



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

APPLICATION OF STATISTICAL SIGNAL PROCESSING TECHNIQUES IN  
NATURAL ROCK TEXTURES CHARACTERIZATION AND ASTROMETRY

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA  
INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

RODRIGO ALEJANDRO LOBOS MORALES

PROFESOR GUÍA:  
JORGE SILVA SÁNCHEZ

PROFESORES CO-GUÍA:  
RENÉ MENDEZ BUSSARD  
JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:  
MARCOS ORCHARD CONCHA  
MATÍAS ZAÑARTU SALAS

SANTIAGO DE CHILE  
2015

RESUMEN DE TESIS PARA OPTAR AL  
TÍTULO DE MAGISTER EN CIENCIAS  
DE LA INGENIERÍA, MENCIÓN ELÉCTRICA  
RESUMEN DE LA MEMORIA PARA OPTAR AL  
TÍTULO DE INGENIERO CIVIL ELÉCTRICO  
POR: RODRIGO ALEJANDRO LOBOS MORALES  
FECHA: JULIO 2015  
PROFESOR GUÍA: JORGE SILVA SÁNCHEZ

## APPLICATION OF STATISTICAL SIGNAL PROCESSING TECHNIQUES IN NATURAL ROCK TEXTURES CHARACTERIZATION AND ASTROMETRY

Tanto en ingeniería en minas como en astronomía existen problemas inversos en los cuales técnicas del procesamiento de señales juegan un rol importante al momento de procesar la información existente de manera óptima. En este trabajo de Tesis los problemas de clasificación de texturas de roca y astronometría son abordados usando técnicas del procesamiento estadístico de señales.

En ingeniería en minas es de gran importancia contar con una buena caracterización del subsuelo. Para ello diversas fuentes de información son utilizadas, encontrándose entre ellas la información visual de la textura de las rocas. Pese al amplio uso de estas fuentes para hacer inferencia del tipo de roca, no se ha logrado el desarrollo de técnicas de procesamiento computacional y automático que las implementen de manera exitosa. En este trabajo de tesis, seis clases de textura de roca son analizadas utilizando técnicas avanzadas del procesamiento de imágenes. Específicamente, para cada clase se propone la extracción de características especialmente diseñadas para esa clase. Las características propuestas ofrecen un alto poder discriminador y baja dimensionalidad. Adicionalmente, se propone un esquema de banco de detectores binarios con el fin de poner a prueba las características diseñadas. Finalmente, el desempeño de clasificación del método propuesto es comparado con métodos en el estado del arte de clasificación de texturas, mostrando ganancias importantes en cuanto a error de clasificación.

El problema de astronometría corresponde a la determinación de la posición de astros mediante dispositivos detectores, comúnmente CCDs (Charged Coupled Devices). Dichos dispositivos presentan fuentes de ruido que afectan negativamente los métodos de localización. En este trabajo de tesis el método de mínimos cuadrados es analizado en detalle. En este contexto dicho método corresponde a un problema de regresión no lineal, por lo cual el desempeño o varianza del estimador resultante no puede ser caracterizado de manera analítica. Para ello se propone un método de aproximación de la varianza del estimador, que permite la comparación analítica con la cota de Cramér-Rao. Finalmente, análisis empíricos son desarrollados utilizando diversas configuraciones experimentales, encontrándose que, en determinadas condiciones de medición, el estimador es eficiente con respecto a la cota de Cramér-Rao.



# Summary

In mining engineering and astronomy there are important inverse problems where information should be optimally processed, and where signal processing techniques play an important role. In this work, the rock texture classification and the astrometry problems are treated using statistical signal processing techniques.

The first contribution of this thesis is in mining engineering, where it is necessary to have a good characterization of the subsoil. Several information sources have been used from the rock image to discriminate its nature. This is an important classification problem where automatic methods have not been systematically developed. In this thesis, six classes of rock texture are analyzed using digital image processing. Particularly, for each class the extraction of a highly discriminating and low dimensional feature is carried out. Additionally, a classifier is proposed to assess the designed features. Finally, the performance of the proposed classification method is compared with state-of-the-art methods in the texture classification problem, showing substantial gains in performance.

The second contribution is in the context of astrometry. The astrometry problem consists of the determination of a star position through the use of a detector, commonly a CCD (Charged Coupled Device). These devices have noise sources that affect the source localization problem, which corresponds to a parameter estimation task. In this thesis, the Least-Squares (LS) method is analyzed in detail. In this context, the LS method corresponds to a non linear regression problem, so its performance or variance cannot be analytically studied. For this reason, an approximation method of the estimator variance is proposed, which allows an analytical comparison with the Cramér-Rao bound. Finally, empirical analyses are developed using several experimental settings. As a result, it is found that, in some measurement regimes, the Least-Squares estimator is efficient with respect to the Cramér-Rao bound.



*A mi madre.*

# Acknowledgements

Luego de nueve años dentro de esta hermosa institución, cuesta reunir en una página a todas aquellas personas que han influido en mi formación como estudiante. Compañeros de aula, amigos, profesores y familia han estado presentes, convirtiendo este primer pasar por la universidad en algo inolvidable. A todos ellos quisiera dedicar este trabajo de Tesis a modo de retribución, que sé, equivale a un porcentaje mucho menor de lo recibido.

Me gustaría nombrar a todos mis amigos: universidad, colegio, y a los del barrio, pero creo es imposible. A aquellos que conocí en Beauchef gracias por brindarme tantas anécdotas fuera y dentro de la facultad. Gracias Mario, Gonzalo, Selene y Renato por mostrarme el lado humano de la universidad. Gracias también a mis amigos del laboratorio que en el último tiempo se convirtieron en mi segunda familia. A ese grupito de compañeros de colegio, que luego de diez años sigo viendo semana a semana en el bar regalón, mil gracias por escuchar mis locuras matemáticas y por hacerme saber que hay amistades que jamás decaen. Y a aquellos del barrio, muchas gracias por enseñarme que las amistades de infancia son las más puras. Gracias por enseñarme Mauricio, donde quiera que estés, que en la vida siempre hay que reír.

Estaré eternamente agradecido a mi profesor Jorge Silva por su constante motivación y formación. Muchas gracias por mostrarme el camino que he decidido seguir, al cual hubiese sido imposible optar sin su constante apoyo y consejo. Quiero agradecer también a los profesores René Mendez, Julián Ortiz y Marcos Orchard, por hacerme ver la pasión y gozo que provoca el descubrir.

Agradecer de manera especial a mi hermana por alegrar mi vida día a día. Te quiero mucho.

Por último, agradecer a aquellos que, sin cuyo esfuerzo, absolutamente nada de esto hubiese sido posible, mis padres. Mil gracias por sus consejos, valores, retos, y apoyo en cualquier hora, día o época del año, pero principalmente por su inmenso cariño. Gracias por haber hecho de su enorme esfuerzo mi oportunidad de poder elegir que camino continuar. Como una canción dice por ahí, soy el himno que mi padre escribió y que mi madre se encargó de cantar.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Analysis and Classification of Natural Rock Textures based on New Transform-based Features</b>	<b>3</b>
2.1	Introduction . . . . .	3
2.2	The Database of Natural Rock Textures . . . . .	5
2.3	The Chain Sequential Classification Approach . . . . .	7
2.3.1	The Image Classification Problem . . . . .	7
2.3.2	The Chain Sequential Classifier . . . . .	7
2.4	Transform-based Feature Design . . . . .	9
2.4.1	First Binary Classifier: Aphanitic Class Detection . . . . .	9
2.4.2	Second Binary Classifier: Phaneritic Class Detection . . . . .	12
2.4.3	Third Binary Classifier: Stockwork Class Detection . . . . .	14
2.4.4	Fourth Binary Classifier: Vein Class Detection . . . . .	15
2.4.5	Fifth Binary Classifier: Porphyry Class Detection . . . . .	19
2.5	Classification Performance . . . . .	21
2.5.1	Experimental Results . . . . .	21
<b>3</b>	<b>Performance Analysis of the Least-Squares Estimator in Astrometry</b>	<b>24</b>
3.1	Introduction . . . . .	24
3.2	Preliminaries . . . . .	26
3.2.1	Astrometry and Photometry based on a Photon Counting Device . . . . .	26
3.2.2	The Cramér-Rao Bound . . . . .	28
3.3	Achievability Analysis of the Cramér-Rao Bound in Astrometry . . . . .	29
3.3.1	Achievability . . . . .	29
3.3.2	Approximating the Performance of the Least-Squares Estimator . . . . .	29
3.3.3	Analysis and Interpretation of Theorem 3.1 . . . . .	30
3.4	Numerical Analysis of the LS Estimator . . . . .	33
3.4.1	Analyzing the Bounded Condition over $W(X^n, \alpha)$ . . . . .	34
3.4.2	Performance Analysis of the LS estimator . . . . .	36
<b>4</b>	<b>Conclusions</b>	<b>41</b>
<b>5</b>	<b>Appendix</b>	<b>43</b>
5.1	The Wavelet and Shearlet Transforms . . . . .	43
5.1.1	The Wavelet Transform . . . . .	43

5.1.2	The Shearlet Transform . . . . .	45
5.2	Wavelet Statistical Texture Modeling . . . . .	46
5.3	Texture-Cartoon Separation . . . . .	46
5.4	Source Separation Technique . . . . .	47
5.5	Proof of Proposition 2 . . . . .	48
5.6	Proof of Theorem 3.1 . . . . .	49
5.7	Proof of Proposition 3 . . . . .	50
<b>Bibliography</b>		<b>53</b>

# List of Tables

2.1	Success probability and SVM kernel of each binary classifier . . . . .	22
2.2	Error probability of each method . . . . .	22



# List of Figures

2.1	Example of the six classes analyzed in this work. Each column shows three images of Aphanitic, Phaneritic, Porphyry, Breccia, Stockwork and Vein respectively. . . . .	6
2.2	Decision-tree-decoding structure of the sequential classification approach . .	8
2.3	Block diagram of the chain sequential classifier . . . . .	9
2.4	Texture-cartoon separation of the Aphanitic class with a small filter . . . . .	10
2.5	Discriminative ability of $f_1$ . . . . .	11
2.6	Wavelet histogram of the cartoon part of a Phaneritic sample . . . . .	13
2.7	Discriminative ability of $f_2$ . . . . .	13
2.8	Source separation of a Stockwork sample . . . . .	16
2.9	Discriminative ability of $f_3$ . . . . .	16
2.10	Anisotropy analysis of a Vein sample . . . . .	18
2.11	Discriminative ability of $f_4$ . . . . .	19
2.12	LBP flatness of a Porphyry sample . . . . .	20
2.13	Discriminative ability of $f_5$ . . . . .	20
2.14	Success probability of each classifier for each class . . . . .	23
3.1	Variance analysis of $W(X^n, x_c)$ considering different SNR . . . . .	35
3.2	Obtained $\delta$ for different SNR . . . . .	35
3.3	Predicted histograms considering different SNR . . . . .	36
3.4	Relative error between $\sigma_{LS}^2$ and $\sigma_{CR}^2$ considering different SNR . . . . .	38
3.5	Relative performance error between the range of performance of the LS estimator and CR considering different SNR . . . . .	39
3.6	Comparison between $\sigma_{LS}^2$ and simulations . . . . .	40



# Chapter 1

## Introduction

In this Thesis, the applicability of signal processing techniques is studied for two main application domains: feature characterization and classification for natural rock textures, and the performance analysis of classical estimators in astrometry. These two problems belong to two quite distinct areas, mining engineering and astromenomy, respectively. Nevertheless, because of its transversality, signal processing offers the tools to cope with both problems. Specifically, for the former, state-of-the-art image processing techniques can be used, whereas for the latter case, classic statistical signal processing tools can be applied.

On the first problem, the rock texture is a visual information source which is used by experts in order to make decisions about geological and geo-mining-metallurgical processes inside a mine. However, in spite of the possible advantages of using this kind of information, rock texture is not commonly used because it is difficult to characterize. For this reason, in this Thesis, signal processing techniques are used to develop a new set of features for six types of rock textures. In addition, a classification approach is implemented based on the idea of a bank of detectors. The main contribution of this part of the work is the development of discriminative and low dimensional features which allow an optimal automatic process of rock textures, including their characterization and classification.

On the second problem, astrometry, is a branch of astronomy which consists of the estimation of the local position of a point source (a star), through an array of digital detector devices. Because of the noise present in these detectors, several estimation approaches have been proposed; however, there is a small number of studies about the relation between the performance of these approaches and optimal performance limits. In this Thesis, the Least-Squares estimator is analyzed in detail, and its performance is compared to the Cramér-Rao minimum variance bound. The main contribution of this part of the work is the development of mathematical expressions that allow the analytical analysis of the variance of the Least-Squares estimator. These expressions can be easily compared with the Cramér-Rao bound, allowing the analysis of the Least-Squares estimator in several observational regimes without the need of having previous observations.

The Thesis is organized as follows. In Chapter 2 a complete analysis of the rock texture classification problem is given, showing the setting of the problem and empirical results. In

Chapter 3 a complete study of the performance of the Least-Squares estimator is given in the context of astrometry. Finally, Chapter 4 provides a summary of the main results and final remarks.

# Chapter 2

## Analysis and Classification of Natural Rock Textures based on New Transform-based Features

### 2.1 Introduction

Resource evaluation is of great interest in mining because it is the first stage in the valuation of the project, where the quantity and quality of the economic ore must be determined. Subsequent stages such as mine design, planning, and process optimization are performed based on the block model defined during this stage. Currently, models go beyond the quantification of concentrations of the elements of interest. A full material characterization is required to assess the performance that the rock will have in a given process during mining or at the metallurgical plant. Therefore, in addition to a proper estimate of the tonnage and grades of interest, other characteristics of the rock must also be determined. The block model is built considering different kinds of information, such as geophysical data, geological mapping and logging, results from analytical and metallurgical tests over the rock and geotechnical surveys. Visual information takes a leading role in this characterization process. Many of the characteristics of the rock are judged from visual inspection: lithology types require a judgement about the texture of the rock; alteration types are typically defined based on the presence of alteration minerals, which may appear on the rock with diverse levels of intensity, and is also logged; clast size and shape is relevant to define the fabric of the rock. Most of this information will finally be stored as classes in a database, and carried downstream through the mining process. A proper characterization of some of these features has been found to be extremely relevant and quite difficult, specially when considering that a consistent characterization requires a trained geologist. Often the mapping and logging is done by more inexperienced geologists, adding variability to the logged data. This motivates the study of rock textures as a key visual feature of the rock [1, 2, 3], and also because the relationship with metallurgical performance has been established in many studies previously [4, 5, 6]. A good rock characterization allows for a better prediction of mining and metallurgical process responses [7], which has a direct impact on the financial performance of the project [8].

Despite its great potential, rock texture is not systematically used in mining. This may be explained because textures are difficult to define, characterize and quantify (for a detailed explanation see [9]). However, textures have been widely investigated in areas such as image processing and computer vision [10, 11, 12, 13], which offers the possibility of bringing many new algorithms and methods to the geosciences. Nevertheless, it should be noted that what we understand by textures in geology, mining engineering and image processing can be different, so it is necessary to bear in mind this fact before applying any method.

In this part of the Thesis, six classes of rock textures frequently found in porphyry copper deposits are studied through digital image processing. Specifically, the problem of classifying a database of images with few training samples is addressed, which consists of assigning a class label (corresponding to one of the six classes) to each image in the database, trying to minimize an error criterion [14]. For this purpose, it is important to extract lower dimensional discriminative features from the raw information (image), which allows an efficient computational processing. Additionally, it is considered that these classes can be divided in two groups, which will be called: *Stationary textures* and *Structure textures*. The first group, which is a typical case studied in image processing [13, pp. 81], contains rock textures defined by a stationary pattern formed by crystals with a particular distribution, shape and size. The second group contains rock textures that, as a consequence of the behavior of the minerals, present structures of different shape, size and orientation, such as clasts (rock fragments) and veins (fractures in rocks filled with minerals) [15]. The existence of the aforementioned two groups does not allow a direct application of state-of-the-art methods of texture analysis, since the structure (or object) information of the structure textures group is not properly identified [10, 11, 12, 13, 16, 17, 18, 19]. In this chapter, a new approach is proposed, which rests on the extraction of low dimensional discriminative features from the digital images, based on transform domains and source separation algorithms. These features are specially made for the six classes of interest. Then, a classification scheme is built, composed by a concatenation of five binary classifiers, which allows to deal with the different nature that a texture image could present, this is, as a stationary or as a structure pattern. Thus, the classification scheme with the feature extraction (FE) process together, are able to analyze the different kinds of information with a low computational cost.

The focus of this part of the Thesis is on the FE process, which is the development of low dimensional discriminative features. Specifically, one feature is developed for each binary classifier, and is specially designed to extract characteristic information of a particular class. For this task, transform domains given by the Wavelet transform (WT) [20] and the Shearlet transform (ST) [21] are used, to analyze properly the stationary patterns and geometric forms present in rock textures, considering different levels of scale and orientation. Additionally, these two phenomena (stationary and geometric patterns) might coexist in an image, therefore, source separation algorithms proposed in [22, 23] are integrated into the feature design. As we will see, this classification approach shows better performance than state-of-the-art approaches used for texture discrimination of natural images. This ratifies the goodness of the proposed features and the fact that the design takes into consideration the adoption of stationary patterns and object information at different levels of scale and orientation, depending on the specific nature (signature) of the considered rock class.

## 2.2 The Database of Natural Rock Textures

The problem of natural rock texture classification is not new, however, solutions are highly case dependent and it is hard to provide a general procedure. In this study, textures of rocks captured from drill hole cores digital photographs, are used to construct a database. Drill hole cores are stored in trays, which are photographed at high resolution. Although it is well known that textures are correlated to some of the process performance responses, given the difficulty in measuring them and classifying them, they are seldom used as a rock feature in the construction of models of the ore deposits. Nonetheless, geologists have a description of the different types of textures that are of interest, thus a database was constructed, where a number of cases were labelled with one of six texture categories. Typically, the rock texture can be described in terms of the matrix texture, which represents the background of the rock, and also in terms of the objects present in the matrix, namely, crystals, clasts, and structures. In this application six categories are identified, as described next:

1. Aphanitic class: this class is characterized by a matrix of the rock with the presence of small crystals which only can be observed through a microscope. Thus, what is seen is a homogeneous surface (the matrix) with slight and regular local variations.
2. Phaneritic class: this class is characterized by larger crystals that constitute the matrix of the rock and that can be observed through the naked eye. Thus again, the surface shows a homogeneous character with rougher variations (at larger scale) than the Aphanitic class.
3. Porphyry class: in this class, a matrix of small sized crystals (microscopic scale) is combined with larger crystals (visible at naked eye scale) called Porphyds. These crystals are much larger than those found in the Phaneritic class.
4. Vein class: a vein is a fracture on the rock with a sheetlike form that is filled other minerals. Thus, this class is characterized by the presence of one or more of such structures, usually depicting the same orientation. Therefore, the characteristic feature for this class is these large sheetlike structures.
5. Stockwork class: If the surface of a rock presents a random array of veins, with varying orientations, the texture is labelled as stockwork.
6. Breccia class: This class is characterized by a homogeneous matrix containing large clasts, that is, objects with angular form.

