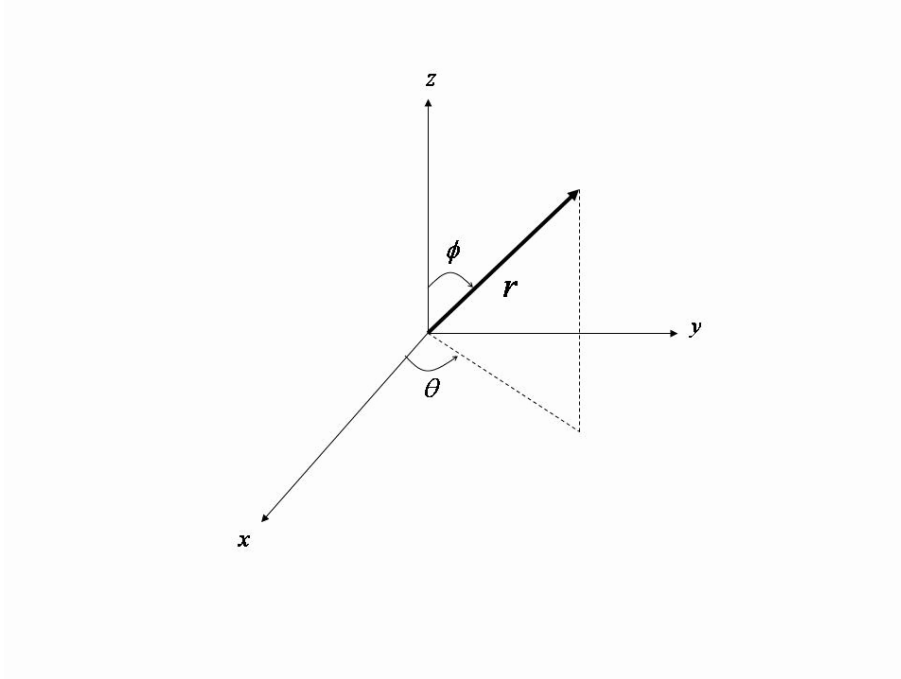
Figure 2.1: The spherical coordinate system, where θ is azimuth, ϕ elevation, and r distance from origin.
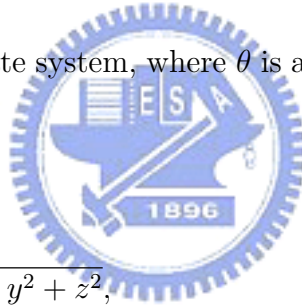
$$r = \sqrt{x^2 + y^2 + z^2},$$

$$x = r\sin\phi\cos\theta, \ y = r\sin\phi\sin\theta, \ z = r\cos\phi,$$

$$\theta = \cos^{-1}\left(\frac{x}{\sqrt{x^2 + y^2}}\right) = \sin^{-1}\left(\frac{y}{\sqrt{x^2 + y^2}}\right),$$

$$\phi = \cos^{-1}\left(\frac{z}{\sqrt{x^2 + y^2 + z^2}}\right).$$

The physics of propagation is described by wave equation for the appropriate medium and boundary conditions. Electromagnetic fields obey Maxwell's equations as follows:

$$\nabla \times \vec{E} = -\frac{\partial(\mu\vec{H})}{\partial t}, \ \nabla \cdot (\epsilon\vec{E}) = 0,$$

$$\nabla \times \vec{H} = -\frac{\partial(\mu\vec{E})}{\partial t}, \ \nabla \cdot (\epsilon\vec{H}) = 0,$$

where $\vec{E}$ is the electric field intensity, $\vec{H}$ is the magnetic field intensity, ϵ is the dielectric permittivity, μ is the magnetic permeability, and ∇ is gradient vector operator defined as

$$\nabla = \frac{\partial}{\partial x}\vec{i}_x + \frac{\partial}{\partial y}\vec{i}_y + \frac{\partial}{\partial z}\vec{i}_z.$$

From Maxwell's equations,

$$\nabla^2 \vec{E} = \frac{1}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2}$$

where ∇^2 denotes the Laplacian operator

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Putting the wave $s(\vec{x}, t)$ into the above equation, we obtain the wave equation as

$$\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} + \frac{\partial^2 s}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 s}{\partial t^2} \tag{2.1}$$

where $c = \dfrac{1}{\sqrt{\epsilon\mu}}$. The solution of the wave equation is a propagating wave. It determines the performance of array processing, which tries to extract useful information from the propagating waves. As a partial differential equation, it is often suggested to hypothesize a separable solution, which means a wave of the form

$$s(x, y, z, t) = f(x)g(y)h(z)p(t).$$

Assume a complex exponential form

$$s(x, y, z, t) = A\exp\{j(wt - k_x x - k_y y - k_z z)\} \tag{2.2}$$

where A is a complex constant and k_x, k_y, k_z, and w are nonnegative real constants. Substituting this form into the wave equation and cancelling the common terms, we have the constraint

$$k_x^2 + k_y^2 + k_z^2 = \frac{w^2}{c^2}. \tag{2.3}$$

Monochromatic, meaning "one color," plane wave thus arises from the solution of (2.2). Plane wave arises because at any instant of time t_0, under the constraint $k_x + k_y + k_z = C$, where C is constant, all points of $s(x, y, z, t_0)$ lie in the same plane. We may simply rewrite the expression as

$$s(\vec{x}, t) = A \exp\{j(wt - \vec{k} \cdot \vec{x})\} \tag{2.4}$$

where $\vec{k} \cdot \vec{x}$ is constant. As time proceeds by δt, planes of constant phase propagate by an amount of $\delta \vec{x}$. The equation $s(\vec{x} + \delta \vec{x}, t + \delta t) = s(\vec{x}, t)$ implies that

$$w\delta t - \vec{k} \cdot \vec{x} = 0. \tag{2.5}$$

As illustrated in Fig. 2.2, the reason why one chooses $\delta \vec{x}$ to have the same direction as $\vec{k}$ is for giving the minimum magnitude. *The direction of propagation $\vec{\zeta}^o$ of the plane wave corresponds to the direction of $\vec{k}$ and it is given by*

$$\vec{\zeta}^o = \frac{\vec{k}}{k}$$

where $k = |\vec{k}|$ is the magnitude of the wavenumber vector. We get $w\delta t = |\vec{k}||\delta \vec{x}|$ due to the simplification of $\vec{k} \cdot \delta \vec{x} = |\vec{k}||\delta \vec{x}|$ and hence $\dfrac{|\delta \vec{x}|}{\delta t} = \dfrac{w}{|\vec{k}|}$ denotes the speed of propagation of the plane wave. Because $|\vec{k}|^2 = w^2/c^2$, the speed of propagation is given by

$$\frac{|\delta \vec{x}|}{\delta t} = c.$$

The distance propagated during one temporal period, defined as $T = 2\pi/w$, is the *wavelength* λ. If δt in (2.5) is substituted by $2\pi/w$, then $\delta \vec{x} = \lambda = 2\pi/|\vec{k}|$, where vector $\vec{k}$ is wavenumber vector and its magnitude indicates the number of cycles (in radians) per meter. The quantity w is temporal frequency; therefore the wavenumber vector can be considered as *spatial frequency variables*, a three dimensional vector.

Rewriting (2.2) from the above simplification, we get

$$s(\vec{x}, t) = A \exp\{j(w(t - \vec{\alpha} \cdot \vec{x})\} \tag{2.6}$$
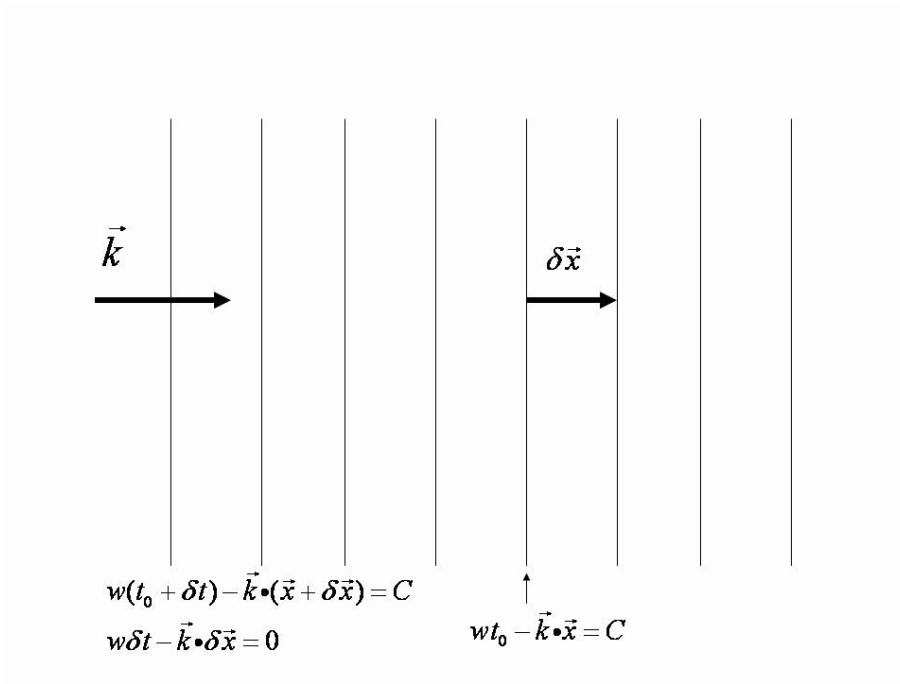
Figure 2.2: The vertical lines represent the plane waves. If one wants to see how much a plane wave moves during δt units, set $\delta \vec{x}$ in the same direction which $\vec{k}$ propagates towards and is known as direction of propagation.

where slowness vector, $\vec{\alpha}$, equals $\vec{k}/w$, the reciprocal of the propagation speed.

For a linear equation, the linear combination of several solutions is also a solution. If $s_1(\vec{x}, t)$ and $s_2(\vec{x}, t)$ are two solutions to the wave equation, then $as_1(\vec{x}, t) + bs_1(\vec{x}, t)$ is also a solution, where a, b are scalars. Sum up the monochromatic solutions gives the waveform

$$s(\vec{x}, t) = s(t - \vec{\alpha} \cdot \vec{x}) = \sum_{-\infty}^{\infty} S_n \exp\{jnw_0(t - \vec{\alpha} \cdot \vec{x})\}. \tag{2.7}$$

By the Fourier theorem, any periodic wave with period $T = 2\pi/w_0$ can be represented in the form. The coefficients S_n are given by

$$S_n = \frac{1}{T} \int_0^T s(u) \exp\{jnw_0 u\} du.$$

Equation (2.7) is a periodic wave propagating arbitrary waveform. It propagates with a speed $c = 1/|\vec{\alpha}|$ and has multiple frequencies $w = nw_0$. Since $s(\vec{x}, t)$ is a complex exponential

waveform, we can use Fourier theory to obtain an alternative expression in the frequency domain as

$$s(\vec{x}, t) = s(t - \vec{\alpha} \cdot \vec{x}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(w) \exp\{jw(t - \vec{\alpha} \cdot \vec{x})\} dw \qquad (2.8)$$

where from the Fourier transform we have

$$S(w) = \int_{-\infty}^{\infty} s(u) \exp\{-jwu\} du. \qquad (2.9)$$

Here we get the conclusion that any signal satisfies the wave equation, no matter what the shape of the signal is. This comes from the following reasons:

1. Arbitrary functions can be expressed as a weighted superposition of complex exponentials.

2. Complex exponentials solve the wave equation.

3. The wave equation is linear and the linearity implies that plane waves propagating in different directions can exist simultaneously.
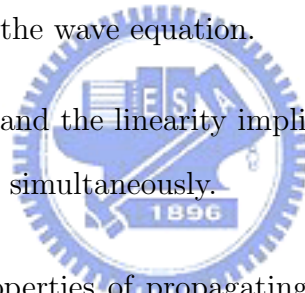
In summary, the following properties of propagating signals can be exploited by array processing algorithms:

- Propagating signals are functions of a single variable, $s(\cdot)$. Based on "local" information, one can infer the signal over all space and time, i.e., reconstruct the signal over space and time. Therefore, signals radiated by the sources can be determined.

- The speed of propagation depends on the physical parameters of the medium. If the characteristics of the medium are known, the speed and direction of propagation can be inferred.

- Signal propagating in a specific direction $\vec{\zeta}^o$ is equivalent to either $\vec{\alpha}$ or $\vec{k}$. At a given time instant, the spatial samples provide a set of values $s(t_0 - \vec{\alpha} \cdot \vec{x}_m)$ that can be used to determine $\vec{\alpha}$.