The database has 16 gray-scale images of  $128 \times 128$  pixels per class, in total 96 images. The images selected show pure classes, that is, images were labelled only when a single class was present in it. Nevertheless, it is known, in practice, that there could be an overlap of two or more classes in a single image.

For analysis purposes, it is relevant to identify two groups within the set of images:

- Stationary textures: it is composed by samples from the Aphanitic, Phaneritic and Porphyry classes, where elements of low scale form a stationary pattern formed by crystals with a particular distribution, shape and size. This group corresponds to textures associated with igneous rocks, that is, rocks created by the crystallization of cooling magmas. It should be noted that the term texture in this case is applied to

describe the shape, size and ordering of the crystals, which form a stationary pattern. Therefore, the main difference between textures in this set is related to their crystals. In the Aphanitic class we have small crystals which can only be observed through a microscope; in the Phaneritic class, larger crystals that can be observed through the naked eye; finally, in the Porphyry class, two types of crystals are found: small size crystals (microscopy scale) and larger crystals (naked eye scale) called Porphyds. Furthermore, this last class has the biggest size crystals among the three classes of the group.

- Structural textures: it is composed by samples from the Vein, Stockwork, and Breccia classes, where structural features define the corresponding texture, based on the shape and relative orientation of these objects within the rock [15]. These textures are also associated to igneous processes, where intrusion of magmas within already formed rock generated the structural feature. In the vein class, the rock shows a fracture that is filled by a different material, therefore, the characteristic feature is the presence of one or few structures with similar orientation. When the rock has been broken not by one, but by several veins in random orientations, that is, when we have a random array of veins, then the class is labelled as Stockwork. Finally, breccias are characterized by the presence of large clasts (fragments of rocks with sharp edges), within a homogeneous matrix.

Examples of the six classes are shown in Fig.2.1.

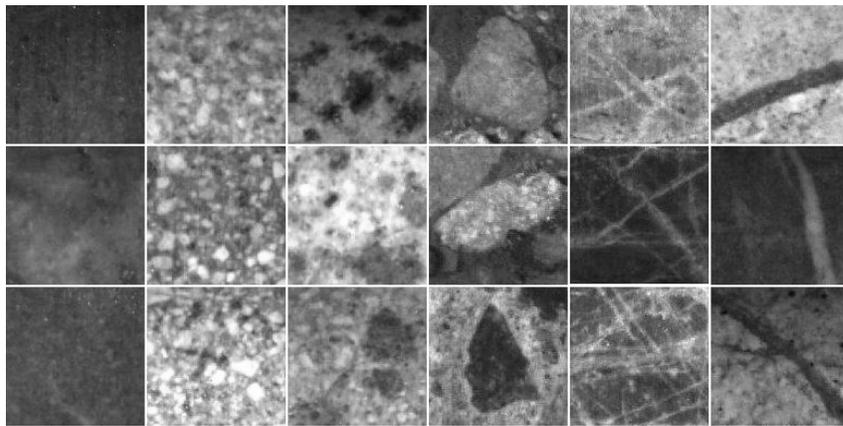


Figure 2.1: Example of the six classes analyzed in this work. Each column shows three images of Aphanitic, Phaneritic, Porphyry, Breccia, Stockwork and Vein respectively.

In the following section, the image classification problem is generally explained, and also the proposed classification approach is presented.

## 2.3 The Chain Sequential Classification Approach

In this section, the adopted classification approach is explained. First, the image classification problem is presented in general, and then the adopted sequential classification scheme is shown.

### 2.3.1 The Image Classification Problem

An image  $z$  can be represented as a two dimensional vector, i.e.,  $z = (z(i, j))_{\{(i, j)\}}$  where  $(i, j) \in \{1, \dots, n_1\} \times \{1, \dots, n_2\}$ . For simplicity, the entries of an image (or pixels) can be resorted in a one dimensional vector of dimension  $n = n_1 \times n_2$ , where each image  $z$  belongs to a subspace  $\mathcal{Z} \subset \mathbb{R}^n$ . Associated to each image  $z$ , there is a hidden (non-observed) class label that is denoted by  $y \in \mathcal{Y} = \{1, \dots, M\}$ , which represents the identity (or class) of the image. The classification task reduces to find the mapping from  $\mathcal{Z}$  to  $\mathcal{Y}$ , which from the observation stipulates the identity or class of the image. In the Bayesian setting, the image is modeled as a realization of a random vector  $Z(w)$  with values in  $\mathcal{Z}$  and the class label is also a random variable  $Y(w)$  with values in  $\mathcal{Y}$ . Then, considering the distribution of the joint vector  $(Z, Y)$ , the *optimal rule* (that minimizes the probability of error) is given by [14]:

$$g^*(z) = \operatorname{argmax}_{i \in \mathcal{Y}} P(Y = i | Z = z), \quad \forall z \in \mathcal{Z}. \quad (2.1)$$

where  $P(Y = i | Z = z)$  denotes the *posteriori* probability of class being  $i$  given the observation  $z$ . This is the well-known maximum a posteriori (MAP) decision rule [14, 24].

Unfortunately, to determine  $g^*(\cdot)$  it is necessary to know the joint distribution  $P(Z, Y)$ , which is not available in practice. In contrast, we have access to i.i.d. samples  $\mathcal{D} = \{(z_1, y_1), (z_2, y_2), \dots, (z_N, y_N)\}$ , i.e., the supervised (training) data. This rises a learning problem, which reduces to approximate the optimal rule empirically from the data  $\mathcal{D}$ . For this, the problem of dimensionality reduction of the observation space  $\mathcal{Z}$  needs to be addressed. More precisely, starting from a big observation space of dimension  $n$ , we need to design a feature transformation (FT)  $f : \mathcal{Z} \rightarrow \mathcal{X} \subset \mathbb{R}^m$  where  $m < n$ . In this context,  $\mathcal{X}$  denotes the feature space, and  $X = f(Z)$  the new feature vector to do the inference about  $Y$ . The application of the FT is known as *feature extraction* (FE) process, which is essential to reduce the learning problem to something that is tractable in terms of dimensionality, given the limited data  $\mathcal{D} = \{(z_1, y_1), (z_2, y_2), \dots, (z_N, y_N)\}$ , with the ideal objective that  $f(Z)$  keeps (in average) almost all the information of  $Z$  to discriminate the hidden variable  $Y$  [14, 24].

### 2.3.2 The Chain Sequential Classifier

We propose to use a sequential approach conformed by a bank of binary classifiers (or detectors). The basic idea is to simplify the decision scheme through: a significant reduction in the dimensionality of the problem (in the FE phase), and the adoption of a sequence of simpler binary classifiers. In this context, the focus of this work is to propose new analysis



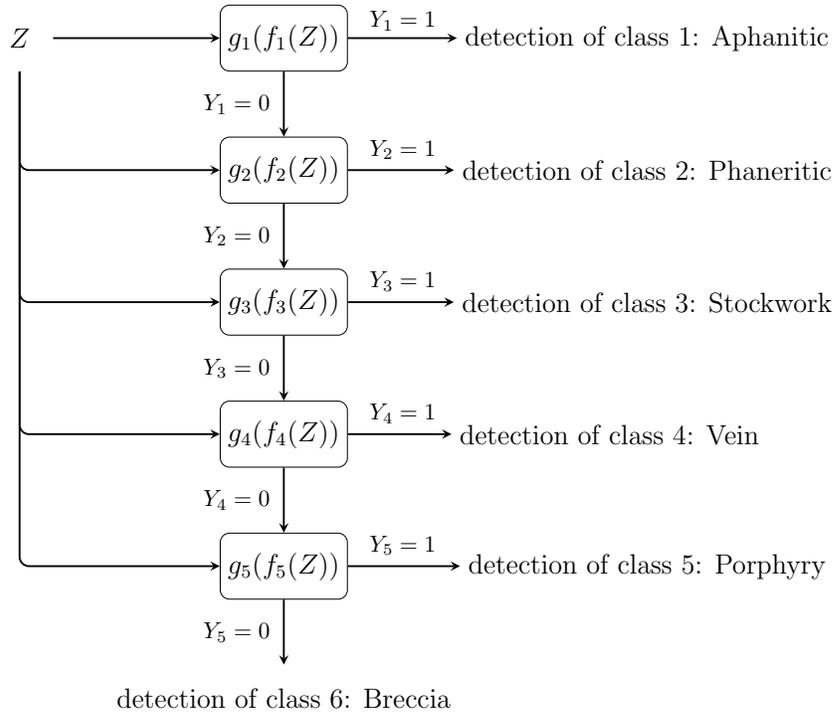


Figure 2.3: Block diagram of the proposed chain sequential classifier.

## 2.4 Transform-based Feature Design

In order to properly discriminate or detect the rock texture class of interest, the focus is put on the FE process, which reduces to design  $\{f_i\}_{i=1}^5$  for each binary classifier in Fig. 2.3. The objective for each stage is to find a parsimonious representation with a good discriminative power to detect the rock class of interest. In this section the FT is developed for each binary classifier in (2.2).

### 2.4.1 First Binary Classifier: Aphanitic Class Detection

Stationary textures present stationary patterns at different scales. In order to detect the presence of these stationary patterns it is necessary to have a technique that analyzes them at different scales or resolutions. For that, the texture-cartoon separation technique proposed in [22] is adopted, which is capable of separating the stationary patterns (or texture) from the geometric forms (or cartoon) according to a pre-established filter of width  $\sigma$ , associated to the scale of the texture-cartoon separation (for more details please see Appendix 5.3). Therefore, changing the value of the width parameter  $\sigma$  allows the separation of stationary patterns at different scales.

In particular the Aphanitic class can be seen as a stationary texture whose stationary pattern is present at a very low scale (microscope scale). For that reason, it is expected

that for a small  $\sigma$  (according to the size of the images a small filter width would correspond to  $\sigma \approx 1$ ) the texture-carton separation technique captures the whole image in the texture component, leaving the cartoon component empty. On the opposite, the cartoon part of the other 5 remaining classes should not be quite altered. Then, focusing just in the cartoon part, we can calculate the relative change of the border content before and after the separation. It is expected that for the Aphanitic images, this relative change should be very high, while for the other classes very small.

More formally, denoting an image by  $z$  and its texture and cartoon parts associated to a filter  $\sigma$  by  $z_\sigma^t$  and  $z_\sigma^u$ , respectively, the border content can be quantified by the total variation (TV) operator [25]. Specifically, for an image its TV is given by  $\|\nabla z\|_{\ell_1}$ , where  $\|\cdot\|_{\ell_1}$  denotes the  $\ell_1$ -norm. Then, the relative change previously described can be used as the feature for this classifier, and can be quantified by:

$$f_1(z) \equiv \frac{|\|\nabla z\|_{\ell_1} - \|\nabla z_\sigma^u\|_{\ell_1}|}{\|\nabla z\|_{\ell_1}} \in [0, 1], \forall z \in \mathcal{Z}. \quad (2.5)$$

To illustrate this idea, Fig. 2.4 shows the texture and cartoon components for an Aphanitic sample and a Stockwork sample, as well as the value of  $f_1$  in both cases. Fig. 2.5 shows the discriminative ability of  $f_1$ . For each sample, we show the corresponding value for  $f_1$ . It can be seen how members of the Aphanitic class are located at high values close to one, whereas the others have small values close to zero. Therefore, our designed one dimensional feature  $f_1$  shows to be a very discriminative indicator of the presence of the Aphanitic class.

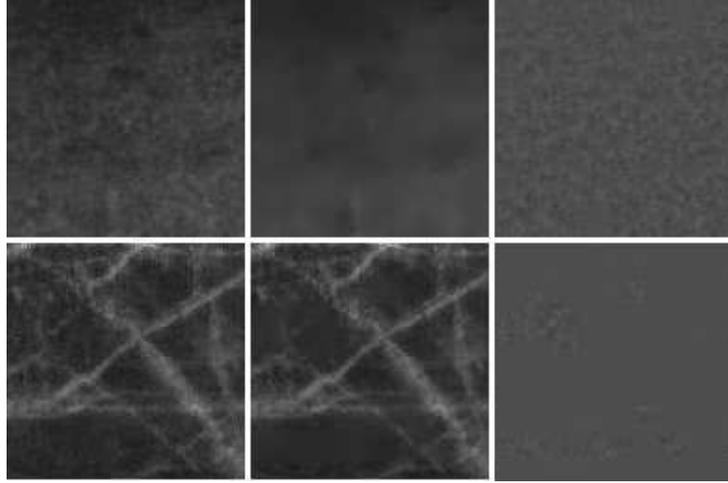


Figure 2.4: Each row shows an original image and its cartoon and texture components for an Aphanitic and a Stockwork samples, respectively, setting  $\sigma = 1$ . In the first case, we have  $f_1 = 0.811$ , and in the second one,  $f_1 = 0.134$ .

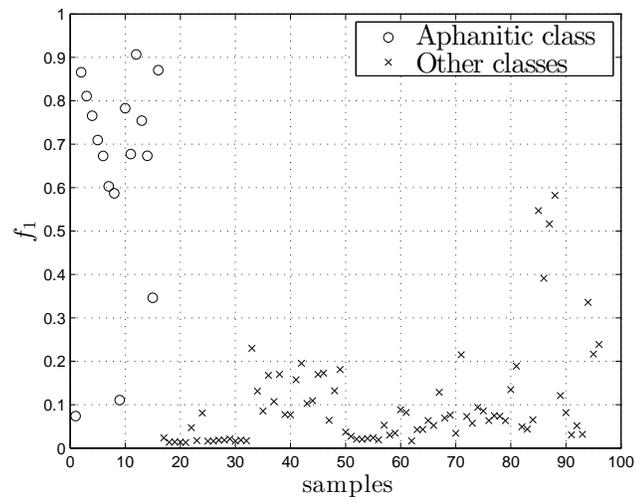


Figure 2.5: Discriminative ability of  $f_1$ . The  $x$ -axis shows different samples. From 1 to 16 correspond to Aphanitic samples, and from 17 to 96 to samples of the other classes. The  $y$ -axis shows the value of  $f_1$  for each sample.

## 2.4.2 Second Binary Classifier: Phaneritic Class Detection

Phaneritic class can be considered a stationary texture as it has stationary patterns. In contrast with the Aphanitic class, these patterns occur at several scales (see examples in Section 2.1). Particularly, there exists a specific scale at which just the Phaneritic class presents a stationary pattern. Then, the objective of the feature extraction part of this stage is to take advantage of this evidence. For that, the stationary pattern is isolated using the texture-cartoon separation technique indicated in Section 2.4.1 fixing a particular  $\sigma$  in the range<sup>1</sup> [5, 6], according to the image size. With this selection, that stationary pattern is captured in the cartoon component. On the contrary, the cartoon component of the other remaining 4 classes should not present any stationary pattern.

To quantify the presence of that particular stationary pattern, a transform domain representation is used, more precisely the *Wavelet Transform* (WT) [20] (more details in Appendix 5.1.1). For an image  $z$  in the pixel-based domain, the WT corresponds to a decomposition (or projection) of its information at several scales and locations. That decomposition is carried out by an orthonormal basis whose elements, called *wavelets*, are denoted by  $\psi_{j,k}$ , where  $j$  and  $k$  correspond to a specific scale and location indexes, respectively. Then, after the projection of  $z$  into the family  $\{\psi_{j,k}\}$ , the WT is the set of *wavelet coefficients*  $\{c_{j,k} = \langle \psi_{j,k}, z \rangle\}$  given by the inner product between  $z$  and each wavelet  $\psi_{j,k}$ . Furthermore, fixing a scale index  $j$ , the *j-wavelet sub-band* corresponds to the wavelet coefficients at all possible locations at scale  $j$ .

It is known that the histogram of coefficients in a wavelet sub-band can be modeled by a *generalized Gaussian density* (GGD) [10], which is parametrized by a *spread* parameter and a *shape* parameter, denoted by  $\alpha$  and  $\beta$ , respectively. An interesting fact is that these parameters allow to infer the form of the pdf associated to a sub-band. A high spread parameter means a wide pdf and a small shape parameter means a fast decay in the tail trend (more details in Appendix 5.2). With this statistical model, we have a probabilistic distribution for each wavelet sub-band of the image. Then, it is possible to quantify the presence of the isolated stationary pattern for the Phaneritic class because when a stationary pattern is present in an image, the pdf of the associated wavelet sub-bands tends to be wide, which means a high value for the spread and the shape parameters. On the other hand, for the rest of the classes where there is not such a stationary pattern, the pdf shape is thin, having smaller scale and shape parameters. In Fig. 2.6 this observation is clearly illustrated.

Finally, for the feature extraction of the second binary classifier, the WT with 3 scales is specifically calculated, which gives a set of 9 spread parameters and 9 shape parameters. Then, the feature vector of the second binary classifier is:

$$f_2 \equiv \begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \vdots & \vdots \\ \alpha_9 & \beta_9 \end{pmatrix}. \quad (2.6)$$

---

<sup>1</sup>A filter size in this range achieves to assign a considerable proportion of stationary patterns of the Phaneritic samples to the cartoon part. This range was chosen since the associated filters allow a better separation of the two classes of the second binary classifier in comparison to other filter sizes (please see Fig. 2.7). Consequently we have a better classification performance.

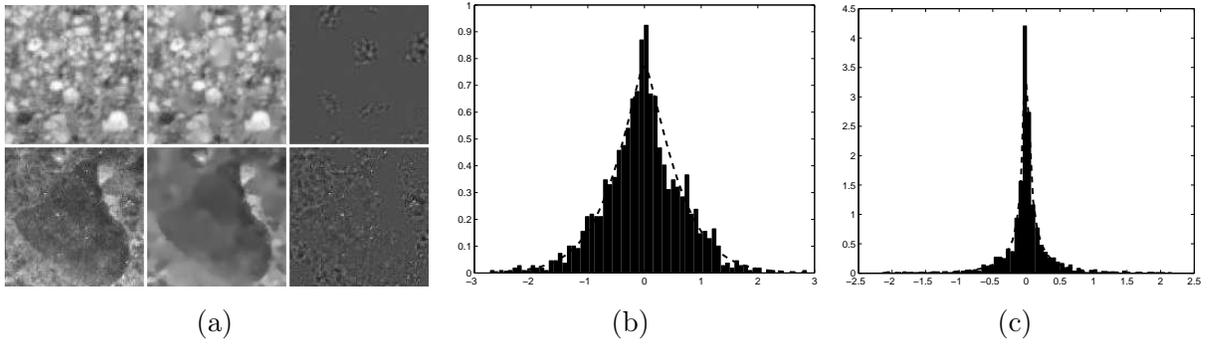


Figure 2.6: (2.6a) Each row shows an original image and its cartoon and texture components for samples of the Phaneritic and Breccia classes, respectively, setting  $\sigma = 5.5$ . (2.6b) Histogram and the fitted model ( $\alpha = 0.683$ ,  $\beta = 1.274$ ) of wavelet coefficients in a particular sub-band at the second scale for the cartoon image of the Phaneritic sample. (2.6c) Histogram and the fitted model ( $\alpha = 0.014$ ,  $\beta = 0.421$ ) of wavelet coefficients in a particular sub-band at the second scale for the cartoon image of the Breccia sample.

Fig. 2.7 illustrates the discriminative ability of  $f_2$ . For sake of space, the parameters of one sub-band are illustrated. Two clusters can be seen, corresponding to the Phaneritic class and the remaining 4 classes. Remarkably, the other sub-bands present a similar discriminative trend.

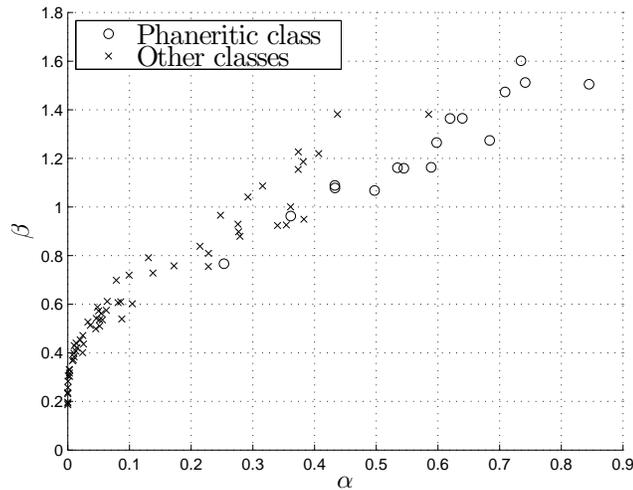


Figure 2.7: Discriminative ability of  $f_2$ . Considering just one wavelet sub-band, the  $x$ -axis shows the value of the spread parameter for 16 samples of the Phaneritic class and 64 samples of the other classes. Analogously, the  $y$ -axis shows the respective values of the shape parameter.

### 2.4.3 Third Binary Classifier: Stockwork Class Detection

The third classifier is intended to detect the Stockwork class. This can be considered as a structure texture in the sense that it is necessary to extract information associated to the presence of a geometric form or signature, which is an array of veins of a sheetlike form, illustrated in Fig. 2.8a. Adversely, the images of this class can present a stationary pattern in some parts, which hinders the analysis of the mentioned geometric forms. For that reason, it is necessary to separate the stationary patterns from the geometric forms. For that, a source separation technique is used, which is based on representing the image in two alternatives transform domains [23, 26, 27].

An interesting concept to introduce is the ability of a basis to reconstruct an image from a few proportion of transform coefficients getting an image very similar to the original. In that case, we say that the image is *compressible* with respect to that transform domain. In our context, the images composed just by geometric forms like veins are compressible in the domain given by the so called *Shearlet Transform* (ST) [21, 28], while the stationary patterns are compressible in the WT domain. The ST is similar to the WT in the sense that it is a collection of basis functions, called *shearlets*, where an image can be decomposed at several scales, locations, and additionally, several orientations. Following a similar notation, we denote the shearlets (basis elements) by  $\psi_{j,k,l}$ , where  $j$  and  $k$  correspond to a specific scale and location, respectively, and  $l$  corresponds to a specific orientation. Analogously to the WT, the ST is the set of *shearlet coefficients*, and fixing a scale  $j$ , and an orientation  $l$ , it is possible to introduce the concept of *shearlet sub-band*, which corresponds to the set of shearlet coefficients at all possible locations in those specific scale and orientation indexes (for more details please see Appendix 5.1.2).

Assuming that a Stockwork image  $z$  is composed by the sum of two independent sources, one associated to the stationary patterns (texture), and the other to the geometric (veins) forms (cartoon), denoted by  $z_s$  and  $z_g$  respectively, the source separation technique uses the fact that these two sources are compressible in the wavelet and the shearlet domains, respectively (for more details please see Appendix 5.4 and [27]). This allows an adequate separation of both sources. Fig. 2.8 shows the result of a separation using the adopted source separation technique, where it can be seen that  $z_g$  captures the geometric forms properly<sup>2</sup>.

After the source separation, if the ST is applied directly to  $z_g$  and a scale  $j$  is fixed, then for each location it is possible to calculate in which orientation the associated shearlet coefficient is maximum [29]. Next, all these found coefficients can be added in one singular value, that will be called *Orientation Energy* (OE), and will be denoted by  $\xi_j$ . More formally, considering the shearlet coefficients of  $z_g$ , denoted by  $\{\mathcal{SH}_{j,l,k}(z_g)\}$ , where  $j$ ,  $l$ , and  $k$  are the scale, orientation and location parameters, respectively, for a pixel  $k = (k_1, k_2)$  and a scale  $j$ , it is possible to define:

$$\xi_{j,k}(l) \equiv |\mathcal{SH}_{j,l,k}(z_g)|, \quad l \in L_j, \quad (2.7)$$

where  $L_j$  corresponds to the orientations associated to the scale  $j$  (for more details please

---

<sup>2</sup>In this example a contrast enhancement was made for the cartoon component of the Stockwork sample in order to observe more adequately the extraction of the random array of veins.

see Appendix 5.1.2). Then, the OE of  $z$  at scale  $j$  corresponds to:

$$\xi_j(z) \equiv \sum_{k \in \mathbb{Z}^2} \max_{l \in L_j} \xi_{j,k}(l). \quad (2.8)$$

In other words,  $\xi_j(z)$  can be seen as the  $\ell_1$ -norm of the concatenated columns of a matrix whose elements are the maximum values among the possible orientations, this is:

$$\xi_j(z) = \|\bar{I}\|_{\ell_1}, \quad (2.9)$$

$$I(j, k) \equiv \max_{l \in L_j} |\mathcal{SH}_{j,l,k}(z_g)|, \quad (2.10)$$

where  $\bar{I}$  is the vector corresponding to the concatenation of the columns of  $I$ . Thus, for a specific scale  $j$ ,  $\xi_j(z)$  gathers the coefficients with the highest energy among all possible orientations.

It happens that  $\xi_j$  tends to be high for the Stockwork class, and small for the remaining three classes. Specifically, for our database, just the OE of the second scale will be used<sup>3</sup>, which provides enough discrimination for the problem. Thus, the feature used by the third binary classifier is given by:

$$f_3 \equiv \xi_2. \quad (2.11)$$

In Fig. 2.9 the discriminative ability of  $f_3$  is shown. For each sample we show the value of  $f_3$ . It can be seen how the Stockwork class is located at high values and the other classes at small values<sup>4</sup>.

## 2.4.4 Fourth Binary Classifier: Vein Class Detection

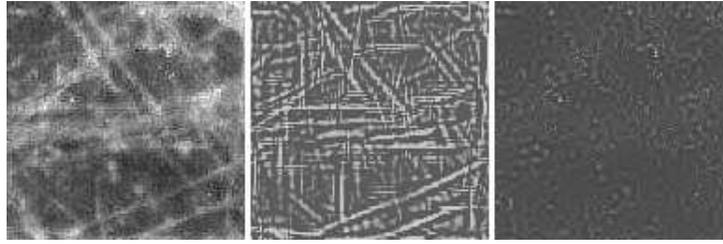
The fourth classifier is intended to detect the Vein class. These images are characterized by the presence of a considerable thick vein, which is an object of a sheetlike form following a preferential orientation. Fig. 2.10a shows a sample of the class and its respective vein with a vertical orientation. Therefore, the proposed feature extraction has to quantify the presence of this signature object and differentiate it from the objects inside images of the other two remaining classes. In fact, it is assumed that in the other classes there are just Breccia and Porphyry samples, whose objects are clasts and porphyds respectively, which have a more isotropic form than veins. Thus, it is that isotropy difference the way to discriminate the class. For that purpose, the ST is used, but in a different manner than was done before for the third classifier.

It is known that if in an image we keep a small proportion of the highest shearlet coefficients, and the rest is zeroed out, these high coefficients are those with locations corresponding to the borders of the image [30]. Then, if inside the image there is an object with a preferential orientation, the shearlet sub-band corresponding to that orientation should

---

<sup>3</sup>Considering just the OE of this scale we achieve a good separation of the Stockwork class with respect to the remaining classes. We prove adding more scales, and therefore increasing the dimensionality of the feature vector, calculating the classification performance for each case. We verified empirically that the classification error was not less for feature vectors of higher dimension.

<sup>4</sup>According to empirical results high values correspond to energies higher than 400.



(a)



(b)

Figure 2.8: (2.8a) Sample of the Stockwork class and its component of stationary patterns and the component of geometric forms, respectively ( $f_3 = 599.53$ ). (2.8b) Sample of the Porphyry class and its component of stationary patterns and the component of geometric forms, respectively ( $f_3 = 312.27$ ).

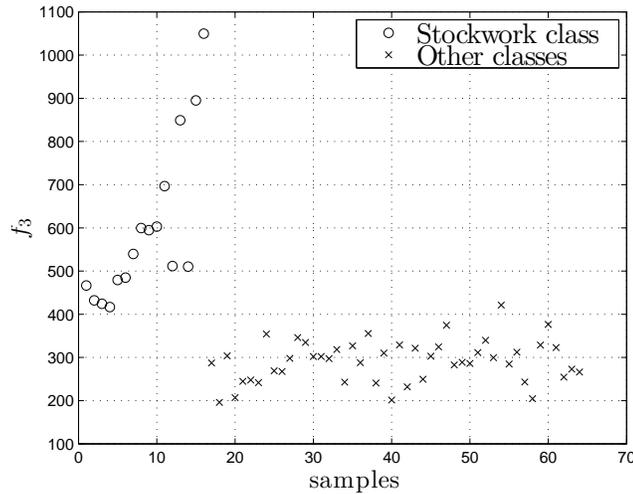


Figure 2.9: Discriminative ability of  $f_3$ . The  $x$ -axis shows different samples. From 1 to 16 correspond to Stockwork samples, and from 17 to 64 to samples of other classes. The  $y$ -axis shows the value of  $f_3$  for each sample.

have a big number of the high coefficients (located at the borders of the object). Also, in the sub-band with an orientation orthogonal to the object orientation, there should not be high coefficients, since the objects do not have borders with another direction (Fig. 2.10 illustrates the idea). Then, it is proposed to sum the absolute value of the coefficients of each sub-band (this corresponds to the  $\ell_1$ -norm of the sub-band), which is known as the *energy* of the sub-band, and the maximum of these energies should be the sub-band with the same orientation as that of the object. Thus, if we take the ratio between the energy of the

sub-band orthogonal to the object orientation and the energy of the sub-band with the same orientation, the value should be close to zero. Conversely, for objects with an isotropic form, the energy of each sub-band should be similar since the object has borders in each direction, and the ratio between any sub-band and its orthogonal counterpart should be close to one.

More formally, given an image  $z$ , its shearlet coefficients  $\{\mathcal{SH}_{j,l,k}(z)\}$ , and a percentage value  $p$ , the following set can be defined

$$\mathcal{SH}^p(z) \equiv \{\mathbb{K}_{\mathcal{P}_p}(\mathcal{SH}_{j,l,k}(z))\mathcal{SH}_{j,l,k}(z), j \leq 0, l \in L_j, k \in \mathbb{Z}^2\}, \quad (2.12)$$

where  $\mathcal{P}_p$  is the set of  $p$ -percentage most significant shearlet coefficients, and

$$\mathbb{K}_{\mathcal{P}_p}(\mathcal{SH}_{j,l,k}(z)) = \begin{cases} 1, & \mathcal{SH}_{j,l,k}(z) \in \mathcal{P}_p, \\ 0, & \mathcal{SH}_{j,l,k}(z) \notin \mathcal{P}_p. \end{cases} \quad (2.13)$$

Thus,  $\mathcal{SH}^p(z)$  are the shearlet coefficients with the  $1 - p$  non significant fraction set to zero. Also, the elements of this set can be denoted by  $\mathcal{SH}_{j,l,k}^p(z) \equiv \mathbb{K}_{\mathcal{P}_p}(\mathcal{SH}_{j,l,k}(z))\mathcal{SH}_{j,l,k}(z)$ , and the sub-band corresponding to the scale parameter  $j$ , and the orientation parameter  $l$  by  $\mathcal{SH}_{j,l}^p(z)$ . Next, for a scale  $j$ , considering its associated orientations  $l \in L_j$ , the set  $\mathcal{K}_j^p \equiv \{\kappa_{j,l}^p \in [0, 1], l \in L_j\}$  can be defined, where

$$\kappa_{j,l}^p \equiv \left( \frac{\min(\|\mathcal{SH}_{j,l}^p(z)\|_{\ell_1}, \|\mathcal{SH}_{j,\bar{l}}^p(z)\|_{\ell_1})}{\max(\|\mathcal{SH}_{j,l}^p(z)\|_{\ell_1}, \|\mathcal{SH}_{j,\bar{l}}^p(z)\|_{\ell_1})} \right)^2, \quad (2.14)$$

with  $\bar{l} \in L_j$  being the closest orientation to the orthogonal orientation of  $l$ . Then, it is defined

$$\kappa_j^{p*} \equiv \kappa_{j,l^*}^p, \quad \text{where} \quad (2.15)$$

$$l^* \equiv \arg \max_{l \in L_j} \|\mathcal{SH}_{j,l}^p(f)\|_{\ell_1}. \quad (2.16)$$

Finally,

$$f_4 \equiv \kappa_j^{p*}, \quad (2.17)$$

where  $j = 2$  and  $p = 0.03$  will be used, since they allow a good discrimination according to the images size.

Using samples of the three classes related to the fourth binary classifier, Fig. 2.10 illustrates how the high shearlet coefficients are located at the borders of images, and how  $f_4$  varies as the isotropy of objects increases. It should be mentioned that, in order to focus the analysis on the signature objects of the class, a texture-cartoon separation was done using  $\sigma = 8$  before applying the ST<sup>5</sup>, because some stationary patterns needed to be removed. Finally, in Fig. 2.11 the discriminative ability of  $f_4$  is shown. For each sample, it can be seen how the Vein class is located at low values and the other classes at high values.

---

<sup>5</sup>This filter size allows us to assign the geometric form (thick vein) to the cartoon component. The stationary patterns present in the Vein and the remaining classes are properly assigned to the texture component.

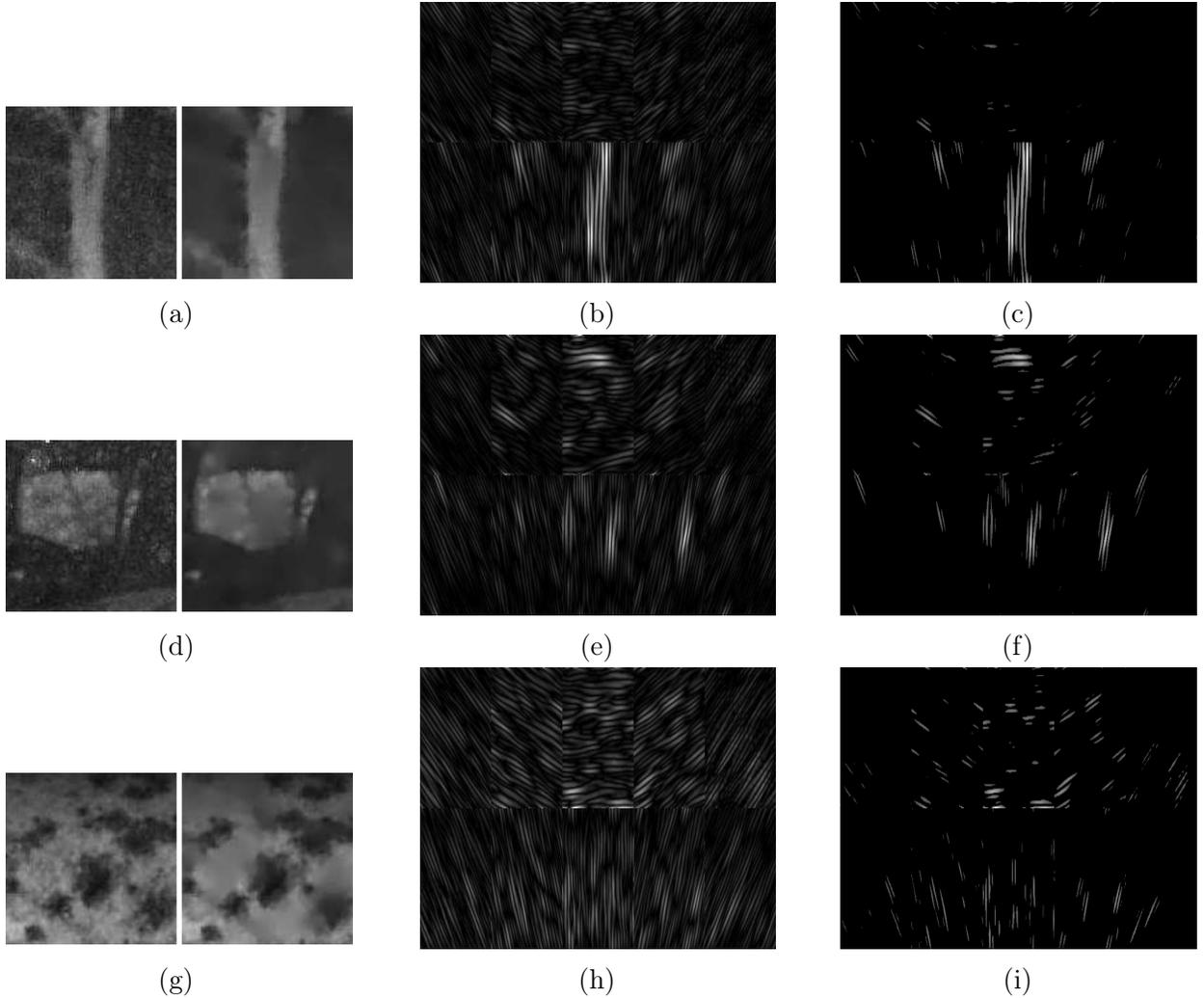


Figure 2.10: (2.10a), (2.10b) and (2.10c) show respectively a sample of the Vein class with its cartoon part ( $\sigma = 8$ ), and its shearlet coefficients at scale  $j = 2$  with ten orientations before and after considering the 3% of most significant coefficients ( $f_4 = 0.0011$ ). (2.10d), (2.10e) and (2.10f) show respectively a sample of the Breccia class with its cartoon part ( $\sigma = 8$ ), and its shearlet coefficients at scale  $j = 2$  with ten orientations before and after considering the 3% of most significant coefficients ( $f_4 = 0.7468$ ). (2.10g), (2.10h) and (2.10i) show respectively a sample of the Porphyry class with its cartoon part ( $\sigma = 8$ ), and its shearlet coefficients at scale  $j = 2$  with ten orientations before and after considering the 3% of most significant coefficients ( $f_4 = 0.7024$ ).