- The *superposition principle* allows several propagating waves to occur simultaneously without interaction.

2.2 Wavenumber-Frequency Space

Fourier transform, a useful tool for analyzing time series, can be used to estimate the frequency content of a signal or a weighted sum of sinusoidal functions. The forward Fourier transform and inverse Fourier transform have similar forms, thus duality exists between the time-domain and frequency-domain representation. When the Fourier analysis is extended to multidimensional signals, the properties still hold. Four-dimensional Fourier transform of a spatiotemporal signal $s(\vec{x}, t)$ can be written as

$$S(\vec{k}, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\vec{x}, t) \exp\{-j(wt - \vec{k} \cdot \vec{x})\} d\vec{x} d\vec{t}. \tag{2.10}$$

The above can be interpreted as an representation of signals in terms of the temporal frequency variable w and the wavenumber vector variable $\vec{k}$, which is considered a spatial frequency variable.

Consider a uniform plane wave

$$s(\vec{x}, t) = \exp\{j(w_0 t - k^o \cdot \vec{x})\}$$

where k^o and w_0 are particular wavenumber vector and temporal frequency. The sinusoidal wave, with period $T = 2\pi/w_0$, spatial frequency $k^o = |\vec{k}|$, wavelength $\lambda = 2\pi/k^o$, speed of propagation $c_0 = w_0/k^o$ and direction of propagation parallel to the wavenumber vector $\vec{k}$,

has its corresponding Fourier transform

$$S(\vec{k}, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{-j(w - w_0)t + (\vec{k} - k^o) \cdot \vec{x}\} d\vec{x} dt.$$

When our space-time signal is a propagating wave, $s(\vec{x}, t) = s(t - \vec{\alpha}^o, \vec{x})$, the Fourier transform is

$$S(\vec{k}, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(t - \vec{\alpha}^o), \vec{x}) \exp\{-j(wt - \vec{k} \cdot \vec{x})\} d\vec{x} dt.$$

Substituting $u = t - \vec{\alpha} \cdot \vec{x}$ into the above equation, we get

$$S(\vec{k}, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(u) \exp\{-j(wu - (\vec{k} - w\vec{\alpha}^o) \cdot \vec{x}\} d\vec{x} dt$$
$$= S(w)\delta(\vec{k} - w\vec{\alpha}).$$

2.3   Random Fields

2.3.1   Stationary Random Fields

Spatiotemporal signals can be described as functions of space and time. Noise plays an important role in determining the characteristics of the propagating signals. Mathematical modeling of noise requires use of the theory of probability and stochastic processes. Spatiotemporal stochastic processes are known as *random fields*. Let $f(\vec{x}, t)$ denote a random field. For particular values of $\vec{x} = \vec{x}^o$ and $t = t_0$, $f(\vec{x}_0, t_0)$ denotes a random variable that is characterized by a probability density function $p_{f(\vec{x}_0, t_0)}(\cdot)$. The following gives the definitions of expected value, correlation function, and covariance function of a random field:

- Expected value: $\varepsilon[f(\vec{x}_0, t_0)] = \int \alpha p_{f_0}(\alpha) d\alpha$.

- Correlation function: $R_f(\vec{x_0}, \vec{x_1}, t_0, t_1) = \varepsilon[f_0 f_1] = \int \int \alpha \beta p_{f_0, f_1}(\alpha, \beta) d\alpha d\beta$.

- Covariance function: $K_f(\vec{x_0}, \vec{x_1}, t_0, t_1) = R_f(\vec{x_0}, \vec{x_1}, t_0, t_1) - \varepsilon[f_0]\varepsilon[f_1]$.

Stationary random fields possess an important and special property for the above and other statistics: The statistics do not vary with *absolute* time and position. The joint probability functions of random variables depend only on the *difference in position and time.* Thus the correlation function varies with the difference $\vec{x}_1 - \vec{x}_0$ and $t_1 - t_0$, denoted $\vec{\chi}$ and τ below, respectively. The correlation function of a stationary random field can be written as

$$R_f(\vec{\chi}, \tau) = \varepsilon[f(\vec{x}_0, t_0)f(\vec{x}_0 + \vec{\chi}, t_0 + \tau)].\tag{2.11}$$

When the stationary random field $f(\vec{x}, t)$ is filtered by a linear, space-invariant and time-invariant filter, we get another stationary random field $y(\vec{x}, t)$ at the output, see Fig. 2.3. The filter is characterized by its impulse response $h(\vec{x}, t)$ or its wavenumber-frequency response $H(\vec{k}, w)$. The autocorrelation of the output is give by

$$R_y(\vec{\chi}, \tau) = \int_{-\infty}^{\infty} R_{hh}(\vec{a}, b) R_f(\vec{\chi} - \vec{a}, \vec{\tau} - b) d\vec{a} db\tag{2.12}$$

where $R_{hh}(\vec{a}, b)$ represents the deterministic autocorrelation function of the impulse response $h(\vec{x}, t)$ as

$$R_{hh}(\vec{a}, b) = \int_{-\infty}^{\infty} h(\vec{x}, t) h(\vec{x} + \vec{a}, t + b) d\vec{x} dt.\tag{2.13}$$

From (2.11) it can be observed that the correlation function of output random field is equal to the convolution between the deterministic autocorrelation function of the filter impulse response and the input correlation function.

The power spectral density function of a random field filtered in space and time can be shown to be

$$S_y(\vec{k}, w) = |H(\vec{k}, w)|^2 S_f(\vec{k}, w).\tag{2.14}$$

We see that the input power spectral density is multiplied by the squared magnitude of the filter's frequency response to obtain the power spectral density function of the (stationary) output random process.

$f(\vec{x},t) \longrightarrow \boxed{H(\vec{k},w)} \longrightarrow y(\vec{x},t)$
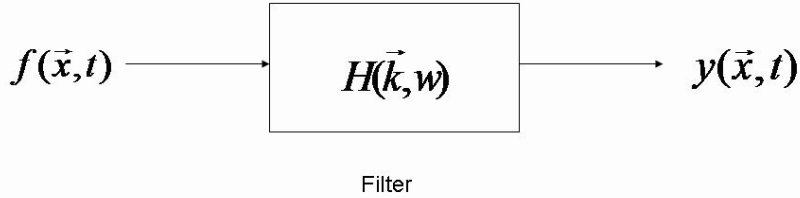
Filter

Figure 2.3: Filtered stationary random field.

2.3.2   Correlation Matrices of Sensor Outputs

The correlation functions computed from the sensor outputs play a crucial role in array processing algorithms, and the corresponding *correlation matrices* are worthy of some discussion. Correlation matrices are classified into two classes: *spatiotemporal* correlation matrices and *spatiospectral* correlation matrices. Below we introduce the former. For more information, see [3].

For convenience in mathematical treatment, is is common to use a matrix form where the sensors are statically positioned, meaning that spatial variables $\vec{x}_1$ and $\vec{x}_2$ are evaluated at the sensor locations $\{\vec{x}_m\}$, $m = 0, ..., M - 1$. Also we need to sample simultaneously at the output of each sensor for $\{t_n\}$, $n = 0, ..., N - 1$. The correlation matrix $\mathbf{R}(t_0)$ is arranged from these MN values by blocking the matrix into M^2 submatrices, each of which

13

being a cross-correlation matrix of the temporal values between pairs of sensor outputs. The $MN \times MN$ *spatiotemporal correlation matrix* is

$$\mathbf{R}(t_0) = \begin{bmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \cdots & \mathbf{R}_{0,M-1} \\ \mathbf{R}_{1,0} & \mathbf{R}_{1,1} & \cdots & \mathbf{R}_{1,M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{M-1,0} & \cdots & \cdots & \mathbf{R}_{M-1,M-1} \end{bmatrix}$$

where $\mathbf{R}_{m_1,m_2}$ represents the $N \times N$ cross-correlation matrix of the field collected by the sensors m_1 and m_2. We have

$$\mathbf{R}_{m_1,m_2} = R_f(\vec{x}_{m_1}, \vec{x}_{m_2}, t_{n_1}, t_{n_2}), \quad n_1, n_2 = 0, ..., N-1.$$

note that $\mathbf{R}(t_0)$ is a Hermitian matrix because the off-diagonal terms are symmetric, i.e., $\mathbf{R}_{m_1,m_2} = \mathbf{R}'_{m_2,m_1}$, and each main-diagonal term $\mathbf{R}_{m,m}$ is equal to the temporal correlation matrix of the observations measured at the mth sensor. When the field is stationary, each submatrix is *Toeplitz*, i.e., a matrix with constant values along each diagonal in the form

$$\begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_{-1} & \cdots & \mathbf{R}_{-(M-1)} \\ \mathbf{R}_1 & \mathbf{R}_0 & \mathbf{R}_{-1} & \mathbf{R}_{-(M-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{M-1,0} & \mathbf{R}_{M-2} & \cdots & \mathbf{R}_0 \end{bmatrix}.$$

A special case occurs when the spatial and temporal correlations are separable. The spatiotemporal correlation can be written as $R_f(\vec{\chi}, \tau) = R_f^{(s)}(\vec{\chi}) \cdot R_f^{(t)}(\tau)$ where the two terms on the right-hand-side are the spatial correlation function and the temporal correlation functions, respectively. In terms of the spatiotemporal covariance matrix, it implies that $\mathbf{R}_{m_1,m_2} = \rho_{m_1,m_2} \mathbf{R}_{0,0}$, where ρ_{m_1,m_2} denotes the spatial correlation coefficient between sensors m_1 and m_2.

If the sampled field is either spatially or temporally white, the correlation matrix's structure is simplified greatly. For the case of a spatially white field, the correlation function becomes $\delta_{m_2-m_1} R_f(\vec{0}, t_{n_2} - t_{n_1})$ and $\rho = \mathbf{I}$, yielding the spatiotemporal covariance matrix

14

$$
\begin{bmatrix}
\mathbf{R}_{0,0} & 0 & \cdots & 0 \\
0 & \mathbf{R}_{0,0} & 0 & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \mathbf{R}_{0,0}
\end{bmatrix}
$$

When the field is temporally white, the spatiotemporal correlation function simplifies to $\delta_{n_2-n_1} R_f(\vec{x}_{m_2} - \vec{x}_{m_1}, 0)$. Matrices $\mathbf{R}_{m_1,m_2}$ are diagonal, equaling $\rho_{m_1,m_2}\sigma^2\mathbf{I}$, so that

$$
\mathbf{R} =
\begin{bmatrix}
\sigma^2\mathbf{I} & \rho_{0,1}\sigma^2\mathbf{I} & \cdots & \rho_{0,M-1}\sigma^2\mathbf{I} \\
\rho_{1,0}\sigma^2\mathbf{I} & \sigma^2\mathbf{I} & \rho_{1,2}\sigma^2\mathbf{I} & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
\rho_{M-1,0}\sigma^2\mathbf{I} & \cdots & \cdots & \sigma^2\mathbf{I}
\end{bmatrix}.
$$

2.4 Assumptions on Signal and Noise

Let the sensor outputs be denoted $\{f(\vec{x}_m), t\}_{m=0}^{M-1}$, which contains a superposition of signals and noise. The noise is usually considered as stationary, zero-mean Gaussian, and statistically independent of the field's signals. The stationarity is assumed in both the temporal and the spatial domains and it means that the observation noise collected at each sensor has identical statistical characteristics and any cross-correlation between the noises at different sensor outputs depends only on the physical distance separating the sensors. The above assumptions implies the *isotropic noise* condition: The noise component of the observations consist of waveforms propagating toward the array from all directions in the field. Noise comes from everywhere and has well-defined power spectrum. Some signals may be jammers, which are unwanted signals transmitting toward the array and resulting in interference on the detecting process for the signals we are interested in. The source locations of jammers may be known before processing. Signals can be realizations of a stochastic process or deterministic. They may convey information and our goal is trying to exploit the information hidden in the signal structure with effective array processing algorithms.

The sampled signal obtained from the mth sensor is a linear combination of signals and

15

stationary noise that can be written as

$$y_m(t_n) = \sum_{i=0}^{N_s-1} s_i(m, t_n) + n_m(t_n), \quad n = 0, ..., N-1, \qquad (2.15)$$

where s_i represents a source that propagates energy. Each signal depends on sensor index m via a propagation-induced delay-per-sensor $\Delta_i(m)$: $s_i(m, t_n) = s_i(t_n - \Delta_i(m))$. It is equal to the scaled dot product between the source's apparent direction vector and the sensor's location as

$$\Delta_i(m) = \frac{\vec{\zeta}_i^o \vec{x}_m}{c}.$$

The propagation speed c varies frequently because of inhomogeneities in the medium. This variation makes our problem complex and complicated to determine the direction of propagation. For given sensor locations, an algorithm can only estimate the ratio of the signal's direction vector and the speed of propagation. Given either the direction or the speed of propagation, we can get the other term without ambiguity.

As mentioned, the signals can be stochastic or deterministic. The matrices of interest consist of all possible cross-correlations between the various signals and their observed values at each sensor. Define an $MN \times N_s$ signal matrix $\mathbf{S}(t_0)$ as

$$\mathbf{S}(t_0) = \begin{bmatrix} \mathbf{s}_0(0) & \mathbf{s}_1(0) & \cdots & \mathbf{s}_{N_s-1}(0) \\ \mathbf{s}_0(1) & \mathbf{s}_1(1) & \cdots & \mathbf{s}_{N_s-1}(1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_0(M-1) & \mathbf{s}_1(M-1) & \cdots & \mathbf{s}_{N_s-1}(M-1) \end{bmatrix}$$

where each column contains a sequence of observations for a specific signal. Each term of the matrix, $\mathbf{s}$, is equal to $col[s-i(m, t_0), ..., s_i(m, t_{N-1})]$, i.e., the output sequence vectors recorded at the mth sensor. The observation vector is written as $\mathbf{y}(n_0 T) = \mathbf{S}(n_0 T)\mathbf{1}$, where $\mathbf{1} = col[1, ..., 1]$. The spatiotemporal correlation matrix $\mathbf{R}(t_0)$ for a deterministic signal is

given by

$$
\begin{aligned}
\mathbf{R}(t_0) &= E[\mathbf{y}(t_0)\mathbf{y}'(t_1)] \\
&= \mathbf{S}(t_0)\mathbf{1}\mathbf{1}'\mathbf{S}'(t_0) + \mathbf{R}_n \quad\quad (2.16)
\end{aligned}
$$

where $\mathbf{y}(t_0) = \mathbf{S}(t_0)\mathbf{1} + \mathbf{n}(t_0)$ and $\mathbf{R}_n$ denotes the correlation matrix of the noise.

Evaluation of the spatiospectral correlation matrix reveals a strikingly similar mathematical form. We define the $M \times N_s$ signal matrix $\mathbf{S}$ as

$$
\mathbf{S} = \begin{bmatrix} e^{-jw\Delta_0(0)} & \cdots & e^{-jw\Delta_{N_s-1}(0)} \\ \vdots & \cdots & \vdots \\ e^{-jw\Delta_0(M-1)} & \cdots & e^{-jw\Delta_{N_s-1}(M-1)} \end{bmatrix}.
$$

Each column of this matrix has the form

$$
\mathbf{e}(\Delta_i) = col[e^{-jw\Delta_i(0)}, ..., e^{-jw\Delta_i(M-1)}]
$$

which represents the propagation-induced phase shift experienced by a unit-amplitude complex exponential signal and measured by the sensors. Define the $N_s \times N_s$ matrix $\mathbf{C}$ as *intersignal coherence matrix* with $c_{i_1,i_2} = \varepsilon[S_{i_1}(\omega)S_{i_1}^*(\omega)]$. The spatiospectral matrix becomes

$$
R(\omega) = \mathbf{S}\mathbf{C}\mathbf{S}' + \mathbf{R}_N.
$$

The similarity between spatiotemporal and spatiospectral correlation matrices allows us to ignore whether a time-domain or frequency-domain algorithm is in question. If the signal is narrow-band, the matrix's smaller size dramatically reduces the amount of computation; if wideband, matrices are required to integrate fully the signal components in the observations and the amount of computation increases proportionally. In short, the signal's characteristics affect the *algorithm type* and the resulting complexity.

Problem Formulation

Consider M sensors composed in an array with arbitrary locations and directions. Assume there are q narrow-band sources centered around a known frequency w_0 coming from different locations $\theta_1, \cdots, \theta_q$. Use the complex envelopes, $p \times 1$ vector, to represent the received signals as

$$x(t) = \sum_{i=1}^{q} \mathbf{a}(\theta_i) s_i(t) + \mathbf{n}(t), \qquad (2.17)$$

where $s_i(t)$ represents the signal of the ith source received at the reference point and $n(t)$ represents the noise at the ith sensor. $\mathbf{a}(\theta)$ is the *steering vector* of the array toward direction θ,

$$\mathbf{a}(\theta) = [a_1(\theta)e^{-jw_0\tau_1(\theta)}, \cdots, a_M(\theta)e^{-jw_0\tau_M(\theta)}]. \qquad (2.18)$$

In the (2.18), $a_i(\theta)$ is the amplitude response of the ith sensor towards direction θ. $\tau_i(\theta)$ represents the propagation delay beteen the reference point and the ith sensor to a wavefront impinging from direction θ. For simplicity, we use matrix notation to replace (2.17) as

$$\mathbf{x}(t) = \mathbf{A}(\theta^{(q)})\mathbf{s}(t) + \mathbf{n}(t), \qquad (2.19)$$

where

$$\mathbf{A}(\theta^{(q)}) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_q)] \qquad (2.20)$$

is the $M \times q$ matrix of the steering vectors. We will discuss more in chapter 3.

18

Chapter 3

Signal Detection Using a Multiple Hypothesis Test

The number of signals is an important factor dominating the performance of many algorithms of signal detection methods. Here we investigate an approach based on *multiple hypothesis test*.

3.1 Eigenanalysis Methods

3.1.1 Eigenanalysis

Eigenanalysis is the determination of the natural coordinate system for a given quantity. It is used frequently in the development of signal processing algorithms. A linear operator's spectrum is defined to be its set of eigenvalues and its "natural" basis is the set of normalized eigenvectors. Let $\mathcal{L}$ be a linear operator, x denote a vector and a represent a scalar: $\mathcal{L}[a_1 x + a_2 x] = a_1 \mathcal{L}[x_1] + a_2 \mathcal{L}[x_2]$. This operation can represent signal processing operations such as linear filtering and the multiplication of a vector by a matrix. One way of viewing the eigensystem is that the eigenvalue corresponding to an eigenvector gives the operator's gain when it acts on the eigenvector. Let the order of indexes from the smallest to the largest, $\lambda_1, \lambda_2, ..., \lambda_M$, where M is the number of eigenvectors, corresponds to the order of

the magnitude of eigenvalues from the largest to the smallest. One common application of a linear operator is to describe a linear, time-invariant system. The *eigensystem* $(\lambda_i, v_i)_{\mathcal{L}}$ refers to the collection of eigenvectors and associated eigenvalues.

The utility of eigenanalysis rests not only on defining the input-reproducing property, but also on the orthogonality of the eigenvectors and their completeness. A vector x can be represented by a linear eigenvector expansion which defines a coordinate system:

$$x = \sum_i a_i v_i \text{ with } a_i = < x, v_i > .$$

Applying the linear operator to defining the system allows the associated linear operator's output to have an easily calculated representation as

$$\mathcal{L}[x] = \sum_i \lambda_i a_i v_i .$$

Eigenvectors of a linear operator do not guarantee to form an orthogonal set for the reason depending on whether the Fourier transform exists, but this linear operators do.

3.1.2　Signal and Noise Subspaces

For many adaptive array processing algorithms, the spatial correlation matrix $\mathbf{R}$ is the crucial feature. Let us begin with the formulation and definition of signal and noise subspaces.

The observation vector $\mathbf{x}(t)$ is the $M \times 1$ vector given by

$$\mathbf{x}(t) = \sum_{i=1}^{q} \mathbf{A}(\Phi_i) s_i(t) + \mathbf{n}(t) \tag{3.1}$$

where $s_i(\cdot)$ is scalar complex waveform referred to as the ith signal, $\mathbf{A}(\Phi)$ denotes an $M \times 1$ complex vector, parameterized by an unknown parameter vector Φ associated with the ith signal, and $\mathbf{n}(\cdot)$ represents an $M \times 1$ complex vector referred to as the additive noise. M is known as the number of sensors. Here we assumed the q signals are complex, stationary, and

ergodic Gaussian random processes with zero mean and positive definite covariance matrices. The noise vector is assumed to be a complex stationary and ergodic Gaussian vector process, independent of the signals, with zero mean and covariance matrix given by $\sigma^2\mathbf{I}$, where σ^2 is an unknown scalar constant and $\mathbf{I}$ is the identity matrix. The problem is how can we estimate the number of signals from a finite set of observation $x(t_1),...,x(t_N)$. Rewrite (3.1) as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{3.2}$$

where $\mathbf{A}$ is the $p \times q$ matrix

$$\mathbf{A} = [\mathbf{A}(\Phi_1) \cdots \mathbf{A}(\Phi_q)] \tag{3.3}$$

and s(t) is the $q \times 1$ vector

$$\mathbf{s}(t) = [s_1(t) \cdots s_q(t)]^T. \tag{3.4}$$

From the previous assumptions that the noise is zero mean, additive, white, Gaussian and independent of the signals, use of eigenanalysis to develop algorithms hinges on the relation between the eigenvectors of spatial correlation matrices and signal vectors. The covariance matrix of $X(t)$ is given by

$$\mathbf{R} = \mathbf{A}\mathbf{C_S}\mathbf{A}' + \sigma^2\mathbf{I} \tag{3.5}$$

where $\mathbf{A}'$ is the conjugate transpose of $\mathbf{A}$ and $\mathbf{C_S} = E[\mathbf{s}(\cdot)\mathbf{s}(\cdot)']$, the covariance matrix of signals. Assume $q < M$ and that both $\mathbf{C_S}$ and $\mathbf{A}$ are of full rank q. The positivity of $\mathbf{R}$ guarantees the following representation

$$\begin{aligned} \mathbf{R} &= \mathbf{A}\mathbf{C_S}\mathbf{A}' + \sigma^2\mathbf{I} \\ &= \sum_{i=1}^{M} \lambda_i v_i v_i' \\ &= \mathbf{U}\Lambda\mathbf{U}' \end{aligned} \tag{3.6}$$

where $\mathbf{U} = [\mathbf{v}_1,...\mathbf{v}_M]$ and $\Lambda = diag[\lambda_1,...,\lambda_M]$ with $\lambda_1 > \cdots > \lambda_M$. Consider the case of $\mathbf{A}\mathbf{C_S}\mathbf{A}'$ being singular, i.e., $rank(\mathbf{A}\mathbf{C_S}\mathbf{A}') < M$. Then the determinant of $\mathbf{A}\mathbf{C_S}\mathbf{A}'$ is

21

equal to zero. Therefore σ^2 is equal to at least one of the eigenvalues of $\mathbf{R}$. But since $\mathbf{A}$ is full-rank and $\mathbf{C_S} > 0$, $\mathbf{AC_S A'}$ must be nonnegative definite. Thus σ^2 can only be the minimum eigenvalue λ_{min}. Since $rank(\mathbf{AC_S A'}) = q$, its smallest eigenvalues are zeros with multiplicity $(M - q)$. In addition, the eigenvalues of $\mathbf{R}$ and those of $\mathbf{AC_S A'} = \mathbf{R} - \sigma^2 \mathbf{I}$ differ from σ in all cases.

Let $\{\mathbf{v}_{q+1}, ..., \mathbf{v}_M\}$ denote the set of eigenvectors associated with $\lambda_{min} = \sigma^2$. Since $\mathbf{R}v_i = (\mathbf{AC_S A'} + \sigma^2 \mathbf{I})v_i = \lambda_i v_i$, we have $\mathbf{AC_S A'}v_i = (\lambda_i - \sigma^2)v_i$. For $\lambda_i = \sigma^2$, $\mathbf{AC_S A'}v_i = 0$ or $\mathbf{A'}v_i = 0$. That is, the noise eigenspace is orthogonal to the column space of $\mathbf{A}$.