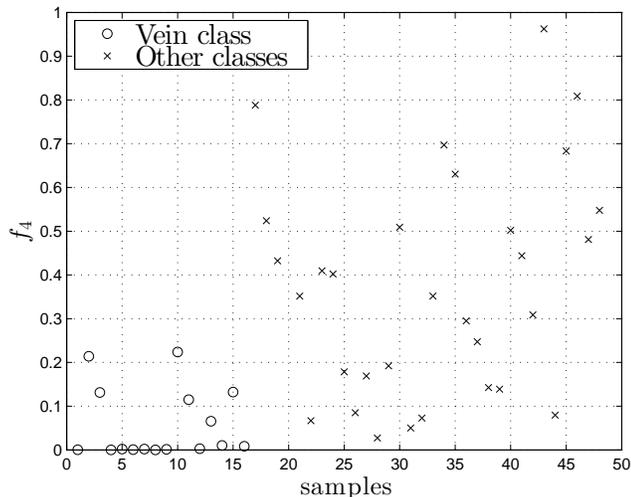


Figure 2.11: Discriminative ability of  $f_4$ . The  $x$ -axis shows different samples. From 1 to 16 correspond to Porphyry samples, and from 17 to 48 to samples of other classes. The  $y$ -axis shows the value of  $f_4$  for each sample.

## 2.4.5 Fifth Binary Classifier: Porphyry Class Detection

The fifth binary classifier is intended to discriminate the Porphyry class from the Breccia class. On one hand, the Porphyry class corresponds to a stationary texture which present stationary patterns of small and big scale. On the other hand, the Breccia class corresponds to a structure texture which is characterized by the presence of big objects called clasts. For the detection, it is intended to extract and quantify the information given by the small scale stationary patterns present in the Porphyry class, which are almost not present in the Breccia class. For that purpose *Local Binary Patterns* (LBP) are used [11]. LBP associates a binary pattern to each pixel of an image. Specifically, for each pixel its value is compared with the value of the other pixels around it considering a certain distance or profile. Next, a binary chain of length equal to the number of compared pixels is assigned to it. The value of each bit depends on whether the compared pixel was higher or smaller. Then, the frequency of the patterns is computed with a histogram.

Particularly, the low scale stationary patterns present in the Porphyry class create very frequent patterns in the LBP domain, which is reflected in the LBP histogram, since it shows an elevated zone. Conversely, the absence of stationary patterns in the Breccia class has as a consequence a flat LBP histogram as illustrated in Fig. 2.12. Then, the feature of the fifth binary classifier is developed quantifying this fact. For that, the kurtosis of the LBP histogram is proposed as feature, since this value is associated to the histogram flatness. Then,  $f_5$  is defined as follows. Given an image and its LBP histogram denoted by  $h$ ,  $f_5$  is given by:

$$f_5 = \frac{\frac{1}{n} \sum_{i=1}^n (h_i - \bar{h})^4}{\left(\frac{1}{n} \sum_{i=1}^n (h_i - \bar{h})^2\right)^2}. \quad (2.18)$$

The right-hand-side of (2.18) is the kurtosis of the LBP histogram, where  $n$  denotes the considered histogram bins,  $h_i$  the histogram value at  $i$ -bin, and  $\bar{h}$  the histogram mean. In

Fig. 2.13 the discriminative ability of  $f_5$  is shown. For each sample it is shown the value of  $f_5$ , and it can be seen how the Porphyry class is located at small values, and the Breccia class at high values.

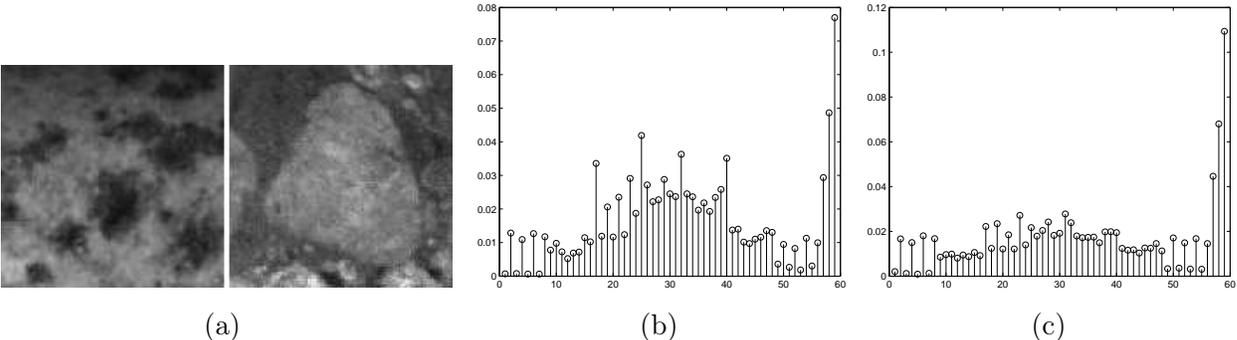


Figure 2.12: (2.12a) An example of Porphyry (left) and Breccia (right) classes, respectively. (2.12b) LBP histogram of the Porphyry image ( $f_5 = 8.465$ ). (2.12c) LBP histogram of Breccia image ( $f_5 = 21.572$ ).

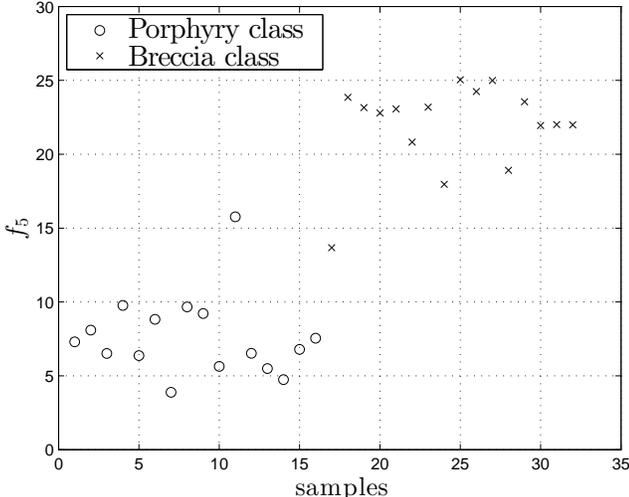


Figure 2.13: Discriminative ability of  $f_5$ . The  $x$ -axis shows different samples. From 1 to 16 correspond to Porphyry samples, and from 17 to 32 to Breccia samples. The  $y$ -axis shows the value of  $f_5$  for each sample.

## 2.5 Classification Performance

In this section we analyze the classification performance of the chain sequential classifier (CSC) presented in Section 2.3. From the binary sequential approach in Eq.(2.4), its probability of error is given by:

$$P(g_{chain}(X) \neq Y) = 1 - P(g_{chain}(X) = Y), \quad (2.19)$$

$$= 1 - \sum_{l=1}^6 P(g_{chain}(X) = Y|Y = l)P(Y = l). \quad (2.20)$$

From (2.20), the problem reduces to calculate the conditional probabilities  $\{P(g_{chain}(X) = Y|Y = l)\}_{l=1}^6$ , which involves the knowledge of the success probability for each binary detector  $\{P(g_i(X_i) = Y_i)\}_{i=1}^5$  in (2.2). More precisely, from the structure in Fig. 2.3, we have that:

$$P(g_{chain}(X) = Y|Y = l) = \prod_{i=1}^l P(g_i(X_i) = Y_i). \quad (2.21)$$

where the random binary vector  $(Y_1, ..Y_5) \in \{0, 1\}^5$  in (2.21) denotes the binary coding of the true class  $Y$ , using for that coding the path structure illustrated in Fig 2.2. Finally, since we have the same number of images per class in our database, we consider that  $P(Y = l) = 1/6, \forall l \in \{1, \dots, 6\}$ . Then, replacing (2.21) in (2.20), we have that:

$$P(g_{chain}(X) \neq Y) = 1 - \frac{1}{6} \cdot \sum_{l=1}^6 \left( \prod_{i=1}^l P(g_i(X_i) = Y_i) \right). \quad (2.22)$$

### 2.5.1 Experimental Results

To measure the performance of the proposed method Eq. (2.22) is used, then the problem reduces to compute empirically the set of probabilities  $\{P(g_i(X_i) = Y_i)\}_{i=1}^5$ , i.e., the performance of every binary detector in the chain. It was decided to proceed empirically using a cross-validation approach, due to the small size of our database, finding empirical estimations of the probabilities of error for each binary classifier.

To implement every detector, we use a discriminative approach based on *support vector machine* (SVM). The justification on using SVM is its excellent performance when very limited data is available, as well as its ability to find good decision regions in the presence of discriminative features. Several kernels for the SVM approach were evaluated, and the selected ones were those with the best empirical performance per classifier (detector).

Table 2.1 reports the performance of each binary classifier, where its success probability is listed as well the selected SVM kernel. Here it is possible to evaluate the goodness of features designed for each of the stages in Section 2.4, where the performance for the binary detectors are very good in general. The only clear exception to this rule is  $g_4$ , which shows a substantial lower accuracy. This can be attributed to the complexity of discriminating the Vein class from the Breccia and Porphyry classes. In fact, from the way the feature extraction was designed

for this detector (detecting if the objects inside the images have a preferential direction), it happens that there are some images from the Breccia class that contain structures with a certain degree of anisotropy, which causes confusions. This can be observed in Fig. 2.11. In general, Table 1 demonstrates the quality of the feature development proposed in this part of the thesis.

Table 2.1: Success probability and SVM kernel of each binary classifier

Binary Classifier	Success Probability ( $P(g_i(X_i) = Y_i)$ )	SVM configuration
$g_1$	0.938	Linear kernel
$g_2$	0.925	Polynomial kernel
$g_3$	0.969	Linear kernel
$g_4$	0.792	Polynomial kernel
$g_5$	0.932	Linear kernel

In order to evaluate the performance of the proposed method against other competing methods, two widely recognized state-of-the-art methods are considered. These are used to discriminate texture in natural images [10, 11], and are based on WT and LBP, respectively. Table 2.2 presents the probability of error of the scheme  $g_{chain}$ , and the other two classifiers. Remarkably, the proposed method shows an important reduction in classification errors when compared with the methods based on WT or LBP. To complement this analysis, Fig. 2.14 shows the success probability per class, where as expected the two natural-texture methods have good performances for stationary texture classes (Aphanitic, Phaneritic and Porphyry), however for the structured textures (Stockwork, Vein and Breccia), where there is a combination of stationary patterns at a given scale (texture) and objects (cartoon), the performance is diminished. This ratifies the motivation of this part of the thesis, in which conventional natural-texture discriminative methods would have problems detecting objects with the presence of both texture patterns and cartoon components. On the contrary, the chain sequential classifier based on the use of features that capture both aspects (texture and object as useful information for discrimination) achieves a more uniform success probability among the classes, and, more importantly, better performance overall.

Table 2.2: Error probability of each method

Method	Probability of Error
CSC	<b>0.241</b>
LBP based method	0.365
Wavelet based method	0.355

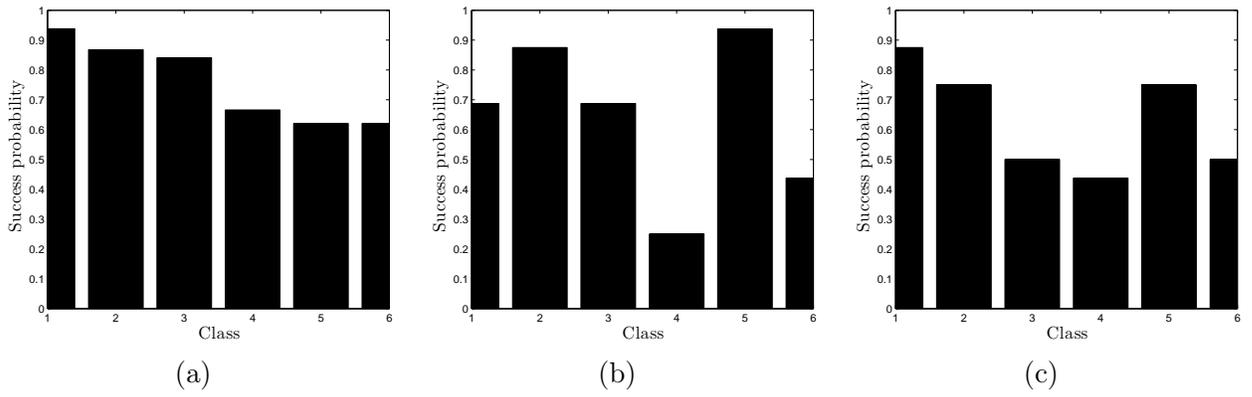


Figure 2.14: (2.14a) Success probability of the CSC for each class. (2.14b) Success probability of the Wavelet based method for each class. (2.14c) Success probability of the LBP based method for each class.

# Chapter 3

## Performance Analysis of the Least-Squares Estimator in Astrometry

### 3.1 Introduction

Astrometry is a branch of observational astronomy that deals with the precise and accurate estimation of angular positions of light-emitting (usually point-like) sources projected against the celestial sphere, and as such, it is the oldest technique employed in the study of the heavens [31, 32, 33]. Repeated measurements of positions, spread over time, allow a determination of the motions and distances of these sources, with astrophysical implications on dynamical studies of stellar systems and the Milky Way as a whole. With the advent of solid-state detectors and all-digital techniques applied to radio-interferometry and specialized ground- and space-based missions, astrometry has been revolutionized in recent years, as we have entered a high-precision era in which this technique has started playing an important role in all areas of astronomy, astrophysics [34], and cosmology [35].

Astrometrically determined distances, when combined with the measurement of the amount of light (or flux) received on earth from the observed sources (usually called “photometry”), allows us to determine the intrinsic luminosities (or electromagnetic power) of the sources, which in turns allows a comparison with theoretical models of the objects being studied to understand their nature, origin, and evolution. But, even if distances are unknown, photometric measurements alone at different wavelengths (through the use of filters that select a narrow window of the electromagnetic spectrum, and block-away the light from outside the selected window) allow reconstructing the “spectral energy distribution” of the objects, placing photometry as a fundamental tool in modern astrophysics.

Current technology, based on two-dimensional discrete digital detectors (such as *charged coupled devices* (CCDs) [36]), record a (noisy) image (on an array of photo-sensitive pixels) of celestial sources, from which it is possible to estimate both their astrometry and photometry, simultaneously. The inference problem associated to the determination of these quantities is at the core of the astrometric and photometric endeavor described previously. A number of techniques have been proposed to estimate the location and flux of celestial sources as

recorded on digital detectors. In this context, estimators based on the use of a least-squares (LS) error function have been widely adopted [37, 38, 39, 40, 41]. The use of this type of decision rule has been traditionally justified through heuristic reasons. First, LS methods are conceptually straightforward to formulate based on the observation model of these problems. Second, they offer computationally efficient implementations and have shown reasonable performance [42, 43]. Finally, the LS approach was the classical method used when the observations were obtained with analog devices [44, 45] (which corresponds to a Gaussian noise model for the observations, different from that of modern digital detectors that is characterized instead by Poisson statistics) and, consequently, the LS method was adopted from the analogous to the digital observational setting naturally.

Driven by the increase in the intrinsic precision available by the new detectors and instrumental settings, it is timely to re-visit the pertinence on the use of LS estimators. Indeed, in the digital setting, where we observe discrete samples (or counts) on a photon integrating device, there is no formal justification that the LS approach is optimal in the sense of minimizing the mean square error (MSE) of the parameter estimation, in particular for astrometry, which is the focus of this part of the thesis.

The question of optimality (in some statistical sense) has always been in the interest of the astronomical community, in particular the idea of characterizing fundamental performance bounds that can be used to analyze the efficiency of the adopted estimation schemes. In this direction, there are some seminal works on the use of the celebrated Cramér-Rao (CR) bound in astronomy by [46, 47, 48, 49]; and [50]. The CR bound is a minimum variance bound for the family of unbiased estimator [51, 52]. In astrometry and photometry this bound has offered meaningful closed-form expressions that can be used to analyze the complexity of the inference task, and its dependency on key observational and design parameters like the position of the object in the array, the intensity of the object, the signal to noise ratio (SNR), and the resolution of the instrument [53, 54, 55, 56, 57]. In particular for photometry, [56] used the CR bound to show that the LS estimator is a good estimator, achieving a performance close to the limit in a wide range of observational regimes, and achieving precisely the bound at low SNR. In astrometry, on the other hand, [54, 55] have recently studied the structure of this bound and has analyzed its dependency with respect to important observational parameters for realistic astronomical observing conditions. In this context, closed-form expressions for the bound were derived in a number of important settings (high pixel resolution, low and high SNR), and their trends were explored across pixel resolutions and the position of the object in the array. As an interesting outcome, the analysis of the CR bound allows us to find the optimal pixel resolution of the array for a given setting, as well as providing formal justification to some heuristic techniques commonly used to improve performance in astrometry, like *dithering* for undersampled images [54, Sec. 3.3].