Since the subspace spanned by $\{\mathbf{v}_{q+1}, ..., \mathbf{v}_M\}$ is orthogonal to the column space of $\mathbf{A}$, $\{\mathbf{v}_1, ..., \mathbf{v}_q\}$ must span the column space of $\mathbf{A}$. Thus we can partition the eigenvector/value pairs into signal eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_q\}$ corresponding to $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \sigma^2$ and noise eigenvectors $\{\mathbf{v}_{q+1}, ..., \mathbf{v}_M\}$ corresponding to $\lambda_{q+1} = \cdots = \lambda_M = \sigma^2$. Let

$$\mathbf{R} = \mathbf{U_s \Lambda_s U_s'} + \mathbf{U_n \Lambda_n U_n'}, \tag{3.7}$$
$$\Lambda_\mathbf{s} = diag[\lambda_1, ..., \lambda_q],$$
$$\Lambda_\mathbf{n} = diag[\lambda_{q+1}, ..., \lambda_M].$$

The number of signals q can be determined from multiplicity of the smallest eigenvalue of $\mathbf{R}$. The problem is that the covariance matrix $\mathbf{R}$ is unknown in practice. Because of the finite sample size, the *observed* eigenvalues are all different with probability one and for this reason, it is difficult to determine the number of signals merely by "observing" the eigenvalues.

3.2 Multiple Hypothesis Test

The problems of interest in array processing are not only bearing estimation, but also the number of signals. In many algorithms, the number of signals is known so that the resolution

of bearing estimation improves greatly; in other words, if we have a credible method of determining the number of sources, we can correctly adopt high resolution algorithms. This is the so-called *source detection problem.*

The algorithms used to detect the number of signals incident on the array, such as the AIC[4], MDL[5], and sphericity test, exploit the form of the spatial correlation matrix $\mathbf{R}$. AIC and MDL are both information theoretic criteria for a model selection problem proposed by Wax and Kailath [6], and the sphericity test is a sequence of hypothesis tests on adjacent groups of eigenvalues of $\mathbf{R}$ to determine how many of the smallest eigenvalues are equivalent.

For M sensors, $\mathbf{R}$ is the $M \times M$ spatial correlation matrix where the obtained vectors are selected from the Fourier transforms of the sensor outputs at particular frequencies of the narrow-band signals. We have discussed the signal and noise subspace in Section 3.1.2. Now we simplify the notation in (3.5) for convenience as

$$
\begin{aligned}
\mathbf{R} &= \mathbf{A}\mathbf{C_S}\mathbf{A}' + \sigma_n^2\mathbf{I} \\
&= \mathbf{A}\mathbf{C_S}\mathbf{A}' + \mathbf{K}_n
\end{aligned}
\tag{3.8}
$$

where $\mathbf{A}$ is the matrix constructed by columns corresponding to the vector form of signals, $\mathbf{A}'$ is the conjugate transpose of $\mathbf{A}$, $\mathbf{C_S}$ is the matrix of the correlation between the signals, and $\mathbf{K}_n$ represents the correlation matrix of the noise. Consider a *ULA* (uniform linear array) where the sensors are spaced apart with constant distance, such as half-wavelength, in a linear way. The columns of $\mathbf{A}$, also called *steering vector* ,are of the Vandermonde form

$$
\mathbf{a}_i^T = [1 \ e^{\gamma_i} \ e^{2\gamma_i} \cdots e^{(M-1)\gamma_i}]
\tag{3.9}
$$

where $\gamma_i = j2\pi(d/\lambda)\sin\theta_i$, with λ denoting the wavelength of the narrow-band signals, θ_i denoting the *DOA* (degree of arrival), i.e., the direction of ith signal, and d representing the spacing between sensors. For the AWGN scenario, we have $K_n = \sigma_n^2\mathbf{I}$ where σ_n^2 is the noise power.

Conceptually, the correlation matrix is shaped from two parts: signal and non-signal (noise) subspaces. If there are fewer sources than sensors, then the rank of $\mathbf{AC_S A}'$ is equal to the number of sources. For the case where there are q sources, $\mathbf{AC_S A}'$ is of rank q with q eigenvalues $\xi_1, \xi_2, \cdots, \xi_q$. On the other hand, the $\sigma_n^2 \mathbf{I}$ contains $M - q$ diagonal terms equal to σ_n^2. Thus the q largest eigenvalues $\xi_1 + \sigma_n^2, \xi_2 + \sigma_n^2, \cdots, \xi_q + \sigma_n^2$ represent the eigenvalues of the signal subspace; the remaining $(M - q)$ eigenvalues of the correlation matrix σ_n^2 correspond to the noise subspace part. For the fewer sources case, i.e., $q < M$, the smallest $M - q$ eigenvalues of $\mathbf{R}$ are all equal. Since we can determine the order of the noise, we can get the number of signals. Finally if there are more sources than sensors, we can only detect at most $M - 1$ sources since the algorithm assumes that the smallest eigenvalue lies in the noise subspace with different value from the others; in other words, the algorithm can only detect $M - 1$ sources.

3.2.1 Multiple Hypothesis

The motivation of such an experiment comes from the following problem. The following paragraph is just one case to exemplify the procedure when we make assumption of hypotheses test. Here are two system, S_0 and S_1, generating different measurements $\mathbf{X}$ distributed as $N[\mathbf{m}_1, \mathbf{R}_1]$ and $N[\mathbf{m}_1, \mathbf{R}_1]$, respectively, where $N[\mathbf{m}_i, \mathbf{R}_i]$ indicates that the measurement is normal distributed with mean $\mathbf{m}_i$ and correlation matrix $\mathbf{R}_i$. Therefore, we have two models, or hypotheses, to test which one is more reliable and consistent. The models are:

$$H_0 : \mathbf{X} : N[\mathbf{m}_0, \mathbf{R}_0],$$

$$H_1 : \mathbf{X} : N[\mathbf{m}_1, \mathbf{R}_1].$$

Generally, we choose the conservative term as null hypothesis, H_0, and the other as the alternative, $H_i, i \neq 0$. In other fields such as biomedical engineering, pharmaceutical test, or neuroimaging analysis, we have strong faith on those conventional methods or experienced

techniques bringing benefit as time goes by. The problem is that which one should we select to minimize the probability of erroneous decisions. The Neyman-Pearson detector [7] provides a scheme of decision according to the likelihood ratio.

Classes of Tests

Let $\mathbf{X}$ denote a random vector with distribution $F_\theta(\mathbf{x})$, where the $\boldsymbol{\theta}$ belongs to the parameter space Θ which is a disjoint covering of the parameter space, $\Theta = \Theta_i \cup \Theta_j$. Take the binary hypothesis case for example. H_0 denotes the hypothesis that $\boldsymbol{\theta} \in \Theta_0$ and H_1 represents $\boldsymbol{\theta} \in \Theta_1$. The test

$$H_0\text{: } \boldsymbol{\theta} \in \Theta_0 \text{ versus } H_1\text{: } \boldsymbol{\theta} \in \Theta_1$$

thus is called *binary test of hypotheses*. Moreover, if $\Theta = \Theta_0 \cup \Theta_1 \cup \cdots \cup \Theta_{M-1}$ is the parameter space, the test

$$H_0 \text{ versus } H_1 \text{ versus } \cdots \text{ } H_{M-1}$$

is called *multiple* or *M-ary hypothesis test*. Hypotheses also are classified into two kinds: *simple* hypotheses and *composite* hypotheses. If a single vector $\boldsymbol{\theta}$ belongs to Θ_i, the hypothesis H_i is termed simple; otherwise, it is composite. For the binary hypothesis case, if H_0 is simple and H_1 is composite, this is a test about a simple hypothesis versus a composite alternative. A typical example is

$$H_0\text{: } \boldsymbol{\theta} = \mathbf{0} \text{ versus } H_1\text{: } \boldsymbol{\theta} \neq \mathbf{0}$$

where we judge whether $\boldsymbol{\theta}$ is equal to zero or not. This kind of binary hypothesis test is called "two-sided" because the alternative H_1 lies on both sides of H_0. Alternatively, if the judgement does not aim on the equality, such as
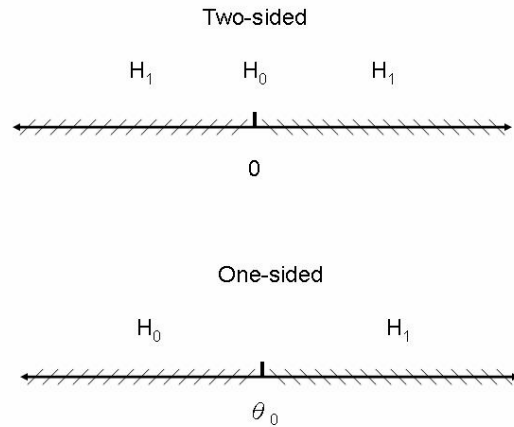
Figure 3.1: Two kinds of binary hypotheses.

$$H_0: \boldsymbol{\theta} > \theta_0 \text{ versus } H_1: \boldsymbol{\theta} < \theta_0,$$

then it is said to be "one-sided" test of a composite hypothesis versus a composite alternative.

False Alarm and Power of A Hypothesis Test

There are two categories of errors: *type I error* and *type II error*. Specifically, let $\mathbf{X}$ denote a random vector with distribution $F_\theta(\mathbf{x})$, $\boldsymbol{\theta} \in \Theta$. If the test is a binary hypothesis of $H_0:\boldsymbol{\theta} \in \Theta_0$ versus $H_1:\boldsymbol{\theta} \in \Theta_1$, the problem can be expressed as

$$\phi(\mathbf{x}) = \begin{cases} 1: & H_1, \ \mathbf{x} \in R, \\ 0: & H_0, \ \mathbf{x} \in A. \end{cases} \tag{3.10}$$

The meaning of the notation is "if the $\mathbf{x}$ belongs to region $\mathbf{R}$, the test function $\phi(\mathbf{x}) = 1$ and hypothesis H_1 is accepted; if the $\mathbf{x}$ lies in region $\mathbf{A}$, the test function $\phi(\mathbf{x}) = 0$ and hypothesis H_0 is accepted, H_1 is rejected."

In signal transmission, the type *I* error can be interpreted as the parameter $\boldsymbol{\theta} \in \Theta_0$ but $\mathbf{x} \in R$, and hence H_0 is rejected. This is called false alarm and the corresponding probability is given by

$$
\begin{aligned}
\alpha &= P_{\boldsymbol{\theta}_0}[\phi(\mathbf{X}) = 1] \\
&= E_{\boldsymbol{\theta}_0}\phi(\mathbf{X}) \\
&= P_{FA}
\end{aligned}
\tag{3.11}
$$

where $E_{\boldsymbol{\theta}_0}\phi(\mathbf{X})$ is the expectation of $\phi(\mathbf{X})$ under the density function $f_{\boldsymbol{\theta}_0}(\mathbf{X})$. Furthermore, if H_0 is composite, then it is defined to be the following supremum:

$$
\alpha = \sup_{\boldsymbol{\theta} \in \Theta_0} E_{\boldsymbol{\theta}}\phi(\mathbf{X}),
\tag{3.12}
$$

which is the worst probability of accepting H_1 when H_0 is in force.

If $\boldsymbol{\theta} \in \Theta_1$ and $\mathbf{x} \in R$, H_1 is accepted when H_1 is really in force. The probability of correctly accepting H_1 is the so-called *power* or *detection probability*

$$
\begin{aligned}
\beta &= P_{\boldsymbol{\theta}_1}[\phi(\mathbf{X}) = 1] \\
&= E_{\boldsymbol{\theta}_1}\phi(\mathbf{X}) \\
&= P_D.
\end{aligned}
\tag{3.13}
$$

The probability of *type II error* can be expressed as:

$$
P_M = 1 - P_D,
\tag{3.14}
$$

which is the probability of accepting H_0 when H_1 is in force. If H_1 is composite, the power is defined for each $\boldsymbol{\theta} \in \Theta_1$ and given by the notation $\beta(\boldsymbol{\theta})$. If $\boldsymbol{\theta}$ is scalar and Θ_1 a continuous set, then the power might look like the curve in Fig. 3.2.
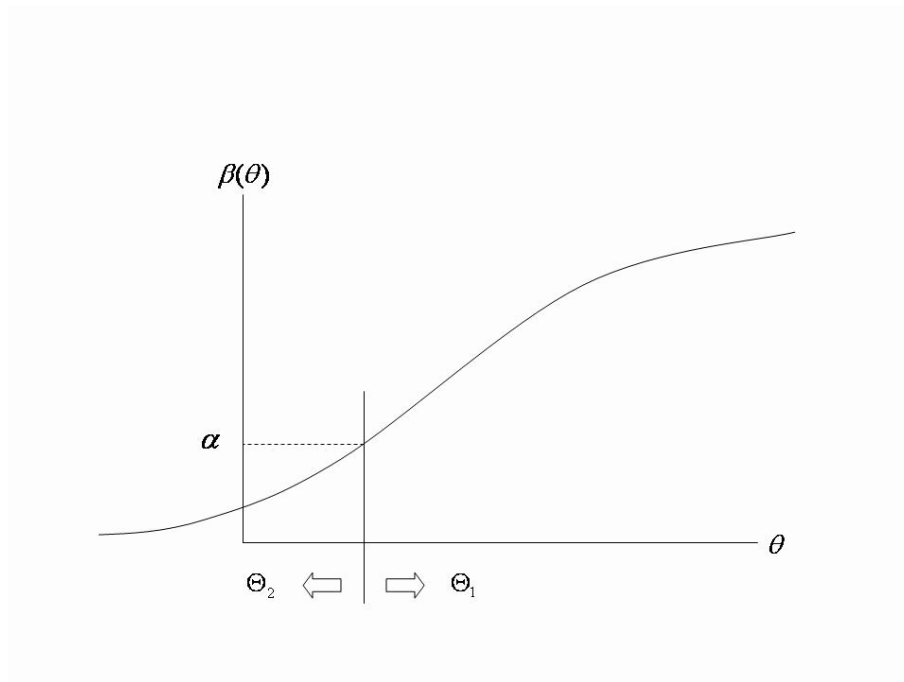
Figure 3.2: A typical power curve.

3.2.2 Sphericity Test

The sphericity test is one kind of well-known algorithms for estimating the number of signals from the eigenvalues of the spatial correlation matrix of the array. The basic procedure of sphericity test is a hypothesis test to determine the equivalence of smallest eigenvalues. James [8] and Lawley [9] have contributed to the test under real correlation matrices; but for array processing application the spatial correlation matrix is complex. Recently, however, several publications, such as, [10], [11], and [12], have considered complex spatial correlation matrices.

Suppose we have the correlation matrix $\mathbf{R}$ of a length M Gaussian random vector. The

test is

$$H_0 : \quad \mathbf{R} \quad = \quad \sigma_n^2 \mathbf{I}_M,$$

$$H_1 : \quad \mathbf{R} \quad \neq \quad \sigma_n^2 \mathbf{I}_M, \tag{3.15}$$

where σ_n^2 is unknown. Because we are interested in the equality of eigenvalues, we rewrite (3.15) as

$$H_0 : \quad \lambda_1 = \lambda_2 = \cdots = \lambda_M,$$

$$H_1 : \quad \lambda_1 > \lambda_M, \tag{3.16}$$

where the λs represent the eigenvalues of the correlation matrix and M is the number of sensors.

For the source detection problem, the number of sources should be less than the number of sensors, which is also the number of eigenvalues. Assume there are k, less than M, sources in the detection range, the $M - k$ smallest eigenvalues would be equal and the others are greater in magnitudes. The test takes a series of nested hypothesis tests to verify the $M - k$ eigenvalues for equality. The hypotheses are

$$H_0(k) : \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_M,$$

$$H_1(k) : \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{k+1} \geq \lambda_{k+2} \geq \cdots \geq \lambda_M. \tag{3.17}$$

Since the detection problem focuses on the equality of a subset of eigenvalues, the result can be deduced from $H_0(k)$ by testing the k value. Once the null hypothesis is true, we can say that the number of sources is k.

Making a decision based on hypotheses testing needs a threshold for judgement. Mauchley

[13] provides the form of sphericity test as

$$T(l_{k+1}, l_{k+2}, \cdots, l_M) = \ln \left[\frac{\left(\frac{1}{M-k} \sum_{i=k+1}^{M} l_i \right)^{M-k}}{\prod_{i=k+1}^{M} l_i} \right] \begin{array}{c} H_0(k) \\ \lessgtr \\ H_1(k) \end{array} \gamma \qquad (3.18)$$

where l_i denotes sampled eigenvalues from the sampled correlation matrix $\hat{\mathbf{R}}$ and, of course, they have no need to be the same as the true ones. In the equation we notice that γ is what we depend on to determine whether the hypotheses is true or not. The threshold γ can be set according to the Neyman-Pearson criterion [14]. A brief introduction is provided in appendix.

This approach needs knowledge of the distribution of the sphericity under the null hypothesis. Box [15] proposes a likelihood ratio test which shows that

$$2\left((N-1) - \frac{2M^2+1}{6M} \right) T(l_1, \cdots, l_M) \qquad (3.19)$$

is asymptotically a chi-square distribution with degrees of freedom $M^2 - 1$, i.e., $\chi^2_{M^2-1}$. Furthermore, Liggett [16] gives the test statistic

$$2\left((N-1) - k - \frac{2(M-k)^2+1}{6(M-k)} \right) T(l_{k+1}, \cdots, l_M), \qquad (3.20)$$

which is asymptotically distributed as $\chi^2_{(M-k)^2-1}$. However, both have constraints in application. Equation (3.19) is only applicable when all of the eigenvalues are being tested for equality and (3.20) fails to account for the influence of the larger eigenvalues on the distribution of the smaller eigenvalues being tested for equality. As a result, they may converge to a different χ^2 distribution or even not a χ^2 distribution. In 1990, Williams and Johnson [12] proposed an improved test statistic to overcome the constraints limiting the performance of

array processing algorithm. It is

$$2\Big((N-1)-k-\frac{2(M-k)^2+1}{6(M-k)}+\sum_{i=1}^{k}\Big(\frac{l_i}{\bar{l}}-1\Big)^{-2}\Big)\cdot\ln\left[\frac{\Big(\frac{1}{M-k}\sum_{i=k+1}^{M}l_i\Big)^{M-k}}{\prod_{i=k+1}^{M}l_i}\right],\qquad(3.21)$$

which is asymptotically distributed as $\chi^2_{(M-k)^2-1}$, where $\bar{l}=\sum_{i=k+1}^{M}\frac{l_i}{M-k}$. Improvement in performance occurs when the smaller eigenvalues of $\hat{\mathbf{R}}$ being tested for equality are close in value to the larger sample eigenvalues; in other words, the larger eigenvalues in the noise subspace are close to the smaller ones in signal subspace. This new sphericity test statistic performs as well as AIC and MDL, even better in many cases. Using (3.21) for passive array processing may solve the problem of source detection when it may be unsolvable by using other source detection algorithms.

3.2.3 Control of the global level

There are often a number of interesting parameters to be estimated or a number of interesting hypotheses to be tested. In some cases the problems may be treated separately without any connection to each other. In most cases, however, the problems are connected and thus we face a multiple statistical inference problem where we need to consider the different problems simultaneously.

The multiple test algorithms have been designed and interpreted from several angles. Some tests are based on decision theoretic concepts, while others are based on probabilities of making wrong decisions. In the following we use the common approach of protection against error of the first kind to avoid any true hypotheses.

Let the hypotheses in a multiple test problem be denoted by H_i and the alternatives to those by A_i. A multiple test procedure is a rule assigning to each outcome a set of rejected

31

hypotheses. That is, for a single null hypothesis H_1 against an alternative A_1, the size of the test is defined as the supremum of the probability of the critical region C_1 when the hypothesis H_1 is true. This probability of error of the first kind is always kept at a small level α. If the null hypothesis H_1 is true, the probability of accepting it is $1 - \alpha$. When we have made a "discovery" by rejecting the null hypothesis, we can be quite confident that the null hypothesis is not true. This implies that we do not make any a "discovery" by accepting the null hypotheses, because we do not have such a protection against errors of the second kind. In such a multiple test, there are many possible combinations of null hypotheses H_1, H_2, $\cdots$, H_M. If the "discoveries" in the form of rejected null hypotheses need to be claimed safely, the probability of rejecting any true null hypotheses must be kept small.

3.2.3.1   The Bonferroni Procedure

Before discuss about Bonferroni procedure, we need to know what is familywise error rate (FWER). In mutiple comparison procedure (MCPs), existence of one or more false results is called familywise error. MCPs aim to control the probability of committing any type I error in families of comparisons under simultaneous consideration. FWER (familywise error rate) is probability of familywise error.

$$
\begin{aligned}
\text{FWER} \ &= P(\text{FWE}) \\
&= P(\text{One or more } H_i | H_o) \\
&= P(\text{Max } H_i | H_0),
\end{aligned}
\tag{3.22}
$$

where $P(\cdot)$ is a probability function, H_0 is null hypothesis and H_i are alternative hypotheses, i $\neq 0$.

It is an important problem in multiple inferences to control of type I error. The traditional concern in multiple hypotheses focuses on the probability of erroneously rejecting any of the true null hypotheses, i.e., FWER. Given a significance level α, the control of FWER requires each of the M tests to be conducted at a lower level. In Bonferroni procedure, the significance

32

level of each test is under α/M. Reference [17] has more discussion about this. Figure 3.3 shows the flow of procedure.

3.2.3.2 A Simple Sequentially Rejective Test

Holm proposed a technique based on the Boole inequality which was developed in the multiple interference theory called the "Bonferroni technique" in Section 3.2.3.1. For this reason the method is called "sequentially rejective Bonferroni test" [1].

When there are M hypotheses $H_1, H_2, \cdots, H_M$ to be tested with the level α/M, the probability of rejecting any true hypotheses is smaller than or equal to α according to the Boole inequality. It constitutes the classical Bonferroni multiple test procedure with the multiple level of significance α.

In the procedure we need some test statistics, $Y_1, Y_2, \cdots, Y_M$, to test the separate hypotheses. Denote by y the the outcome of these test statistics. The critical level $\hat{\alpha}_k(y)$ for the test statistic Y_k is equal to the supremum of the probability $P(Y_k \leq y)$ when the hypothesis H_k is true. We denote $B_k = \hat{\alpha}_k(Y_k)$ so that the Bonferroni procedure can be achieved by comparing the levels $R_1, R_2, \cdots, R_M$ with α/M (see Fig. 3.3).

The sequentially rejective Bonferroni test has a slight difference by modified version than the classic one. Let $R_{(i)}$s be the sorted representation of R_is, i.e., $R_{(i)}$s are the ordered set of the original R_is in the sense that $R_{(1)} \leq R_{(2)} \leq \cdots \leq R_{(M)}$. In the procedure, we compare the obtained levels with the numbers $\dfrac{\alpha}{M}, \dfrac{\alpha}{n-1}, \cdots, \dfrac{\alpha}{1}$ instead of $\dfrac{\alpha}{M}$. This means that the probability of rejecting any set of hypotheses using the classical Bonferroni test is smaller than or equal to that using the sequentially rejective Bonferroni test based on the same test statistics. The classical Bonferroni test has been applied mainly in cases where no other multiple test procedure is presented. We can always replace the classic Bonferroni test by the corresponding sequentially rejective Bonferroni test without loss.
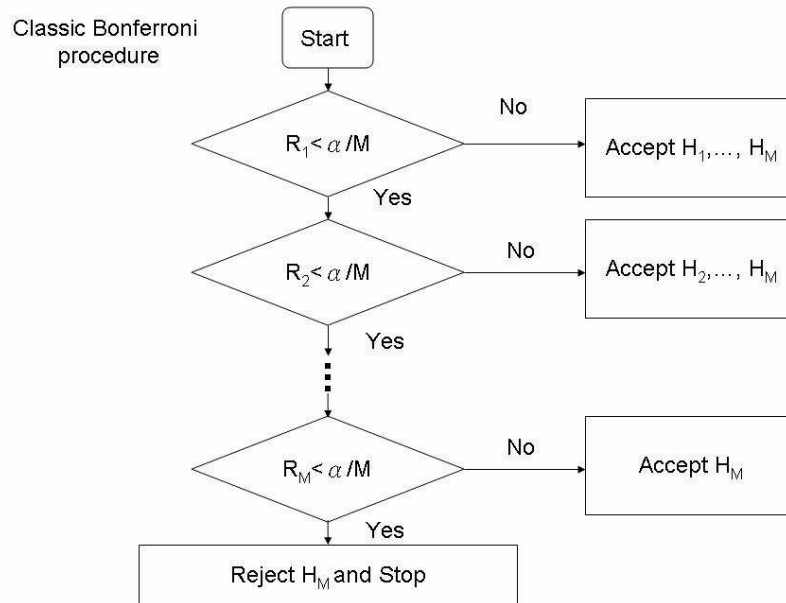
Figure 3.3: Classic Bonferroni scheme.