The specific problem of evaluating the existence of an estimator that achieves the CR bound has not been covered, and remains an interesting problem. On this, [54] have assessed empirically (using numerical simulations) the performance of two LS methods and the maximum likelihood estimator, showing that their variances follow very closely the CR limit in some specific regimes. In this part of the thesis, it is analyzed in detail the performance of the LS estimator with respect to the CR bound, with the goal of finding concrete regimes, if any, where this estimator achieves the CR bound and, consequently, where it can be considered an

efficient solution to the astrometric problem. This application is a challenging one, because estimators based on a LS type of objective function do not have a closed-form expression in astrometry. In fact, this estimation approach corresponds to a non linear regression problem, where the resulting estimator is implicitly defined. As a result, no expressions for the performance of the LS estimator can be obtained analytically. To address this issue, the result given by Theorem 3.1 derives expressions that bound and approximate the variance of the LS estimator. The approach is based on the work by [58], where the authors addressed the problem of approximating the bias and mean square error of general estimators that are the solution of an optimization problem. In [59] another methodology is given to approximate the variance and mean of implicitly defined estimators, which has been applied to medical imaging and acoustic source localization [60].

The main result of this part of the Thesis is a refined version of the result presented in [58], where one of its key assumptions that is not applicable in the astrometry estimation problem, is reformulated. In this process, lower and upper bounds for the MSE performance of the LS estimator are derived. Using these bounds, we analyze how close the performance of the LS estimator is to the CR bound across different observational regimes. It is shown that for high SNR there is a considerable gap between the CR bound and the performance of the LS estimator. Remarkably, it is shown that for the more challenging low SNR observational regime (weak astronomical sources), the LS estimator is near optimal, as its performance is arbitrarily close to the CR bound.

## 3.2 Preliminaries

This section introduces the problem of astrometry as well as concepts and definitions that will be used throughout this part of the Thesis. For simplicity, just the 1-D scenario of a linear array detector is studied, as it captures the key conceptual elements of the problem<sup>1</sup>.

### 3.2.1 Astrometry and Photometry based on a Photon Counting Device

The specific problem of interest is the inference of the position of a point source. This source is parameterized by two scalar quantities, the position of object  $x_c \in \mathbb{R}$  in the array<sup>2</sup>, and its intensity (or brightness, or flux) that we denote by  $\tilde{F} \in \mathbb{R}^+$ . These two parameters induce a probability  $\mu_{x_c, \tilde{F}}$  over an observation space that we denote by  $\mathbb{X}$ . More precisely, given a point source represented by the pair  $(x_c, \tilde{F})$ , it creates a nominal intensity profile in a photon counting device (PCD), typically a CCD, which can be generally written as:

$$\tilde{F}_{x_c, \tilde{F}}(y) = \tilde{F} \cdot \phi(y - x_c, \sigma), \quad (3.1)$$

---

<sup>1</sup>This analysis can be extended to the 2-D case as shown in [54].

<sup>2</sup>This can be related to an angular position in the sky, measured in seconds of arc (arcsec thereafter), through the “plate-scale”, which is an optical design feature of the instrument plus telescope configuration.

where  $\phi(y - x_c, \sigma)$  denotes the one dimensional normalized point spread function (PSF) and where  $\sigma$  is a generic parameter that determines the width (or spread) of the light distribution on the detector (typically a function of wavelength and the quality of the observing site, see Section 3.4) [54, 55].

The profile in (3.1) is not observed directly, but through three sources of perturbations. First, an additive background noise which captures the photon emissions of the open (diffuse) sky, and the noise of the instrument itself (the read-out noise and dark-current [61, 62]) modeled by  $\tilde{B}_i$  in Eq.(3.2). Second, an intrinsic uncertainty between the aggregated intensity (the nominal object brightness plus the background) and actual measurements, which is modeled by independent random variables that follow a Poisson probability law. Finally, the spatial quantization process associated with the pixel-resolution of the PCD as specified in Eqs. (3.2) and (3.3). Including these three effects, we have a countable collection of independent and not identically distributed random variables (observations or counts)  $\{X_i : i \in \mathbb{Z}\}$ , where  $X_i \sim \text{Poisson}(\lambda_i(x_c, \tilde{F}))$ , driven by the expected intensity at each pixel element  $i$ , given by:

$$\lambda_i(x_c, \tilde{F}) \equiv \mathbb{E}\{X_i\} = \underbrace{\tilde{F} \cdot g_i(x_c)}_{\equiv \tilde{F}_i(x_c, \tilde{F})} + \tilde{B}_i, \quad \forall i \in \mathbb{Z} \quad (3.2)$$

and

$$g_i(x_c) \equiv \int_{y_i - \Delta y/2}^{y_i + \Delta y/2} \phi(y - x_c, \sigma) dy, \quad \forall i \in \mathbb{Z}, \quad (3.3)$$

where  $\{y_i : i \in \mathbb{Z}\}$  denotes the standard uniform quantization of the real line-array with resolution  $\Delta y > 0$ , i.e.,  $y_{i+1} - y_i = \Delta y$  for all  $i \in \mathbb{Z}$ . In practice, the PCD has a finite collection of measured elements (or pixels)  $X_1, \dots, X_n$ , then a basic assumption here is that we have a good coverage of the object of interest, in the sense that for a given position  $x_c$ :

$$\sum_{i=1}^n g_i(x_c) \approx \sum_{i \in \mathbb{Z}} g_i(x_c) = \int_{-\infty}^{\infty} \phi(y - x_c, \sigma) dy = 1. \quad (3.4)$$

At the end, given the source parameters  $(x_c, \tilde{F})$ , the probability mass function (pmf) of the joint random observation vector  $X_1^n = (X_1, \dots, X_n)$  (with values in  $\mathbb{N}^n$ ) is given by<sup>3</sup>:

$$f_{(x_c, \tilde{F})}(x_1^n) = p_{\lambda_1(x_c, \tilde{F})}(x_1) \cdot p_{\lambda_2(x_c, \tilde{F})}(x_2) \cdots p_{\lambda_n(x_c, \tilde{F})}(x_n), \quad \forall x_1^n \in \mathbb{N}^n, \quad (3.5)$$

where  $p_\lambda(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$  denotes the pmf of the Poisson law [63].

To conclude, the astrometric and photometric estimation problem can be posted as the problem of characterizing a decision rule  $\tau_n() : \mathbb{N}^n \rightarrow \Theta$ , with  $\Theta = \mathbb{R}^2$  being a parameter space, where given an observation  $x_1^n$  the parameters to be estimated are  $(\hat{x}_c(x_1^n), \hat{\tilde{F}}(x_1^n)) = \tau_n(x_1^n)$ .

---

<sup>3</sup>Note that from Eq.(3.5) the observations are independent but not identically distributed.

### 3.2.2 The Cramér-Rao Bound

The Cramér-Rao bound offers a performance bound on the variance of the family of unbiased estimators [51, 52, 54, 55]. More precisely, let  $X_i$  (with  $i = 1, \dots, n$ ) be a collection of independent observations that follow a parametric pmf  $f_{\bar{\theta}}$  defined on  $\mathbb{N}$ . The parameters to be estimated from  $X^n = \{X_i : i = 1, \dots, n\}$  will be denoted in general by the vector  $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_m) \in \Theta = \mathbb{R}^m$ . Let  $\tau_n(X^n) : \mathbb{N}^n \rightarrow \Theta$  be an unbiased estimator<sup>4</sup> of  $\bar{\theta}$ , and  $L(x^n; \bar{\theta}) = f_{\bar{\theta}}(x_1) \cdot f_{\bar{\theta}}(x_2) \cdots f_{\bar{\theta}}(x_n)$  be the likelihood of the observation  $x^n \in \mathbb{N}^n$  given  $\bar{\theta} \in \Theta$ . Then, the Cramér-Rao bound [51, 52] establishes that if

$$\mathbb{E}_{X^n \sim f_{\bar{\theta}}^n} \left\{ \frac{\partial \ln L(X^n; \bar{\theta})}{\partial \theta_i} \right\} = 0, \quad \forall i \in \{1, \dots, m\}, \quad (3.6)$$

then,

$$\text{Var}(\tau_n(X^n)_i) \geq [\mathcal{I}_{\bar{\theta}}(n)^{-1}]_{i,i}, \quad (3.7)$$

where  $\mathcal{I}_{\bar{\theta}}(n)$  is the *Fisher information* matrix given by:

$$[\mathcal{I}_{\bar{\theta}}(n)]_{i,j} = \mathbb{E}_{X^n \sim f_{\bar{\theta}}^n} \left\{ \frac{\partial \ln L(X^n; \bar{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln L(X^n; \bar{\theta})}{\partial \theta_j} \right\} \quad \forall i, j \in \{1, \dots, m\}. \quad (3.8)$$

In particular, for the scalar case ( $m = 1$ ), we have that for all  $\theta \in \Theta$

$$\min_{\tau_n(\cdot) \in \mathcal{T}^n} \text{Var}(\tau_n(X^n)) \geq \mathcal{I}_{\theta}(n)^{-1} = \mathbb{E}_{X^n \sim f_{\theta}^n} \left\{ \left[ \left( \frac{d \ln L(X^n; \theta)}{d\theta} \right)^2 \right] \right\}^{-1}, \quad (3.9)$$

where  $\mathcal{T}^n$  is the collection of unbiased estimators and  $X^n \sim f_{\theta}^n$ .

Returning to the problem in Section 3.2.1, [54, 55] have characterized and analyzed the Cramér-Rao bound for the isolated problem of astrometry and photometry, respectively, as well as the joint problem of photometry and astrometry. Particularly, the following results are highlighted:

**Proposition 1** ([55, pp. 800]) Let us assume that  $x_c \in \mathbb{R}$  is fixed and known, and we want to estimate  $\tilde{F}$  (fixed but unknown) from  $X^n \sim f_{(x_c, \tilde{F})}$  in (3.5). In this scalar parametric context, the Fisher information is given by:

$$\mathcal{I}_{\tilde{F}}(n) = \sum_{i=1}^n \frac{g_i(x_c)}{\tilde{F} g_i(x_c) + \tilde{B}_i}, \quad (3.10)$$

which from Eq.(3.9) induces a minimum variance (MV) bound for the *photometry estimation problem*. On the other hand, if  $\tilde{F} \in \mathbb{R}^+$  is fixed and known, and we want to estimate  $x_c$  (fixed but unknown) from  $X^n \sim f_{(x_c, \tilde{F})}$  in (3.5), then the Fisher information is given by:

$$\mathcal{I}_{x_c}(n) = \sum_{i=1}^n \frac{\left( \tilde{F} \frac{dg_i(x_c)}{dx_c} \right)^2}{\tilde{F} g_i(x_c) + \tilde{B}_i}, \quad (3.11)$$

which from Eq.(3.9) induces a MV bound for the *astrometric estimation problem*, and where  $\sigma_{CR}^2 \equiv \mathcal{I}_{x_c}(n)^{-1}$  denotes the (astrometric) CR bound.

<sup>4</sup>In the sense that, for all  $\bar{\theta} \in \Theta$ ,  $\mathbb{E}_{X^n \sim f_{\bar{\theta}}^n} \{\tau_n(X^n)\} = \bar{\theta}$ .

At this point it is relevant to study if there is any practical estimator that achieves the CR bound presented in (3.10) and (3.11) for the photometry and astrometry problem, respectively. For the photometric case, [56, Appendix A] has shown that the classical least-squares (LS) estimator is near-optimal, in the sense that its variance is close to the CR bound for a wide range of experimental regimes, and furthermore, in the low SNR regime, when  $\tilde{F}g_i(x_c) \ll \tilde{B}_i$ , its variance (determined in closed-form) achieves the MV bound  $\mathcal{I}_{\tilde{F}}(n)^{-1}$  in Eq.(3.10). This is a formal justification for the goodness of LS as a method for doing isolated photometry in the setting presented in Section 3.2.1. An equivalent analysis has not been conducted for the astrometric problem, which is the focus of the next section.

### 3.3 Achievability Analysis of the Cramér-Rao Bound in Astrometry

We first evaluate if the CR bound for the astrometric problem,  $\sigma_{CR}^2$  in (3.11), is achieved by any unbiased estimator. Then, we focus on the widely used LS estimation approach, to evaluate its performance in comparison to the astrometric MV bound presented in Proposition 1.

#### 3.3.1 Achievability

Concerning achievability, it can be said that for astrometry there is no estimator that achieves the CR bound in any observational regime. More precisely,

**Proposition 2** For any fixed and unknown parameter  $x_c \in \mathbb{R}$ , and any unbiased estimator  $\tau_n$

$$\text{Var}(\tau_n(X^n)) > \sigma_{CR}^2, \quad (3.12)$$

where  $X^n$  follows  $f_{x_c}$  in (3.5). (The proof is presented in Appendix 5.5).

#### 3.3.2 Approximating the Performance of the Least-Squares Estimator

For astrometry, a version of the least-squares estimator (given the model presented in Section 3.2.1) corresponds to the solution of:

$$\tau_{LS}(x^n) = \arg \min_{\alpha \in \mathbb{R}} \underbrace{\sum_{i=1}^n (x_i - \lambda_i(\alpha))^2}_{\equiv J(\alpha, x^n)}, \quad (3.13)$$

where  $\lambda_i(\alpha) = \tilde{F}g_i(\alpha) + \tilde{B}_i$ , and where  $\alpha$  is a generic variable representing the astrometric position, and  $g_i(\cdot)$  is given in (3.3). The solution to (3.13) does not have a closed-form

expression and, consequently, a number of iterative approaches have been adopted [37, 64] to solve or approximate  $\tau_{LS}(x^n)$ . Hence as  $\tau_{LS}(x^n)$  is implicit, it is not possible to compute its mean, its variance, nor its estimation error directly.

In this context, the problem of computing the mean square error (MSE) of an estimator that is the solution of an optimization problem has been recently addressed by [58] using a general framework. Their basic idea was to provide sufficient conditions on the objective function, in our case  $J(\alpha, x^n)$ , to derive a good approximation for  $\mathbb{E}_{X^n \sim f_{x_c}} \{(\tau_{LS}(X^n) - x_c)^2\}$ . Based on this idea, a refined result (specialized to our astrometry problem) is provided, which relaxes one of the idealized assumptions proposed in [58, Eq.(5)], and which is not strictly satisfied in our problem. As a consequence, this new result offers upper and lower bounds for the bias and MSE of  $\tau_{LS}(x^n)$ , respectively.

**Theorem 3.1** *Let be  $x_c \in \mathbb{R}$  a fixed and unknown parameter, and that  $X^n \sim f_{x_c}$ . In addition, defining the residual random variable<sup>5</sup>  $W(X^n, \alpha) \equiv \frac{J''(\alpha, X^n) - \mathbb{E}_{X^n \sim f_{x_c}} \{J''(\alpha, X^n)\}}{\mathbb{E}_{X^n \sim f_{x_c}} \{J''(\alpha, X^n)\}}$ . If there exists  $\delta \in (0, 1)$  such that  $\mathbb{P}(W(X^n, x_c) \in (-\delta, \delta)) = 1$ , then*

$$|\mathbb{E}_{X^n \sim f_{x_c}} \{\tau_{LS}(X^n)\} - x_c| \leq \epsilon(\delta), \quad (3.14)$$

$$\mathbb{E}_{X^n \sim f_{x_c}} \{(\tau_{LS}(X^n) - x_c)^2\} \in \left( \frac{\sigma_{LS}^2(n)}{(1 + \delta)^2}, \frac{\sigma_{LS}^2(n)}{(1 - \delta)^2} \right), \quad (3.15)$$

where

$$\sigma_{LS}^2(n) \equiv \frac{\mathbb{E}_{X^n \sim f_{x_c}} \{J'(x_c, X^n)^2\}}{(\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\})^2} \quad (3.16)$$

and

$$\epsilon(\delta) \equiv \frac{\mathbb{E}_{X^n \sim f_{x_c}} \{|J'(x_c, X^n)|\}}{\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\}} \cdot \frac{\delta}{1 - \delta}. \quad (3.17)$$

(The proof is presented in Appendix 5.6).

### 3.3.3 Analysis and Interpretation of Theorem 3.1

**Remark 1** Theorem 3.1 is obtained under a bounded condition (with probability one) over the random variable  $W(X^n, x_c)$ . Therefore, it is important to derive an explicit expression for  $W(X^n, x_c)$ . Adopting (3.13) it follows that:

$$J''(\alpha, x^n) = 2 \sum_{i=1}^n \left[ \left( \frac{d\lambda_i(\alpha)}{d\alpha} \right)^2 + (\lambda_i(\alpha) - x_i) \frac{d^2\lambda_i(\alpha)}{d^2\alpha} \right], \forall x^n,$$

and, consequently,  $\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\} = 2 \sum_{i=1}^n \left( \frac{d\lambda_i(\alpha)}{d\alpha} \Big|_{\alpha=x_c} \right)^2$ . Then

---

<sup>5</sup> As a short hand notation:  $J'(x_c, x^n) \equiv \frac{dJ(\alpha, x^n)}{d\alpha} \Big|_{\alpha=x_c}$ , and  $J''(x_c, x^n) \equiv \frac{d^2J(\alpha, x^n)}{d^2\alpha} \Big|_{\alpha=x_c}$ .

$$W(X^n, x_c) = \sum_{i=1}^n (\lambda_i(x_c) - X_i) \cdot \left[ \lambda_i''(x_c) / \sum_{j=1}^n (\lambda_j'(x_c))^2 \right]. \quad (3.18)$$

Then,  $W(X^n, x_c)$  is not bounded almost surely as  $X_i$  could take any value in  $\mathbb{N}$  with non-zero probability. However,  $\mathbb{E}_{X^n \sim f_{x_c}} \{W(X^n, x_c)\} = 0$  and its variance in closed-form is:

$$\text{Var}(W(X^n, x_c)) = \sum_{i=1}^n \lambda_i(x_c) \cdot \left[ (\lambda_i''(x_c))^2 / \left( \sum_{j=1}^n (\lambda_j'(x_c))^2 \right)^2 \right]. \quad (3.19)$$

From this, it can be evaluated how far is the bounded assumption of Theorem 3.1. To do this, we can resort to the *Markov's inequality* [65], where  $\mathbb{P}(W(X^n, x_c) \notin (-\rho, \rho)) \leq \text{Var}(W(X^n, x_c)) / \rho^2$ . Then for any  $\epsilon \in (0, 1)$ , a critical  $\delta(\epsilon) > 0$  can be characterized such that  $\mathbb{P}(W(X^n, x_c) \in (-\delta(\epsilon), \delta(\epsilon))) > 1 - \epsilon$ . Using this and Theorem 3.1, the conditional bias and conditional MSE of  $\tau_{LS}(x^n)$  can be bounded using (3.14) and (3.15), respectively. In Section 3.4, a numerical analysis is conducted, where it is shown that the bounded assumption for  $W(X^n, x_c)$  is indeed realistic for a number of important experimental settings in astrometry (with very high probability).