Figure 3.4 shows the sequentially rejective Bonferroni procedure. We need to notice that the judgement of accepting or rejecting null hypotheses depends on the sorted $R_{(i)}$ corresponding to $H_{(i)}$. Once we reject or accept some $H_{(k)}$ where k is between 1 and M, it implies that which one we choose is actually the original hypothesis corresponding to $H_{(k)}$.

The great advantage with the sequentially rejective Bonferroni test is flexibility. There are no restrictions on the type of test except we need the required level for each separate test.

3.2.3.3 False Discovery Rate

Though the above-described classical multiple comparison procedures (MCPs) have been used since the early 1950s, they have not yet been adopted widely. Despite of their use in many institutions such as the FDA of the USA or some journals, these schemes overlook
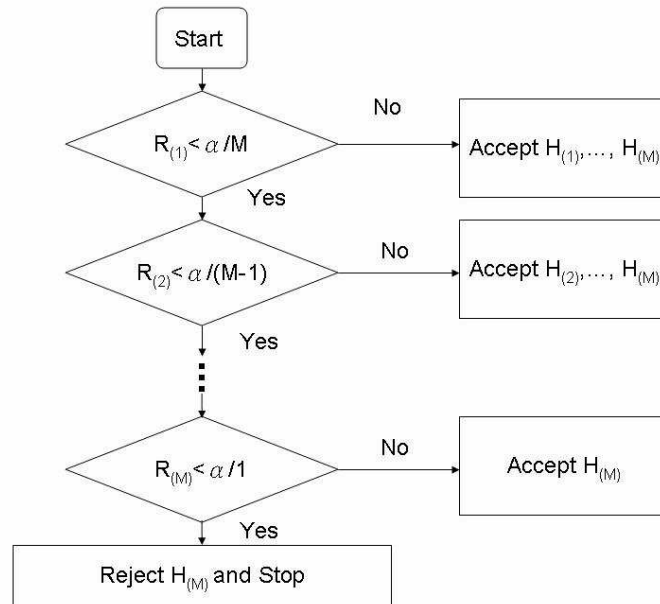
34

Figure 3.4: The Sequentially rejective multiple test procedure

various kind of multiplicity and results in exaggerated treatment differences. Classical MCPs aim to control the probability of committing any type I error in families of comparisons under simultaneous consideration. The control of familywise error rate (FWER) is usually required in a strong sense, say, under all configurations of the true and false hypotheses tested.

Classical MCPs have some difficulties causing constraints in research. First of all, classical procedures controlling the FWER in the strong sense tend to have substantially less power than the per comparison procedure of the same levels. Secondly, the control of the FWER is not so required to be adopted in any kind of application. The control of FWER is quite substantial when a decision from the various individual inferences is likely to be erroneous when at least one of them is. For example, several new medical treatments are compared against a standard. One single treatment is selected from the sets of treatment which are declared significantly better than the standard. However, a treatment group and a control

group are often compared by testing various aspects of the effect. The overall result that the treatment is superior need not be erroneous even if some of the null hypotheses are falsely rejected. Finally, the FWER controlling MCPs concerns comparisons of multiple treatments and families whose test statistics are multivariate normal. In practice, problems are not of the multiple-treatments type and test statistics are not multivariate normal. Actually, the families are often combined with statistics of different types.

The last issue has been partially addressed by the advancing Bonferroni-type procedures we have mentioned previously. It adopted observed individual p-value but still follows the concept of FWER control. The first two problems still present a serious issue. Some approaches called per comparison error rate (PCER), which neglects the multiplicity problems altogether, are still recommended by others. A brief introduction about p-value is provided in appendix.

To keep the balance between type I error control and power, Benjamini and Hochberg [2] have proposed a new point of view on the problem of multiplicity. The traditional concern in multiple hypotheses testing problems has been about controlling the probability of erroneously rejecting even one of the true null hypotheses, i.e., the FWER mentioned previously. The power of detecting a specific hypothesis is greatly declined when the number of hypotheses in the family increases. However, the new point of view is more powerful than FWER controlling procedures and can be used for many multiple testing problems. The new procedures are as flexible as the Bonferroni is. Here we consider the new procedures.

The Benjamini Hochberg Procedures

Let H_1, H_2, $\cdots$, H_M be the hypotheses or treatments in medical test, and P_1, P_2, $\cdots$, P_m are p-values corresponding to them. Sort these p-values into an ordered set $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}$ and each $P_{(i)}$ corresponds to a specific hypothesis $H_{(i)}$. The Bonferroni multiple

testing procedure is as follows:

$$k = \max\left\{i : P_{(i)} \leq \frac{i}{M}\,\gamma\right\} \tag{3.23}$$

where the value k implies the meaning: $H_{(i)}$ for $i=1, 2, \cdots, k$ is rejected. The procedure controls the false discovery rate (FDR) at γ.

At first this was mentioned by Simes [18] as an extension to his procedure for rejection of the intersection hypothesis that all null hypotheses are true if $P_{(i)} \leq \frac{1}{M}\,\alpha$. Though Simes showed that this extension controls the FWER under the intersection null hypothesis, Hommel [19] discussed the invalidity for individual hypotheses, i.e., the probability of erroneous rejections is greater than α. Afterwards, Hochberg [20] offered another procedure in order to control the FWER in the above invalid circumstance:

$$k = \max\left\{i : P_{(i)} \leq \frac{1}{M+1-i}\,\alpha\right\}, \tag{3.24}$$

and still reject all $H_{(i)}$ for $i = 1, 2, \cdots, k$.

Let $\gamma = \alpha$ in (3.23). Both (3.23) and (3.24) start by comparing $P_{(M)}$ with α. If $P_{(M)}$ is smaller than α, all hypotheses are rejected; otherwise, the procedure proceeds to $P_{(M-1)}$ and begins comparison again until it finds a smaller p-value that satisfies the condition; see Fig. 3.5. At the two ends of both Benjamini Hochberg procedure and FDR procedure, both their compared values are equivalent, which means both procedures are the same. Between the two ends of testing procedures, in FDR controlling procedure the set of $P_{(i)}$ are compared with $\frac{1-(i-1)}{M}\alpha$ instead of $\frac{1}{M+1-i}\alpha$ in (3.24). For the reason of the decreasing linearity between two ends, the FDR controlling procedure rejects at least as many hypotheses as Hochberg's procedure, where the compared values between two ends declines hyperbolically, and has also greater power than other FWER controlling methods.
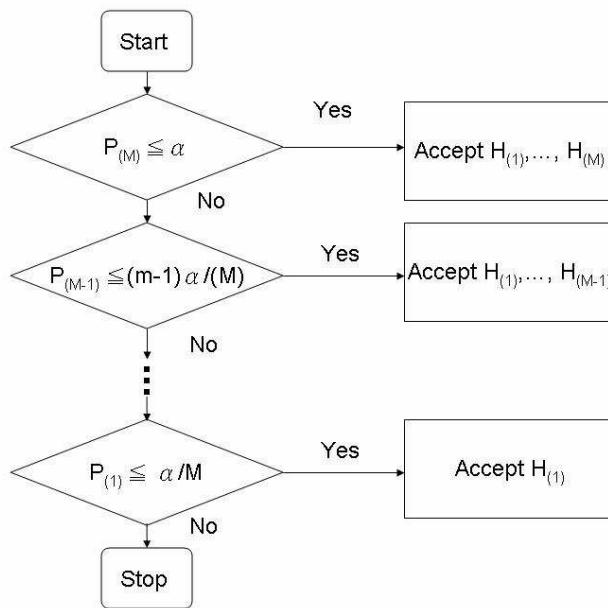
Figure 3.5: The Benjamini Hochberg procedure.

Chapter 4

Simulation

In this chapter, we demonstrate the advantages of the approach by simulation. From the algorithm in chapter 3, we know how to decide the number of signals. Take Fig. 3.5 for example, if the comparing procedure does not meet the corresponding criterion, move to the next step and test; otherwise, we accept the corresponding set of hypotheses. From Fig. 4.1, it is clear that the sorted p values increasingly climb from 0 to 1, and the compared criterion value is a linearly model. Under our setting of 3 sensors, we see the sorted p value is greater than compared value since the number of sources is 3; that is, when compared value is smaller than sorted value, we reject the two corresponding hypotheses and thus the result of detection is 3.

The whole simulation procedure is demonstrated as Fig. 4.2. First of all, we generate what kind of scenario we need to distinguish the advantages we have. Can it be outperform under closely located sources , few number of sensors, or weak sources situations? The source generation step decides what kind of scenario we try to test. Then, we collect data, use eigenanalysis to get sampled eigenvalues. Take these eigenvalues into the test statistics, depending on different proposed formulas we have discussed in chapter 2. After we adopt the test statistics, use the asymptotically chi-square distribution to get p value. Appendix will give some concept of how to get p value and why it is used in statistics area. Finally we
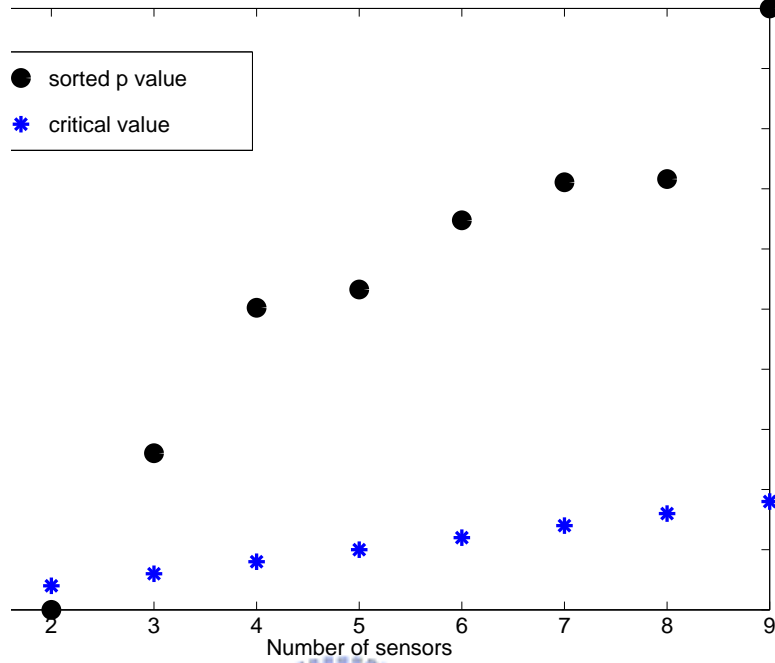
Figure 4.1: From the algorithm, the hypotheses set are rejected by the maximum number of the set which meet the criterion of procedure; that is, from the figure, the number of signals is decided from the minimum value of sorted p values greater than the corresponding compared values.

adopt Benjamini-Hochberg procedure to control the error rate.

Consider different data collected from different scenarios. First, the data is collected from an array with 7 sensors with inter-element spacings of half a wavelength. Three sources are generated from different directions, θ_1, θ_2 and θ_3, respectively. We perform 100 Monte-Carlo runs and detected the number of signals decided by different approaches. Considering 7 sensors with spatial distance d, direction of arrivals (DOA) are 10°, 50° and 80° and snapshots $T = 50$. The significance level α is 0.01. The SNR varies from -5 dB to 0 dB in 0.5 dB steps. The setting of arrival angles are general case, which means that all the sources are separate with sufficient distance. If we use other set of DOA (Degree Of Arrival), such as
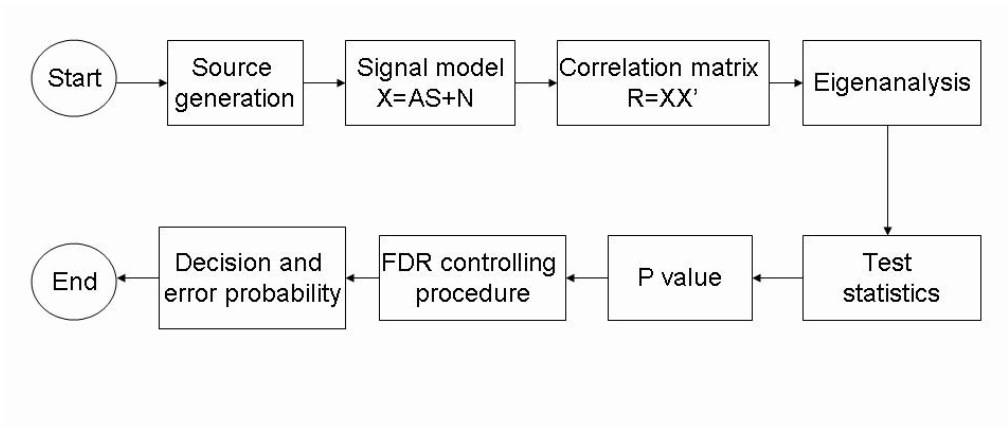
Figure 4.2: Simulation flow diagram.

10° and 50°, we still can get the same result only if the locations of sources are not too close. (Generally, closely located sources have inference on the clarity of collected data. In order to avoid this inference, we separate the sources with sufficient degree, such 10° or wider, in this work.)

Fig. 4.3 shows that when the SNR is under -2 dB, the performance of FDR controlling is better than sphericity test which is without controlling of global level of multiple test. In this figure, we add conventional approach, MDL ($\varnothing$ line), for comparison to reinforce the advantage of this approach. When the probability goes to one, the slight difference between these curves can be seen as the result of the setting of significance level and random parameters in simulation. Significance level is the largest type I error probability we can accept. Once we set the level, we can control the error probability at least under the level. Smaller significance level comes with smaller power. Generally, it is suggest to set the level between 0.2 to 0.01 depends on trade off. Under the same SNR, the proposed procedure has higher correct probability (increase around 5% to 15%) than the classical one.

Since one benefit of sphericity test comes from the closed locations of sources, we collect data from 20° and 30° with 10 sensors and other parameters remain unchanged. Figure
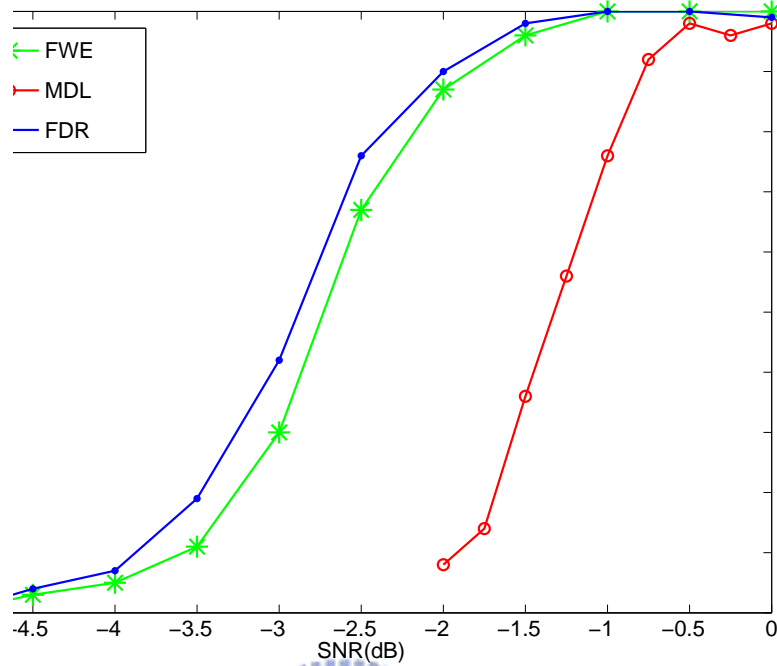
41

Figure 4.3: Three sources are generated from 10°, 50° and 80° direction of arrival, respectively. The array consists 7 sensors with snapshots $T = 50$. $\alpha = 0.01$ and SNR $= [-5{:}0.5{:}0]$ dB

4.4 shows the advantage by using the multiple hypotheses control procedure with sphericity test. The only difference between the two simulations in Fig. 4.4 is the number of sensors. The underside plot is simulated under 5 sensors and the upper is 10. Under few number of sensors, we can see that even under the situation of few number of sensors, its performance still maintains better. If one moves the locations more close, such as the 25° and 30°, the difference between the two curves will be closer; however, if we maintain the same difference between degrees, such as 40° and 50° (the difference remains 10°), the gap between two curves will not changed with their absolute degrees greatly.
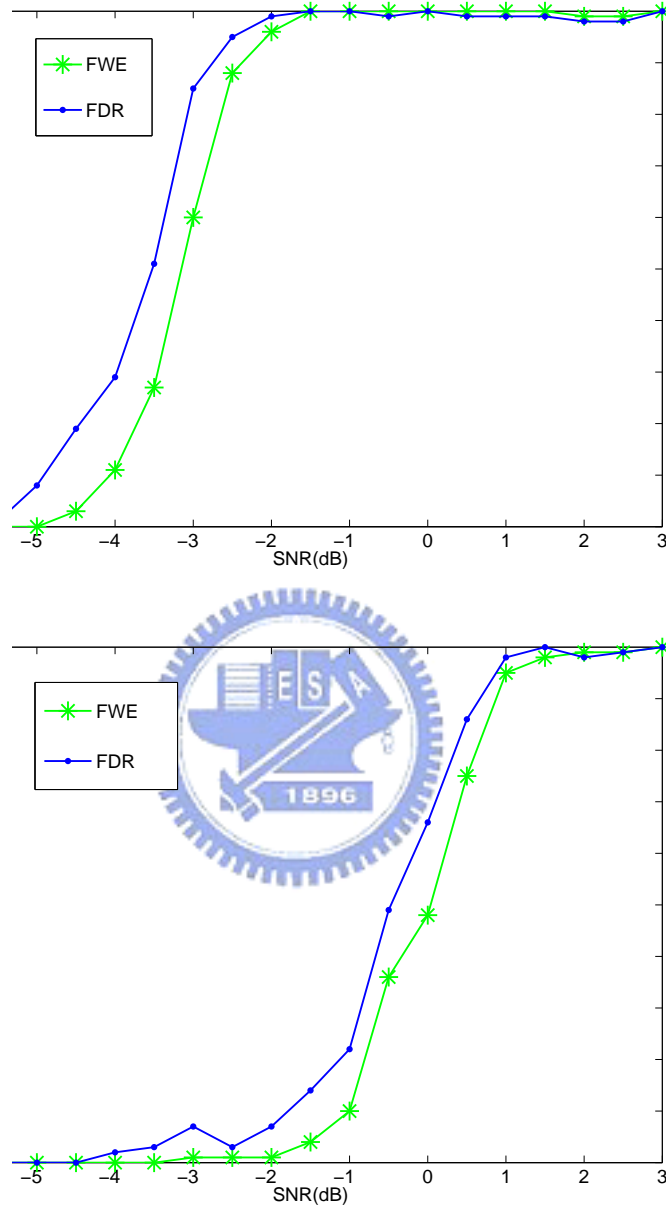
Figure 4.4: Two sources are located closely from 20°, 30° direction of arrival, respectively. The upper simulation contains 10 sensors and the underside one contains 5 only. SNR = $[-6:0.5:3]$ dB, Other parameters remain unchanged.

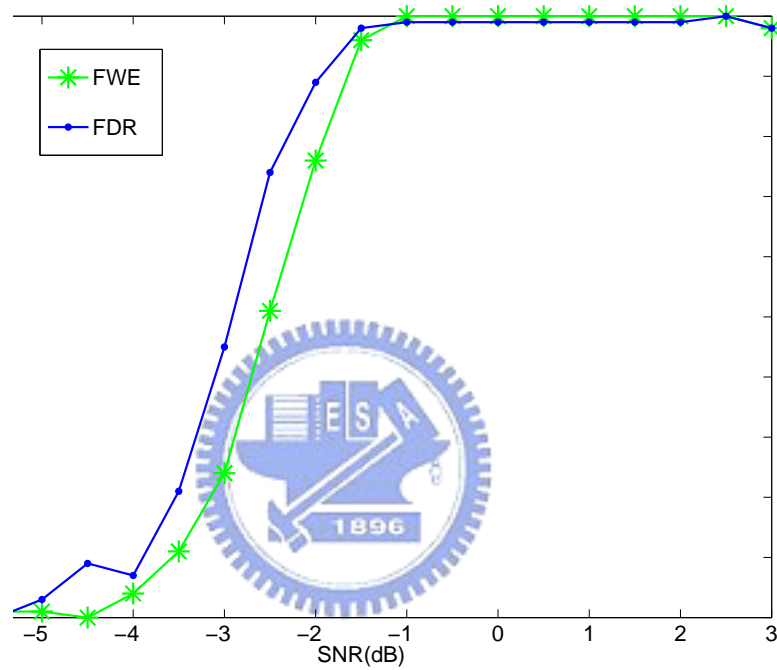Figure 4.5: Under the same scenario as Fig. 4.4, the only difference is that the sources are from 40° and 50°. It can be observed that if the difference of DOA (Degree Of Arrivals) does not change, the result almost remains unchanged.

Chapter 5

Conclusion and Future Work

Conclusion

The Sphericity test was a well-known algorithm for estimating the number of sources from the eigenvalues of the spatial correlation matrix of the array. A new sphericity test statistic was proposed for the complex case whose distribution approaches the asymptotic chi-square distribution much faster than previously proposed statistics. Experiments have shown the better performance in situations where the sources are very difficult to detect such as closely spaced or weak sources.

The approach to multiple hypotheses testing is different from the classical approaches. The classical ones need the control of the familywise error-rate (FWER) in the strong sense, a consecutive type I error rate control against any configuration of the hypotheses. FWER, as a traditional concern in multiple hypothesis problems, controls each of the tests to be conducted a certain level α. When the number of test increases, the power of the FWE-controlling procedure is substantially reduced. This is the cost of multiplicity control.

The false discovery rate (FDR) is a completely different point of view for considering the errors in multiple testing. The power of the FDR controlling method is uniformly larger than that of the other methods. It showed advantage increasingly with the number of alternative

hypotheses. In this work, we proposed that use a global FDR controlling procedure on the sphericity test to verify what we have expected.

Experiments were simulated under different scenarios where the sources were difficult to detect such as closely located and sparse number of sensors. In chapter 4, we simulated many results under such situations and the numerical simulations showed that the FDR-controlling procedure had higher probability of detection than the test statistics without global control. The improvement made the detection of sources more accurate while it is undetectable with other sphericity test statistics and sources detection algorithms. Though the performance of new test statistics has been well-proved, the control of global level of multiple test made it more accurate and without any loss in power.

Future Work

For recent years, researches on false discovery rate control have been developed in many fields. The Benjamini-Hochberg approach is a fast and effective technique for gaining power while still maintaining control of errors. There are several directions where the Benjamini-Hochberg approach can be extended.
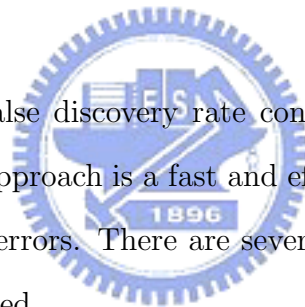
In [2], controlling the FDR has only focused on presenting and motivating the new approach. A different direction is to develop an adaptive method which incorporates the ideas of Hochberg and Benjamini.

Confidence thresholds, which have practical advantages for false discovery control, produces a threshold that controls the probability that FDP (False discovery Proportion [21]) exceeds a pre-specified level, i.e., instead of the Benjamini-Hochberg mean bound FDR $\leq \alpha$, the procedure guarantees that $P\{FDP \leq \alpha\} \geq 1 - \gamma$ small for chosen $0 < \alpha < 1, 0 < \gamma < 1$, where FDP is an unobserved random variable. It has not yet been proved, confidence en-

velopes show to be valid under the same positive dependence conditions as work for FDR control [22].

Secondly, based on a random field definition of FDP, constructs a threshold that controls the proportion of falsely identified regions of active voxels [22].

47

Appendix A

Neyman Pearson Lemma

For a hypothesis test problem, we always need a threshold to make judgement, whether the hypothesis is true or not. Neyman-Pearson criterion tells us how to find it to meet the most powerful test of size α. Let $\Theta = \{\boldsymbol{\theta_0}, \boldsymbol{\theta_1}\}$ and denote the distribution of $\mathbf{X}$ by $F_{\boldsymbol{\theta}}(\mathbf{x})$ for $\boldsymbol{\theta}$. The distribution has density function $f_{\boldsymbol{\theta}_i}(\mathbf{x})$. Let us see a powerful test of size $\alpha > 0$ for testing $H_0 : \boldsymbol{\theta_0}$ versus $H_1 : \boldsymbol{\theta_1}$. In general,

$$\phi(\mathbf{x}) = \begin{cases} 1, & f_{\boldsymbol{\theta}_1}(\mathbf{x}) > k f_{\boldsymbol{\theta}_0}(\mathbf{x}) \\ \gamma, & f_{\boldsymbol{\theta}_1}(\mathbf{x}) = k f_{\boldsymbol{\theta}_0}(\mathbf{x}) \\ 0, & f_{\boldsymbol{\theta}_1}(\mathbf{x}) < k f_{\boldsymbol{\theta}_0}(\mathbf{x}) \end{cases} \tag{1}$$

where k is nonnegative number and $0 \leq \gamma \leq 1$. If set α equals zero, we have

$$\phi(\mathbf{x}) = \begin{cases} 1, & f_{\boldsymbol{\theta}_0}(\mathbf{x}) = 0 \\ 0, & f_{\boldsymbol{\theta}_0}(\mathbf{x}) > 0 \end{cases} \tag{2}$$

is the most powerful test. When $\phi(\mathbf{x}) = 1$ we choose H_1, otherwise we choose H_0 when $\phi(\mathbf{x}) = 0$. What is the meaning of $\phi(\mathbf{x}) = \gamma$? It means we flip a γ coin to select H_1 with probability γ.