**Remark 2** Concerning the MSE of the LS estimator, Eq.(3.15) offers a lower and upper bound in terms of a *nominal value*  $\sigma_{LS}^2(n)$  (Eq.(3.16)) and an interval around it. In the interesting regime where  $\delta \ll 1$  (this regime approaches the ideal case  $\delta = 0$  studied by [58] in which case the variable  $W(X^n, x_c)$  becomes deterministic), we have that  $\tau_{LS}(\cdot)$  is an unbiased estimator, from (3.14), and, furthermore,

$$\text{Var}(\tau_{LS}(X^n)) = \mathbb{E}_{X^n \sim f_{x_c}} \{(\tau_{LS}(X^n) - x_c)^2\} = \sigma_{LS}^2(n) \geq \sigma_{CR}^2,$$

from (3.15) and (3.9). Then, it is interesting to provide an explicit expression for  $\sigma_{LS}^2(n)$ . First we have that,  $J'(x_c, x^n) = \sum_{i=1}^n 2 \frac{d\lambda_i(x_c)}{dx_c} (\lambda_i(x_c) - x_i)$ , and then

$$\begin{aligned} (J'(x_c, x^n))^2 = & 4 \left( \sum_{i=1}^n \sum_{j=1, j \neq i}^n (x_i x_j - x_i \lambda_j - x_j \lambda_i + \lambda_i \lambda_j) \frac{d\lambda_i}{dx_c} \frac{d\lambda_j}{dx_c} \right) \\ & + 4 \left( \sum_{i=1}^n (x_i^2 - 2x_i \lambda_i + \lambda_i^2) \left( \frac{d\lambda_i}{dx_c} \right)^2 \right). \end{aligned}$$

Therefore  $\mathbb{E}_{X^n \sim f_{x_c}} \{J'(x_c, X^n)^2\} = 4 \sum_{i=1}^n \lambda_i(x_c) \cdot (\lambda_i'(x_c))^2$ , which implies that:

$$\sigma_{LS}^2(n) = \frac{\sum_{i=1}^n \lambda_i(x_c) \cdot (\lambda_i'(x_c))^2}{\left( \sum_{i=1}^n (\lambda_i'(x_c))^2 \right)^2} = \frac{\sum_{i=1}^n (\tilde{F} g_i(x_c) + \tilde{B}_i) \cdot (g_i'(x_c))^2}{\left( \tilde{F} \sum_{i=1}^n (g_i'(x_c))^2 \right)^2}. \quad (3.20)$$

In the next section, it is provided a numerical analysis to compare (3.20) with the CR bound from (3.11). Also, it is analyzed if this nominal value is representative of the performance of the LS estimator.

**Remark 3** (Idealized Low SNR regime) Following the ideal scenario where  $\delta \ll 1$ , the weak signal case is explored, in which  $\tilde{F} g_i(x_c) \ll \tilde{B}_i$ , considering a constant background across the

pixels, i.e.,  $\tilde{B}_i = \tilde{B}$  for all  $i$ . Then adopting (3.20) we have that:

$$\sigma_{LS}^2(n) \approx \frac{\tilde{B}}{\tilde{F}^2 \sum_{i=1}^n (g'_i(x_c))^2}. \quad (3.21)$$

On the other hand, from (3.11) we have that  $\mathcal{I}_{x_c}(n) \approx \tilde{F}^2/\tilde{B} \sum_{i=1}^n (g'_i(x_c))^2$ . Remarkably in this context, the LS estimator is optimal and achieves the CR bound when a weak signal is observed. This results is consistent with the numerical simulations in [54, Table 3].<sup>6</sup>

**Remark 4** (Idealized High SNR regime) For the high SNR regime, assuming again that  $\delta \ll 1$ , it is considered the case where  $\tilde{F}g_i(x_c) \gg \tilde{B}_i$  for all  $i$ . In this case:

$$\sigma_{LS}^2(n) \approx \left[ \tilde{F} \frac{(\sum_{i=1}^n g'_i(x_c))^2}{\sum_{i=1}^n g_i(x_c)g'_i(x_c)^2} \right]^{-1} \quad \text{and} \quad \sigma_{CR}^2 \approx \left[ \tilde{F} \sum_{i=1}^n (g'_i(x_c))^2/g_i(x_c) \right]^{-1}. \quad (3.22)$$

Therefore, in this strong signal scenario, there is no match between the variance of the LS estimator and the CR bound, and consequently, we have that  $\sigma_{LS}^2(n) > \sigma_{CR}^2$ . To provide more insight on the nature of this performance gap, in the next proposition a closed-form expression is offered for this mismatch in the realistic high-resolution scenario where the source is oversampled, and the size of the pixel is a small fraction of the width parameter  $\sigma$  of the PSF in (3.1).

**Proposition 3** Assuming the idealized high SNR regime, if we have a Gaussian-like PSF and  $\Delta x/\sigma \ll 1$ , then

$$\frac{\sigma_{LS}^2(n)}{\sigma_{CR}^2} \approx \frac{8}{3\sqrt{3}} > 1. \quad (3.23)$$

(The proof is presented in Appendix 5.7).

This result shows that there is a very significant performance gap between the CR bound and the MSE of the LS estimator in the high SNR regime. This result should motivate the exploration of alternative estimators that could approximate more closely the CR bound in this regime.

---

<sup>6</sup>We note that this achievability result can be considered the astrometric counterpart of what has been shown in photometry by [56], where the LS estimator achieves the CR bound in the low SNR regime.

### 3.4 Numerical Analysis of the LS Estimator

In this section the implications of Theorem 3.1 in astrometry are explored, through the use of simulated observations. First, it is analyzed if the bounded condition over  $W(X^n, \alpha)$  adopted in Theorem 3.1 is a valid assumption for the type of settings considered in astronomical observations. After that, it is compared how efficient is the LS estimator proposed in (3.13) as a function of the SNR and pixel resolution, adopting for that purpose Theorem 3.1 and Proposition 1.

To perform the simulations, the problem is modeled adopting some realistic design variables as well as astronomical observing conditions [54, 55]. For the PSF, various analytical and semi-empirical forms have been proposed, see for instance the ground-based model in [66] and the space-based model in [67]. For the analysis, a Gaussian PSF is adopted, where  $\phi(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x)^2}{2\sigma^2}}$  in (3.3), and where  $\sigma$  is the width of the PSF, assumed to be known. This PSF has been found to be a good representation for typical astrometric-quality ground-based data [68]. In terms of nomenclature,  $FWHM \equiv 2\sqrt{2 \ln 2} \sigma$  measured in arcsec, denotes the *Full-Width at Half-Maximum* (FWHM) parameter, which is an overall indicator of the image quality at the observing site [69].

The background profile, represented by  $\{\tilde{B}_i, i = 1, \dots, n\}$ , is a function of several variables, like the wavelength of the observations, the moon phase (which contributes significantly to the diffuse sky background), the quality of the observing site, and the specifications of the instrument itself. It will be considered a uniform background across pixels underneath the PSF, i.e.,  $\tilde{B}_i = \tilde{B}$  for all  $i$ . To characterize the magnitude of  $\tilde{B}$ , it is important to first mention that the detector does not measure photon counts [ $e^-$ ] directly, but a discrete variable in “*Analog to Digital Units* (ADUs)” of the instrument, which is a linear proportion of the photon counts [61]. This linear proportion is characterized by the gain of the instrument  $G$  in units of [ $e^-/ADU$ ].  $G$  is just a scaling value, where we can define  $F \equiv \tilde{F}/G$  and  $B \equiv \tilde{B}/G$  as the intensity of the object and noise, respectively, in the specific ADUs of the instrument. Then, the background (in ADUs) depends on the pixel size  $\Delta x[arcsec]$  as follows [54]:

$$B = f_s \Delta x + \frac{D + RON^2}{G} [ADU], \quad (3.24)$$

where  $f_s$  is the (diffuse) sky background in [ $ADU/arcsec$ ], while  $D$  and  $RON$ , both measured in [ $e^-$ ] model the dark-current and read-out-noise of the detector on each pixel, respectively. Note that the first component in (3.24) is attributed to the site, and its effect is proportional to the pixel size. On the other hand, the second component is attributed to errors of the counting device (detector), and it is pixel-size independent. This distinction is central when analyzing the performance as a function of the pixel resolution of the array (see details in [54, Sec. 4]). More important is the fact that in typical ground-based astronomical observation, long exposure times are considered, which implies that the background is dominated by diffuse light coming from the sky (the first term in the RHS of (3.24)), and not from the detector [54, Sec. 4].

For the experimental conditions, the considered scenario corresponds to a ground-based station located at a good site with clear atmospheric conditions and the specifications of

current science-grade CCDs, where  $f_s = 1502.5[ADU/arcsec]$ ,  $D = 0$ ,  $RON = 5[e^-]$ ,  $FWHM = 1[arcsec]$  and  $G = 2[e^-/ADU]$  (with these values  $B = 313[ADU]$  for  $\Delta x = 0.2[arcsec]$  using Eq.(3.24)). In terms of scenarios of analysis, different pixel resolutions for the CCD array are explored  $\Delta x \in [0.1, 0.7]$  measured in arcsec, and different signal strengths  $\tilde{F} \in \{1080, 3224, 20004, 60160\}$ , measured in photon counts  $[e^-]$ <sup>7</sup>. Note that increasing  $\tilde{F}$  implies increasing the SNR of the problem, which can be approximately measured by the ratio  $\tilde{F}/\tilde{B}$ . On a given detector plus telescope setting, these different SNR scenarios can be obtained by changing appropriately the exposure time (open shutter) that generates the image.

### 3.4.1 Analyzing the Bounded Condition over $W(X^n, \alpha)$

To validate how realistic is the bounded assumption on  $W(X^n, \alpha)$  in the problem, we first evaluate the variance of  $W(X^n, x_c)$  in (3.19) presented in Fig. 3.1 for different SNR regimes and pixel resolutions in the array. Overall, the magnitudes are very small considering the admissible range  $(0, 1)$  for  $W(X^n, x_c)$  stipulated in Theorem 3.1. Also, given that  $W(X^n, x_c)$  has zero mean, the bounded condition happens with high probability. Complementing this, Fig. 3.2 presents the critical  $\delta$  across different pixel resolutions and SNR regimes<sup>8</sup>. For this, a small value of  $\epsilon (= 10^{-3})$  is fixed, and the  $\delta$  such that  $W(X^n, x_c) \in (-\delta, \delta)$  with probability  $1 - \epsilon$  is calculated. From the curves obtained, it can be said that the bounded assumption is holding (values of  $\delta$  in  $(0, 1)$ ) for a wide range of representative experimental conditions and, consequently, Theorem 3.1 can be used to provide a range on the performance of the LS estimator. Note that the idealized condition of  $\delta \approx 0$  is obtained only for the very high SNR regime (strong signals).

---

<sup>7</sup>These are the same values explored in [54, Table 3].

<sup>8</sup>These values were computed empirically (frequency counts) using 5,000 realizations of the random variable  $W(X^n, x_c)$  for the different SNR regimes and pixel resolutions.

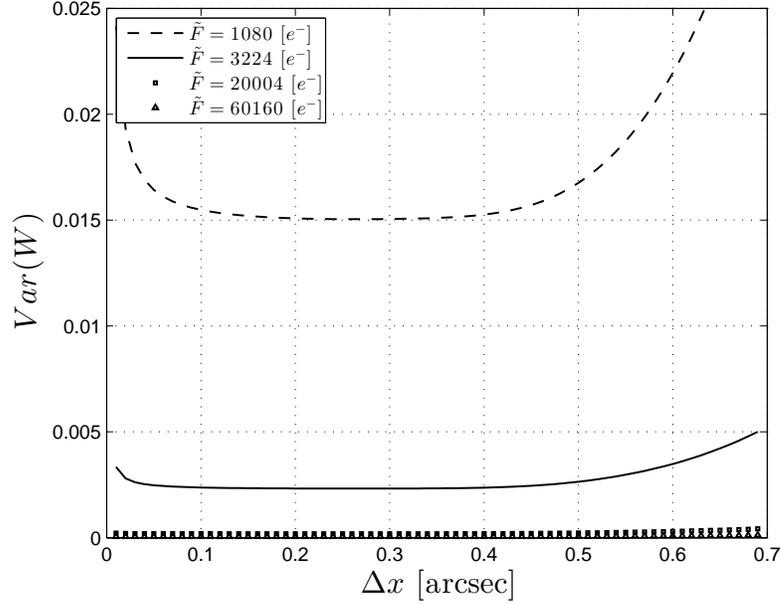


Figure 3.1: Variance of the residual random variable  $W(X^n, x_c)$  (dimensionless) as given by Eq.(3.19), as a function of the pixel resolution  $\Delta x[\text{arcsec}]$  for different realistic SNR scenarios (function of  $\tilde{F}$ ) encountered in ground-based astronomical observations. Since the admissible range for  $\delta$  is the interval  $(0,1)$ , the small computed values indicates that the bounded assumption in Theorem 1 can be considered as valid under these conditions.

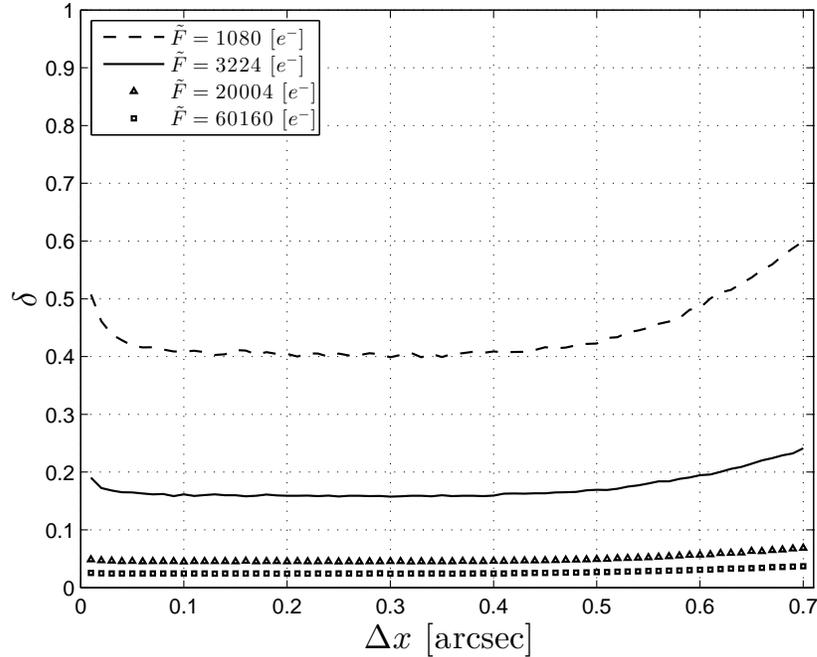


Figure 3.2: Numerical computation of the critical  $\delta$  (dimensionless) such that  $P(W(X^n, x_c) \in (-\delta, \delta)) > 1 - 10^{-3}$ . In all the scenarios (SNR, and  $\Delta x$ ), 5,000 realizations of the random variable  $W(X^n, x_c)$  are used to estimate the probability distribution for  $W(X^n, x_c)$ , and  $\delta$ , from frequency counts. As  $\delta$  decreases, we have a smaller bias (Eqs. (3.14) and (3.17)) and a narrower range for the MSE of the LS estimator (Eq.(3.15)).

### 3.4.2 Performance Analysis of the LS estimator

Adopting Eq.(3.15) it is provided an admissible range for the MSE performance of the LS estimator. For that the critical  $\delta$  obtained in Fig. 3.2 is used. These curves for the different SNR regimes and pixel resolutions are shown in Fig. 3.3. Following the trend reported in Fig. 3.2, the nominal value  $\sigma_{LS}^2$  is a precise indicator for the LS estimator performance for strong signals (matching the idealized conditions stated in Remark 4), while on the other hand, Theorem 3.1 can not indicate if  $\sigma_{LS}^2$  is accurate for low SNR, as we deviate from the idealized case elaborated in Remark 3. Nevertheless, it will be seen, based on some complementary empirical results reported in what follows, that even for low SNR, the nominal  $\sigma_{LS}^2$  predicts precisely the performance of the LS estimator.