Next we are going to prove why it is said to be the most powerful. Assume $\phi'(\mathbf{x})$ is any test such that $0 \leq \phi'(\mathbf{x}) \leq 1$ and such that its size is less than or equal to the size of $\phi(\mathbf{x})$:

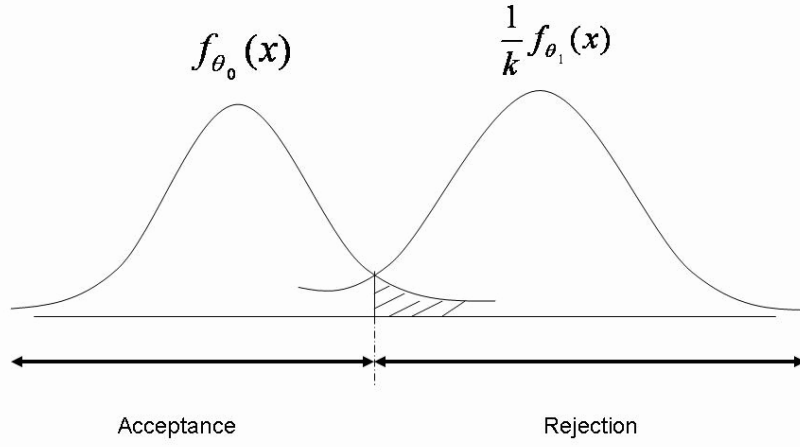$$\alpha' = E_{\boldsymbol{\theta}_0} \phi'(\mathbf{X}) \leq E_{\boldsymbol{\theta}_0} \phi(\mathbf{X}) = \alpha.$$

Figure 1: Neyman-Pearson Criterion representation. The ratio of shadowed area to whole area of $f_{\boldsymbol{\theta_0}}(\mathbf{x})$ is α, the size of test.

From the above inequality and (1):

$$\int d\mathbf{x}[\phi(\mathbf{X}) - \phi'(\mathbf{X})][f_{\boldsymbol{\theta_1}}(\mathbf{x}) - kf_{\boldsymbol{\theta_0}}(\mathbf{x})] \geq 0$$

$$\Rightarrow \int [\phi(\mathbf{X}) - \phi'(\mathbf{X})]f_{\boldsymbol{\theta_1}}(\mathbf{x})d\mathbf{x} \geq \int [\phi(\mathbf{X}) - \phi'(\mathbf{X})]kf_{\boldsymbol{\theta_0}}(\mathbf{x})d\mathbf{x}$$

$$\Rightarrow E_{\boldsymbol{\theta_1}}\phi(\mathbf{X}) - E_{\boldsymbol{\theta_1}}\phi'(\mathbf{X}) \geq k[E_{\boldsymbol{\theta_0}}\phi(\mathbf{X}) - E_{\boldsymbol{\theta_0}}\phi'(\mathbf{X})]$$

$$\Rightarrow \beta - \beta' \geq k(\alpha - \alpha').$$

Since $\alpha' \leq \alpha$, the above conclusion leads to $\beta - \beta' \geq 0$.

Here we are only introducing the basic idea of this criterion. More detail can be found in [14] and [7].

49

Appendix B

Concept of P value

The p-value is a measurement of evidence we have against the null hypotheses. The definition of p-value, for the observed sample data, is the smallest significance level which we depend on to reject null hypotheses. Significance level is the largest probability of type I error we can accept under hypotheses test. The smaller the level is, the more narrow rejective region is. If null hypotheses can be rejected under a specific significance level, it may be rejected under smaller level. However, once the level is too low to reject null hypotheses, the test has no ability of defense. For the Neyman-Pearson criterion, it was mentioned to set a significance level to perform the hypotheses test and choose the best one according to the outcomes. However, the problem is what size should the level be set? Choosing the significance level is a subjective parameter for any test. That is why we treat p-value as an evidence to support whether the null hypotheses are rejected or not.

How do we use p-value to implement a hypotheses test? The meaning of p-value is the lowest significance level which reject the null hypotheses, i.e., if the p-value is greater than the significance level, the null hypotheses cannot be rejected because the level is smaller than the critical point of rejecting H_0; otherwise, if the p-value is smaller than the significance level, the null hypotheses are rejected. The reason is that since the reject region corresponding to significance level can reject the critical point (reject hypotheses), the reject region

corresponding to p-value (smaller than significance level, smaller reject region) can reject it with no doubt.
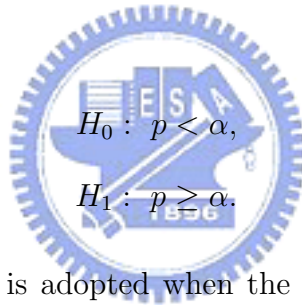
Example

Before the example, we need to translate the above illustration of p-value to mathematical formulas for convenience while discussing the problem. The rule of detection by using p-value is that

$$p\text{-value} < \alpha, \text{ then reject } H_0,$$
$$p\text{-value} \geq \alpha, \text{ then accept } H_0,$$

where α means the significance level. The above rule is based on right-tailed test which means

$$H_0 : \ p < \alpha,$$
$$H_1 : \ p \geq \alpha. \tag{1}$$

See Fig. 1. The right-tailed test is adopted when the analyzers predict the true value is greater than the critical point (assumed value). In statistics, there are other kinds of tailed test besides right-tailed test. Left-tailed test is used when we assume the true value is smaller than the critical point. It can be written as

$$H_0 : \ p \geq \alpha,$$
$$H_1 : \ p < \alpha, \tag{2}$$

and two-tailed test is applied when the both value are difficult to predict so that the test is

$$H_0 : \ p = \alpha,$$
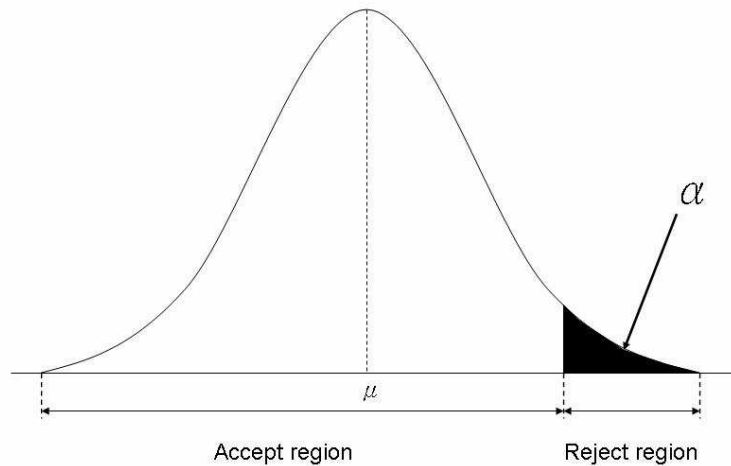$$H_1 : \ p \neq \alpha. \tag{3}$$

Figure 1: Right-tailed test.

Let us look at the example. Company A has new food product Y which need to be print-tagged outside packages. According to the factory's report, the weight of the product is 3 kg, which is a standard number. The quality department has survey of the weights to ensure the tags correspond to the product. From the factory product, which is normally distributed, 49 products are randomly sampled. The mean (μ) and the standard variation (S) are 2.95 kg and 0.38, respectively. The significance level $\alpha = 0.05$. The quality department wants to test whether average of the weights on the tag is reliable.

Because X is normally distributed and the number of samples is quite big, we use the normal distribution for testing the average. The problem can be written as

$$\begin{aligned} H_0: & \quad \mu \geq 3, \quad \text{(the product is standard and acceptable for consumer)}, \\ H_1: & \quad \mu < 3, \quad \text{(the product is not standard)}. \end{aligned} \quad (4)$$

We use the left-tailed test for calculating p-value because the decision of alternative hypothesis depends on whether $\mu < 3$. The p-value is

$$
\begin{aligned}
p &= \mathbf{P}(\bar{X} \leq \bar{X}_0 | \mu = 3) = \mathbf{P}(\bar{X} \leq \bar{X}_0 | \mu = 3) \\
&= \mathbf{P}\left(\frac{\bar{X} - 3}{0.0542} \leq \frac{2.95 - 3}{0.0542}\right) \\
&= \mathbf{P}(Z \leq -0.92) = 0.178,
\end{aligned}
\tag{5}
$$

where $\bar{X}$ and $\bar{X}_0$ represent the average of X and X_0, respectively. $\mathbf{P}(\cdot)$ denotes the probability function. The number 0.0542 comes from $S/\sqrt{n}$. At last we can find p-value through look-up tables. Obviously we can tell the result $p > \alpha$ from the calculation. The null hypothesis can not be rejected and the company maintains the quality of product well and the tag is fine.

Bibliography

[1] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, pp. 65–70, 1979.

[2] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B*, no. 1, pp. 289–300, 1995.

[3] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. New Jersey: Prentice-Hall, 1993.

[4] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd International Symposium on Information Theory* (B. N. Petrov and F. Caski, eds.), pp. 267–281, 1973.

[5] J. Rissanen, "Modeling by the shortest data description," *Automatica*, pp. 465–471, 1978.

[6] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustic Speech Signal Processing*, pp. 387–392, Apr. 1985.

[7] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.

[8] A. T. James, *Multivariate Analysis-II*. New York: Academic, 1969.

[9] D. N. Lawley, "Test of significance for the latent roots of covariance and correlation matries," *Biometrika*, no. 8, pp. 128–136, 1956.

[10] D. B. Williams, "Eigenvalues analysis for source detection with narrowband passive arrays," master's thesis, Rice University, Houston, TX, Sep. 1986.

[11] L. C. Zhao and P. R. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoustics Speech Signal Processing*, pp. 129–132, Feb. 1987.

[12] D. B. Williams and D. H. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE Trans. Acoustics Speech Signal Processing*, pp. 2008–2014, Nov. 1990.

[13] J. W. Mauchley, "Significance test for sphericity of a normal n-variate distribution," *Ann. Math. Stat.*, no. 11, pp. 204–209, 1940.

[14] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

[15] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, pp. 317–346, Dec. 1949.

[16] W. S. Liggett, "Passive sonar: Fitting models to multiple time series," in *Signal Processing: Proc. NATO ASI Signal Processing*, pp. 327–345, New York: Academic, 1973.

[17] P.-J. Chung, J. F. Böhme, A. O. Hero, and C. F. Mechlenbräuker, "Signal detection using a multiple hypothesis test," in *Proc. Third IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2004.

[18] R. J. Simes, "An improved bonferroni procedure for multiple tests of significance," *Biometrika*, no. 73, pp. 751–754, 1986.

[19] G. Hommel, "A stagewise rejective multiple test procedure based on a modified Bonferroni test," *Biometrika*, no. 75, pp. 383–386, 1988.

[20] Y. Hochberg and A. C. Tamhane, *Multiple Comparison Procedure*. New York: Wiley, 1987.

[21] C. R. Genovese and L. Wasserman, "A stochastic process approach to false discovery control," *Annals Statistics*, vol. 32, no. 3, pp. 1035–1061, 2004.

[22] C. R. Genovese and E. P. Merriam, "New procedures for false discovery control," in *Biomedical Imaging: Macro to Nano*, pp. 1498–1501, Apr. 2004.

作者簡歷

　　林宗熹，民國七十年生於宜蘭縣。民國九十三年畢業於國立成功大學工程科學系，同年進入國立交通大學電子研究所，從事訊號處理相關研究。民國九十五年取得碩士學位，論文題目為「多重假設測試於訊號偵測之應用」。研究興趣為通訊系統、統計訊號處理及數位訊號處理。