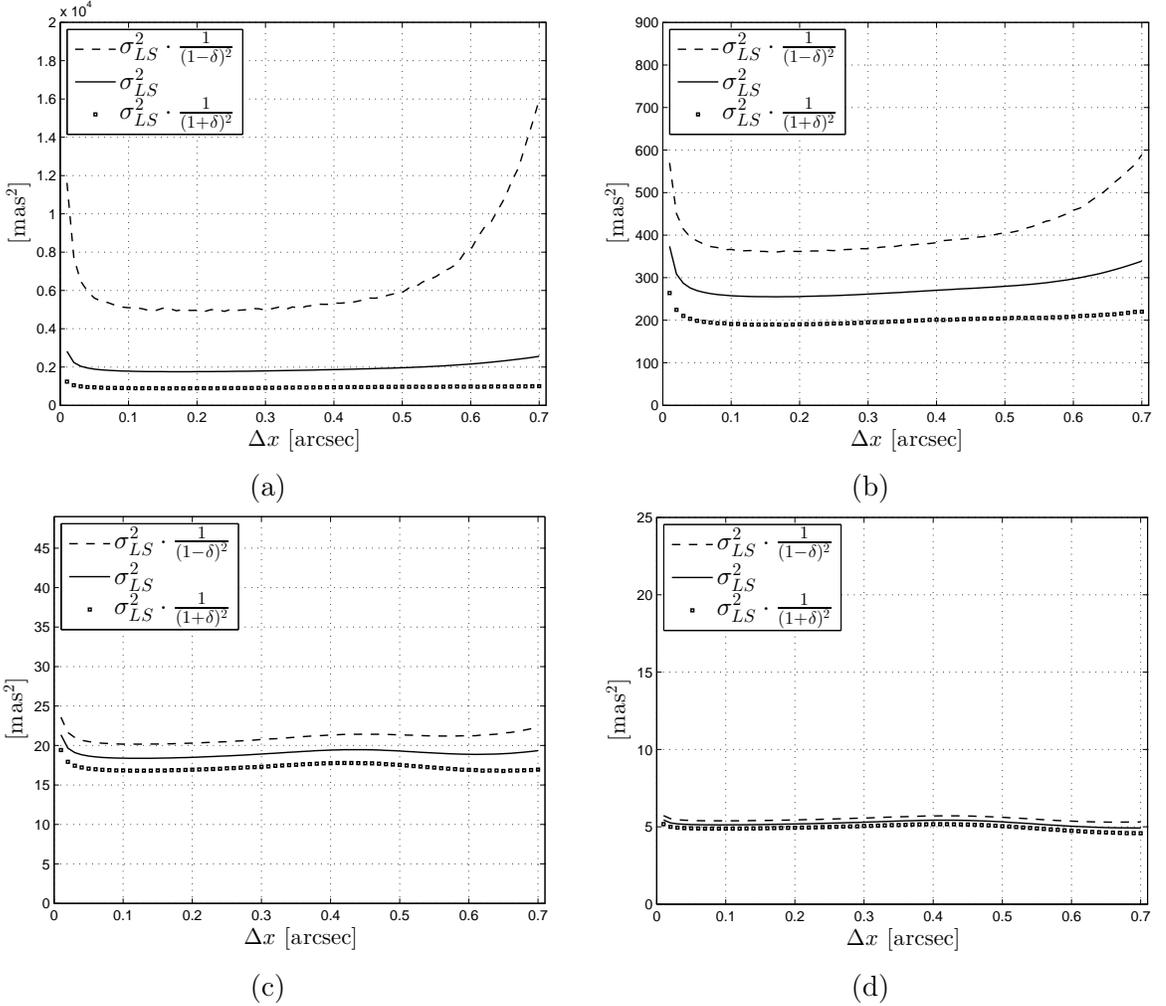


Figure 3.3: Range of  $MSE$  performance (in  $milli-arcsec^2$  ( $mas^2$ )) for the LS method in astrometry predicted by Theorem 3.1 (Eqs. (3.15) and (3.16)). Results are reported for different representative values of  $\tilde{F}$  and across different pixel sizes: (From top-left to bottom-right)  $\tilde{F} = 1080 [e^-]$ ;  $\tilde{F} = 3224 [e^-]$ ;  $\tilde{F} = 20004 [e^-]$ ;  $\tilde{F} = 60160 [e^-]$ .

Assuming for a moment the idealized case in which  $\delta \ll 1$ , the performance analysis can be reduced to measuring the gap between the nominal value predicted by Theorem 3.1 (Eq.(3.16)), and the CR bound in Proposition 1. Fig. 3.4 shows the relative difference given by  $e\% = 100 \cdot \frac{\sigma_{LS}^2 - \sigma_{CR}^2}{\sigma_{CR}^2}$ . From the figure it can be clearly seen that, in the low SNR regime, the relative performance differences tends to zero and, consequently, the LS estimator approaches the CR bound, and it is an efficient estimator. This matches what has been observed in Remark 3. On the other hand for high SNR, a performance gap that is non negligible is observed (up to  $\approx 27\%$  relative difference above the CR for  $\tilde{F} = 60160 [e^-]$ , and  $\approx 15\%$  above the CR for  $\tilde{F} = 20004 [e^-]$  for  $\Delta x = 0.2[arcsec]$ ). This is consistent with what has been observed in Remark 4. Note that in this regime, the idealized scenario in which  $\delta \ll 1$  is valid (see Fig. 3.2) and, thus,  $\mathbb{E}_{X^n \sim f_{x_c}} \{(\tau_{LS}(X^n) - x_c)^2\} \approx \sigma_{LS}^2$ , which is not strictly the case for the low SNR regime (although see Fig. 3.6). To refine the relative performance analysis presented in Fig. 3.4, Fig. 3.5 shows the feasible range (predicted by Theorem 3.1) of performance gap considering the critical  $\delta$  obtained in Fig. 3.2. Four cases are reported: from very low to very high SNR regimes to illustrate the trends. From this figure, it can be seen that the deviations from the nominal value are quite significant for the low SNR regime, and that, from this perspective, the range obtained from Theorem 3.1 is not sufficiently small to conclude about the goodness of the LS estimator in this context. On the other hand, in the high SNR regime, the nominal comparison can be considered quite precise.

The results of the previous paragraph motivate an empirical analysis to estimate the performance of the LS estimator empirically from the data, with the goal of resolving the low SNR regime illustrated in Fig. 3.5. For this purpose, 1,000 realizations were considered for all the SNR regimes and pixel sizes, and the performance of the LS estimator was computed using the empirical MSE. Large number of samples to guarantee convergence to the true MSE error were used, as a consequence of the law of large numbers [63]. Remarkably, it was observed in all cases that the estimated performance matches quite tightly the nominal  $\sigma_{LS}^2$  characterized by Theorem 3.1. This is illustrated in Fig. 3.6, which considers the most critical low SNR regime. Consequently, from this empirical analysis, the ambiguity present in the low SNR regime can be resolved, and it can be concluded that the comparison with the nominal result reported in Fig. 3.4, and the derived conclusion about the LS estimator in the low and high SNR regimes, can be considered valid. The simulations also show that the LS estimator is unbiased.

To conclude, this empirical evidence suggests that Theorem 3.1 could be improved, perhaps by imposing milder sufficient conditions, in order to prove that  $\sigma_{LS}^2$  is indeed a precise indicator of the MSE of the LS estimator at any SNR regime.

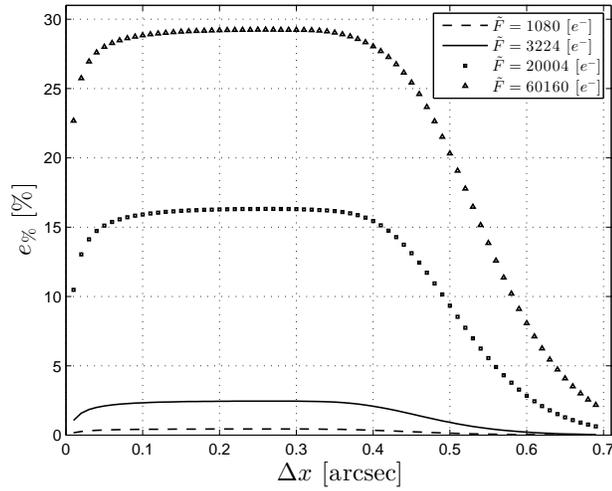


Figure 3.4: Relative performance differences between  $\sigma_{LS}^2$  in Theorem 3.1 (Eq.(3.16)) and the CR bound  $\sigma_{CR}^2$  in Proposition 1 (Eq.(3.11)). Results are reported for different SNR and pixel sizes. A significant performance gap between the LS technique and the CR bound is found for  $FWHM/\Delta x < 1$  (good sampling of the PSF) at high SNR, indicating that, in this regime, the LS method is sub-optimal, in agreement with Proposition 3 (see also Eq.(3.23)). This gap becomes monotonically smaller as the SNR decreases.

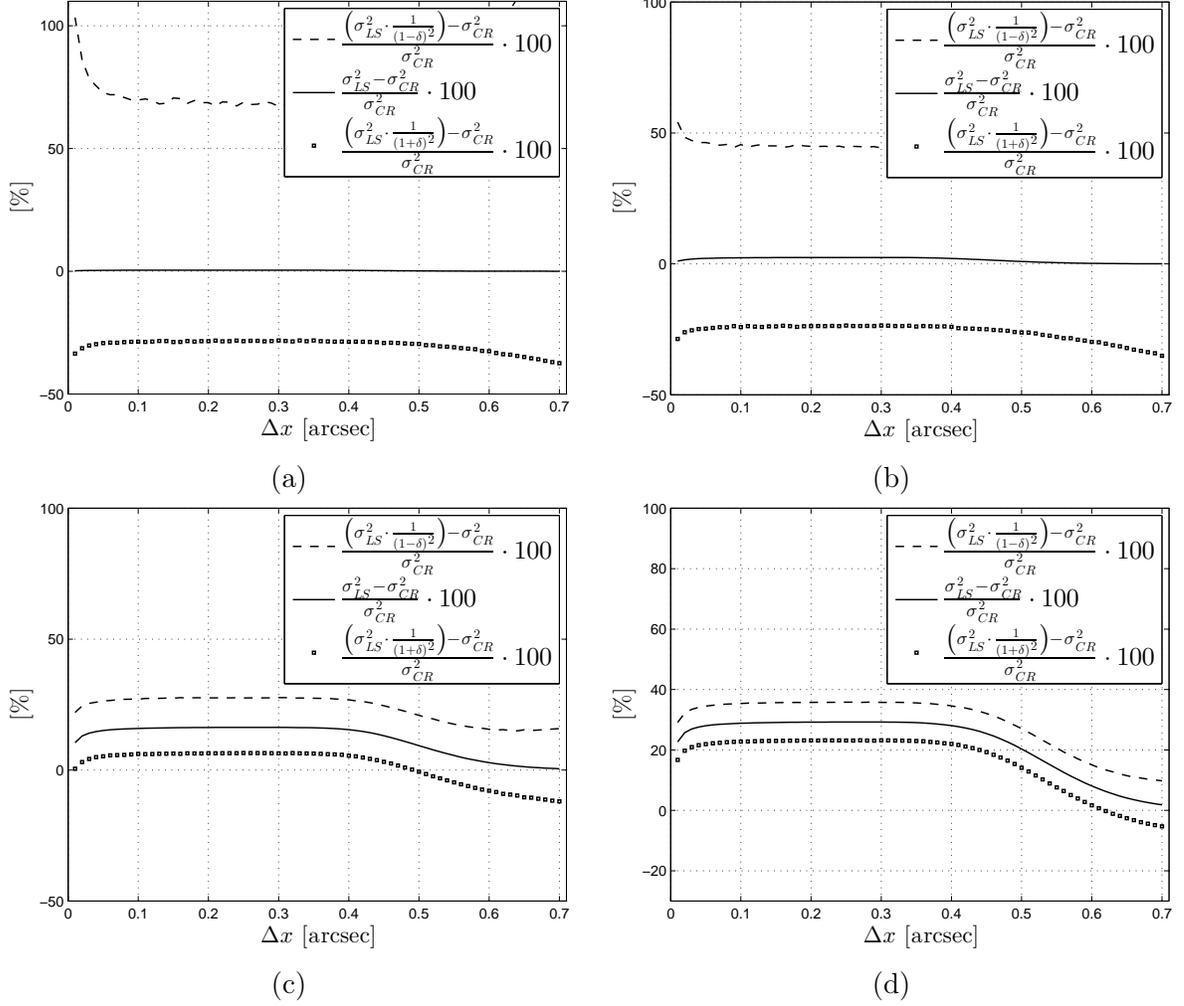


Figure 3.5: Relative performance differences between the range of performance  $\left(\frac{\sigma_{LS}^2(n)}{(1+\delta)^2}, \frac{\sigma_{LS}^2(n)}{(1-\delta)^2}\right)$  stipulated in Theorem 3.1 for the LS estimator (Eqs. (3.15) and (3.16)) and the CR bound  $\sigma_{CR}^2$  in Proposition 1 (Eq.(3.11)). Results are reported for different  $\tilde{F}$  and across different pixel sizes: (From top-left to bottom-right)  $\tilde{F} = 1080$  [e<sup>-</sup>];  $\tilde{F} = 3224$  [e<sup>-</sup>];  $\tilde{F} = 20004$  [e<sup>-</sup>];  $\tilde{F} = 60160$  [e<sup>-</sup>]. The 0% level corresponds to having achieved the CR bound. Note that as the SNR decreases, the bias (see Eqs. (3.14) and (3.17)), and the possible range for the MSE of the LS method (see Eq.(3.15)) increase (see also Fig. 3.2).

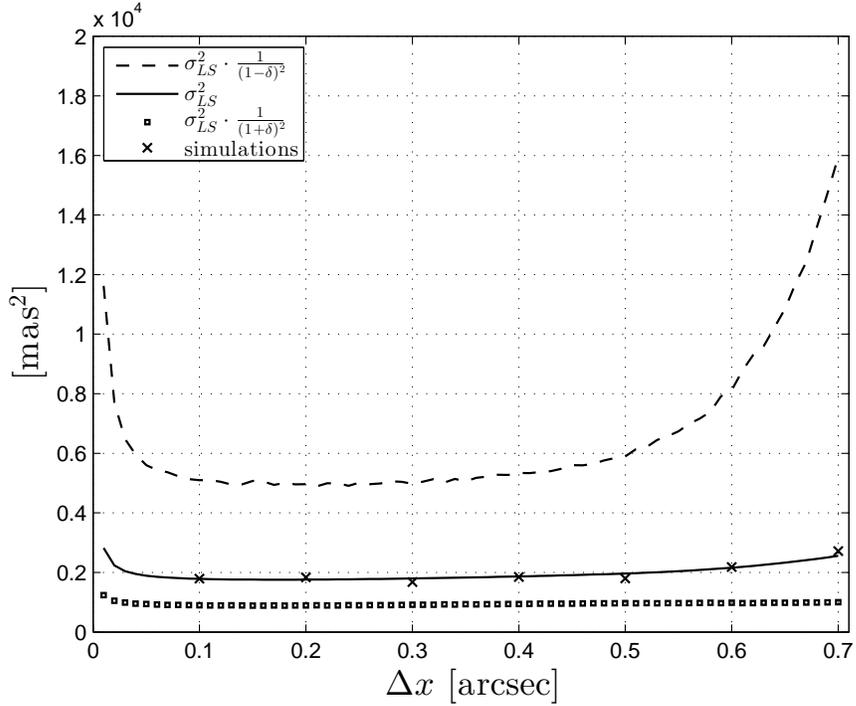


Figure 3.6: Comparison between the nominal value  $\sigma_{LS}^2$ , the performance range  $\left(\frac{\sigma_{LS}^2(n)}{(1+\delta)^2}, \frac{\sigma_{LS}^2(n)}{(1-\delta)^2}\right)$  stipulated in Theorem 3.1, and the empirical estimation of  $\mathbb{E}_{X^n \sim f_{x_c}}\{(\tau_{LS}(X^n) - x_c)^2\}$  from simulations for a low SNR regime of  $\tilde{F} = 1080$  [e<sup>-</sup>]. The fact that the simulations follow closely the nominal value, even at low SNR, justifies the use of  $\sigma_{LS}$  as given by Eqs.(3.16) and (3.20) as a benchmark of the LS method at any SNR.

# Chapter 4

## Conclusions

In this Thesis, signal processing techniques have been applied in two quite different problems: the characterization and automatic classification of natural rock textures and astrometry.

On the first problem studied in Chapter 2, a new set of discriminative features is proposed for detecting geologic rock textures based on the extraction of parsimonious texture and object information in transform based domains. In addition, a chain sequential classification approach is proposed, that, on the one hand, facilitates the design of isolated features, and on the other, provides a way to address the practical issue of limited data. The proposed scheme has the ability to discriminate pure rock textures as well as rocks with both objects and texture components, which is a challenging technical problem. Because of this ability, the proposed framework offers very good performance for the problem of automatic rock texture classification for 6 classes, showing significant improvements with respect to well-known state-of-the-art techniques used conventionally for texture classification (indexing) of nature images. This findings ratify how central for rock texture image classification is the development of specific features for this task, where it is shown that the new proposed features offer important gains with respect to standard approaches for texture analysis and classification.

On the second problem studied in Chapter 3, new results are provided to characterize the performance of the widely used Least-Squares (LS) estimator for the problem of astrometry, when observations are derived from digital discrete array detectors. The main result (Theorem 3.1) provides (in closed-form) a nominal value ( $\sigma_{LS}^2(n)$ ), and a range around it, for the mean square error (MSE) of the LS estimator as well as its bias. From the predicted nominal value, it is analyzed how efficient is the LS estimator in comparison with the minimum variance Cramér-Rao bound. In particular, it is shown that the LS estimator is efficient in the regime of low SNR (a point source with a weak signal), in the sense that it approximates very closely the Cramér-Rao bound. On the other hand, it is shown that at high SNR there is a significant gap in the performance of the LS estimator with respect to the minimum variance bound. It is believed that this sub-optimal behaviour is caused by the Poissonian nature of the detection process, in which the variance per pixel increases as the signal itself. Since the LS method is very sensitive to outliers, the large excursions caused by the large pixel intensity variance at high SNR make the LS method less efficient (from the point of view of

its MSE), than allowed by the Cramér-Rao bound. These performance analyses complement and match what has been observed in photometric estimation, where only in the low SNR regime the LS estimator has shown to achieve the Cramér-Rao bound. These fact should motivate the exploration of alternative estimators in this high SNR observational regime. However, an inspection of [54, Table 3] suggests that either a (Poisson variance-) weighted LS or a Maximum-Likelihood approach do not exhibit this loss of optimality at high-SNR, and should be preferred to the unweighted LS analyzed in this research effort.

# Chapter 5

## Appendix

### 5.1 The Wavelet and Shearlet Transforms

In this section we present two mathematical tools: the *Wavelet Transform (WT)* and the *Shearlet Transform (ST)*, which allow the analysis of an image at different scales, locations and orientations. We introduce briefly their mathematical background showing only the continuous case for the sake of brevity (for a detailed explanation of the discrete case please see [20, Sec. 7.7, pp. 347] and [28]).

#### 5.1.1 The Wavelet Transform

The WT corresponds to a decomposition of a 1-D function  $f$  by a family of functions  $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}^2}$ , which form an orthonormal basis of  $\mathbf{L}^2(\mathbb{R})$ , and provide a partition of the function spectrum in several scales [70, 20]. These functions are obtained through the dilation and translation of a kernel function  $\psi$ , called *mother wavelet*, as follows:

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), \quad (j, k) \in \mathbb{Z}^2. \quad (5.1)$$

The mother wavelet is constructed from a *scaling function*  $\phi$  which satisfies the two-scale difference equation [71]. The construction is given by:

$$\psi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} g(n)\phi(2x - n), \quad (5.2)$$

where  $g(n) = (-1)^n h(1 - n)$  and  $h$  is a filter that have to meet several conditions in order to obtain an adequate family  $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}^2}$  [71]. Thus, the decomposition of a function  $f$  is given by a set of *wavelet coefficients*  $\{d_j[k] \equiv \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{+\infty} f(x)\psi_{j,k}(x)dx : (j, k) \in \mathbb{Z}^2\}$ , where each coefficient captures the information that the function has between two subsequent resolution levels at a particular position. Fortunately, to calculate these coefficients it is not necessary to know explicitly  $\psi_{j,k}$ , but it is enough to have access to the filter  $h$ . In fact, given

a scale  $j$  it is possible to calculate the coefficients of the next scale  $j + 1$  by

$$d_{j+1}[k] = \sum_{n=-\infty}^{+\infty} g[n - 2k]a_j[n], \quad (5.3)$$

where

$$a_{j+1}[k] = \sum_{n=-\infty}^{+\infty} h[n - 2k]a_j[n]. \quad (5.4)$$

The coefficients  $\{a_j[k] \equiv \langle f, \phi_{j,k} \rangle = \int_{-\infty}^{+\infty} f(x)2^{-j/2}\phi(2^{-j}x - k)dx, (j, k) \in \mathbb{Z}^2\}$  are known as *approximate coefficients*, and capture the content that a function has at a particular scale level.

In the case of a 2-D function  $f(x)$  (like an image), where  $x = (x_1, x_2)$ , to keep an orthonormal decomposition there are three kinds of wavelet coefficients given by:

$$d_{j+1}^1[k] = \sum_{(n_1, n_2) \in \mathbb{Z}^2} h_{HH}(n_1 - 2k_1, n_2 - 2k_2)a_j[n_1, n_2], \quad (5.5)$$

$$d_{j+1}^2[k] = \sum_{(n_1, n_2) \in \mathbb{Z}^2} h_{HL}(n_1 - 2k_1, n_2 - 2k_2)a_j[n_1, n_2], \quad (5.6)$$

$$d_{j+1}^3[k] = \sum_{(n_1, n_2) \in \mathbb{Z}^2} h_{LH}(n_1 - 2k_1, n_2 - 2k_2)a_j[n_1, n_2], \quad (5.7)$$

where

$$a_{j+1}[k] = \sum_{(n_1, n_2) \in \mathbb{Z}^2} h_{LL}(n_1 - 2k_1, n_2 - 2k_2)a_j[n_1, n_2], \quad (5.8)$$

$k = (k_1, k_2) \in \mathbb{Z}^2$ ,  $h_{HH}(k) = g(k_1)g(k_2)$ ,  $h_{HL}(k) = g(k_1)h(k_2)$ ,  $h_{LH}(k) = h(k_1)g(k_2)$ , and  $h_{LL}(k) = h(k_1)h(k_2)$ . Further, the approximate coefficients are given by  $\{a_j[k] = \int_{-\infty}^{+\infty} f(x)2^{-j/2}\phi^2(2^{-j}x - k)dx, (j, k) \in \mathbb{Z}^3\}$ . It is important to mention that to use the algorithm given by Eqs. (5.5) to (5.8) the initial approximate coefficients  $a_0[\cdot]$  are needed. In our problem these coefficients are related to the resolution of the camera with which the images were obtained, and are given by a fraction of the digital image values (for more details see [20, Sec. 7.7, pp. 347]). Finally, just considering the parameter  $i \in \{1, 2, 3\}$  and the scale parameter  $j$ , we introduce the concept of *wavelet sub-band* corresponding to the set  $W_j^{(i)}(f) = \{d_j^{(i)}[k], k \in \mathbb{Z}^2\}$ .

## 5.1.2 The Shearlet Transform

The ST corresponds to a 2-D function decomposition by analyzing functions, so called *shearlets*, which are waveforms indexed by scales, orientations, and locations [28, 21].

The shearlet functions correspond to the set  $\{\psi_{j,l,k}^{(d)} : d = 0, 1, j \geq 0, -2^j \leq l \leq 2^j - 1, k \in \mathbb{Z}^2\}$ , where each element is obtained by the action of a family of operators on a single function  $\psi$  defined by the sum of two functions  $\psi^{(1)}$  and  $\psi^{(2)}$ , whose frequency supports are restricted to  $\mathcal{D}_0 = \{(w_1, w_2) : |w_1| \geq 1/8, |w_2/w_1| \leq 1\}$  and  $\mathcal{D}_1 = \{(w_1, w_2) : |w_2| \geq 1/8, |w_1/w_2| \leq 1\}$ , respectively (for more details please see [28, Sec. 2]).

The operations used on  $\psi$  correspond to: scaling operators, associated to matrices  $A_0$  and  $A_1$  called *anisotropic dilation matrices*; shearing operators, associated to matrices  $B_0$  and  $B_1$  called *shearing matrices*; and the translation operator. Thus, a shearlet at scale  $j$ , orientation  $l$ , and location  $k$  is given by:

$$\psi_{j,l,k}^{(d)}(x) = 2^{\frac{3j}{2}} \psi^{(d)}(B_d^l A_d^j x - k), \quad (5.9)$$

where  $A_0 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$ ,  $A_1 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$ ,  $B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ , and  $B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ .

It should be noted the similarity between shearlets and wavelets since both are derived from analogous processes and both give a complete description of  $\mathbf{L}^2(\mathbb{R}^2)$  (in the sense that wavelets form an orthonormal basis and shearlets form a Parseval frame of  $\mathbf{L}^2(\mathbb{R}^2)$  [28, Th. 2.1]). Nevertheless, shearlets allow a more refined analysis since they consider the orientation index ( $l$ ).

The *shearlet coefficients* for a 2-D function  $f$  are given by:

$$\{\mathcal{SH}^{(d)}(f)_{j,l,k} = \langle f, \psi_{j,l,k}^{(d)} \rangle : d = 0, 1, j \geq 0, -2^j \leq l \leq 2^j - 1, k \in \mathbb{Z}^2\}, \quad (5.10)$$

where  $\langle \cdot \rangle$  denotes the inner product in  $\mathbf{L}^2(\mathbb{R}^2)$ . The different orientations are subjected to what part of the support of  $\psi$  is being considered which is associated to the parameter  $d$ . For this reason, it is convenient for the notation to eliminate this parameter by the next re-labeling of the orientation parameter in the shearlet coefficient expression [30, Sec. 3]:

$$\mathcal{SH}(f)_{j,l,k} = \begin{cases} \mathcal{SH}^{(d)}(f)_{j,(l-1-2^j),k}, & 1 \leq l \leq 2^{j+1}, \\ \mathcal{SH}^{(d)}(f)_{j,(3 \cdot 2^j - l),k}, & 2^{j-1} < l \leq 2^{j+2}, \end{cases}$$

wherewith  $l \in L_j = \{1, \dots, 2^{j+2}\}$ . Finally, considering just the scale and orientation parameters, we introduce the concept of *shearlet sub-band* corresponding to the set  $\mathcal{SH}_{j,l}(f) = \{\mathcal{SH}_{j,l,k}(f), k \in \mathbb{Z}^2\}$ . More details in [21, 28].

## 5.2 Wavelet Statistical Texture Modeling

It is shown in [10] that a good statistical model for a natural texture is that the wavelet coefficients for a given sub-band follow a i.i.d distribution. More precisely, a suitable probability density function (PDF) for a specific wavelet sub-band  $W_j^{(i)}(z)$  is given by the *Generalized gaussian density (GGD)* [10]:

$$P(W_j^{(i)}(z) = w) = p(w; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|w|/\alpha)^\beta}, \quad (5.11)$$

where  $\Gamma(\cdot)$  is the Gamma function, and  $\alpha$  as well as  $\beta$  are real parameters that describe the form of the PDF.  $\alpha$  is called the *spread* parameter, related to the standard deviation, where a high value means a wide PDF. On the other hand,  $\beta$  is called the *shape* parameter and is related to the tail asymptotic behavior, where a small value implies a light tail. Both parameters can be estimated by a maximum-likelihood approach using the coefficients in  $W_j^{(i)}(z)$ . Thus, for instance, if a wavelet decomposition of 3 scale levels is applied to a stationary texture image, it is possible to assign to each wavelet sub-band the parameters  $\alpha$  and  $\beta$ , wherewith a model of 18 parameters is assigned to the image.

## 5.3 Texture-Cartoon Separation

This method is composed by two stages [22]. In the first one, the image  $z$  is filtered by a  $\sigma$ -sized low-pass filter  $L_\sigma$  forming  $f_\sigma = L_\sigma * z$ , where  $*$  denotes the convolution operator. For each pixel  $x$  the value  $\Delta LTV(x) = |\nabla z_\sigma(x) - \nabla z(x)|$  is calculated, which corresponds to the change of the local total variation (LTV) after the filtering process. If  $\Delta LTV(x)$  has a high value, the cartoon image pixel  $z_u(x)$  will be equal to the filtered image pixel  $z_\sigma(x)$ , and if it does not,  $z_u(x) = z(x)$ . It should be noted that to have a good separation it is necessary to fix  $\sigma$  adequately. For this task, the following orthogonality measure between two images  $z_1$  and  $z_2$  is used:

$$ADE(z_1, z_2) = \left| \arcsin \left( \frac{\langle \bar{z}_1, \bar{z}_2 \rangle}{\| \bar{z}_1 \|_{\ell_2} \cdot \| \bar{z}_2 \|_{\ell_2}} \right) \right|, \quad (5.12)$$

where  $\bar{z}_1$  and  $\| \cdot \|_{\ell_2}$  denote the vector formed by the concatenation of columns of  $z_1$  and the  $\ell_2$ -norm, respectively.

Next, an interval  $[s_1, s_2]$  of possible  $\sigma$ -values is fixed, and the image  $z$  is divided in  $B$  blocks. Each block is separated with the previous process for each  $\sigma_i \in [s_1, s_2]$  forming  $z_{u,\sigma_i}^b$  and  $z_{v,\sigma_i}^b$ . Next, the blocks  $z_u^b$ ,  $b \in \{1, \dots, B\}$ , that compose the cartoon image are given by:

$$z_u^b = z_{u,\sigma_m}^b, \quad (5.13)$$

where

$$\sigma_m = \arg \min_{\sigma_i \in [s_1, s_2]} (ADE(z_{u,\sigma_i}^b, z_{v,\sigma_i}^b)). \quad (5.14)$$

Finally, the cartoon and texture images for this first stage are denoted by  $z_u^I$  and  $z_v^I$ , respectively, and it is possible to start the second stage. In this part, a new cartoon image denoted

by  $z_u^{II}$  is created from  $z_u^I$ . The process begins with a partition of  $z_u^{II}$  in  $B$  blocks, and then for each  $b \in \{1, \dots, B\}$  the following diffusion equation is applied [72] :

$$z_u^{II,b}(x, t + 1) = z_u^{II,b}(x, t) + \frac{\lambda}{|\eta(x)|} \sum_{(x') \in \eta(x')} \nabla^{(x')} \times (g(\nabla^{(x')} z_u^{I,b}(x))) \nabla^{(x')} z_u^{II}(x, t), \quad (5.15)$$

where  $\lambda \in \mathbb{R}$  is an scalar constant that determines the rate of diffusion,  $|\eta(x)|$  is the number of elements of the spatial neighborhood of pixel  $x$ ,  $g$  is a weigh function that allows not blurring cartoon edges and  $\nabla^{(x')} z_u^{II}(x, t)$  is an approximation of the image gradient at a particular direction given by:

$$\nabla^{(x')} z_u^{II}(x, t) = z_u^{II}(x', t) - z_u^{II}(x, t). \quad (5.16)$$

The last step is finding for each block the iterations  $t_{ADE}$  in which the separation carried out in (5.15) is optimum, for which is used the orthogonality measure given in (5.12). Thus, the final blocks  $u_{II}^b$ ,  $b \in \{1, \dots, B\}$ , of the second stage are calculated as follows:

$$z_u^{II,b}(x) = z_u^{II,b}(x, t_{ADE}), \quad (5.17)$$

where

$$t_{ADE} = \arg \min_{i \in \{1, \dots, I_{max}\}} (ADE(z_u^{II,b}(x, t_i), z_v^{II,b}(x, t_i))), \quad (5.18)$$

and  $I_{max}$  is a previously fixed number of diffusion iterations. Finally, denoting by  $z_u$  and  $z_v$  the final cartoon and texture images respectively, we have that:

$$z_u = z_u^{II}, \quad (5.19)$$

$$z_v = z - z_u. \quad (5.20)$$

## 5.4 Source Separation Technique

For this method it is supposed that the image model has the form:

$$z = z_{S_1} + z_{S_2}, \quad (5.21)$$

where  $z_{S_1}$  and  $z_{S_2}$  are two information sources which are sparse in different transform domains [23]. An important point to highlight is that each of these sources should not have a good sparsity in the domain in which the other source is sparse. Denoting by  $\Phi_1$  and  $\Phi_2$  the synthesis matrices associated to the domains in which the sources are sparse, the separation process is carried out by solving the following optimization problem [26]:

$$(\hat{z}_{S_1}, \hat{z}_{S_2}) = \arg \min_{z_{S_1}, z_{S_2}} \| \Phi_1^T z_{S_1} \|_{\ell_1} + \| \Phi_2^T z_{S_2} \|_{\ell_1} + \lambda \| z - z_{S_1} - z_{S_2} \|_{\ell_2}^2, \quad (5.22)$$

where  $\lambda \in \mathbb{R}$  and  $\| z - z_{S_1} - z_{S_2} \|_{\ell_2}^2$  correspond to a regularization parameter and a term that penalizes a possible noise component in the image, respectively.

## 5.5 Proof of Proposition 2

PROOF. The proof is based on the well-known fact that the CR bound is achieved by an unbiased estimator, if and only if, the following decomposition holds [73, pp. 12]

$$\frac{d \ln L(x^n; \theta)}{d\theta} = A(\theta, n) \cdot (f(x^n) - \theta), \quad (5.23)$$

where  $A(\theta, n)$  is a function of  $\theta$  and  $n$  (i.e., it does not depend on the data), while  $f(x^n)$  is a function of the data exclusively (i.e., it does not depend on the parameter). Furthermore, if the achievability condition in (5.23) is satisfied, then  $A(\theta, n) = \mathcal{I}_\theta(n)$ , and  $f(x^n)$  is an unbiased estimator of  $\theta$  that achieves the CR bound.

The proof follows by contradiction, assuming that (5.23) holds. First using (3.5), we have that:

$$\frac{d \ln L(x^n; x_c)}{dx_c} = \sum_{i=1}^n \left[ \frac{x_i}{\lambda_i(x_c)} \frac{d\lambda_i(x_c)}{dx_c} - \frac{d\lambda_i(x_c)}{dx_c} \right] = \sum_{i=1}^n \frac{x_i}{\lambda_i(x_c)} \frac{d\lambda_i(x_c)}{dx_c}, \quad (5.24)$$

the last equality comes from the fact that  $\sum_{i=1}^n g_i(x_c) = 1$  from the assumption in (3.4). Then adopting (5.24) and (3.11) in (5.23),

$$f(x^n) = \sum_{i=1}^n \frac{\frac{x_i}{\tilde{F}g_i(x_c) + \tilde{B}_i} \frac{dg_i(x_c)}{dx_c}}{\tilde{F} \sum_{i=1}^n \frac{1}{\tilde{F}g_i(x_c) + \tilde{B}_i} \left( \frac{dg_i(x_c)}{dx_c} \right)^2} + x_c, \quad (5.25)$$

which contradicts the assumption that  $f(x^n)$  should be a function of the data alone. Furthermore, if we consider the extreme high SNR regime, where  $\tilde{F}g_i(x_c) \gg \tilde{B}_i$  for all  $i$ , and the low SNR regime, where  $\tilde{F}g_i(x_c) \ll \tilde{B}_i$  for all  $i$ , it follows that,

$$f(x^n) = \sum_{i=1}^n x_i \frac{d \ln g_i(x_c)}{dx_c} \cdot \left[ \tilde{F} \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 / g_i(x_c) \right]^{-1} + x_c \quad (5.26)$$

and

$$f(x^n) = \sum_{i=1}^n \frac{x_i}{\tilde{B}_i} \frac{dg_i(x_c)}{dx_c} \cdot \left[ \tilde{F} \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 / \tilde{B}_i \right]^{-1} + x_c, \quad (5.27)$$

respectively. Therefore a contradiction remains even in these extreme SNR regimes.

□

## 5.6 Proof of Theorem 3.1

PROOF. The approach of [58] uses the fact that the objective function  $J(\alpha, x^n)$  in (3.13) is two times differentiable, which is satisfied in our context. As a short-hand, if we denote by  $\hat{x}_c$  the LS estimator solution  $\tau_{LS}(x^n)$ , then the first order necessary condition for a local optimum requires that  $J'(\hat{x}_c, x^n) \equiv \frac{dJ(\alpha, x^n)}{d\alpha}|_{\alpha=\hat{x}_c} = 0$ . The other key assumption in [58] is that  $\hat{x}_c$  is in a close neighborhood of the true value  $x_c$ . In our case, this has to do with the quality of the pixel-based data used for the inference, and it is assumed it offers a good estimation of the position [38, 43]. Then using a first order Taylor expansion of  $J'(\hat{x}_c, x^n)$  around  $x_c$ , the following key approximation can be adopted [58, Eq.(4)]:

$$0 = J'(\hat{x}_c, x^n) = J'(x_c, x^n) + (\hat{x}_c - x_c)J''(x_c, x^n) \Leftrightarrow (x_c - \hat{x}_c) = \frac{J'(x_c, x^n)}{J''(x_c, x^n)}, \quad (5.28)$$

where  $J''(x_c, x^n) \equiv \frac{d^2J(\alpha, x^n)}{d^2\alpha}|_{\alpha=x_c}$ . If we consider  $X^n \sim f_{x_c}$ , then from (5.28),

$$x_c - \tau_{LS}(X^n) = \frac{J'(x_c, X^n)}{J''(x_c, X^n)}. \quad (5.29)$$

The second step in the approximation proposed by [58] is to bound  $\frac{J'(x_c, X^n)}{J''(x_c, X^n)}$  by  $\frac{J'(x_c, X^n)}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}}$ . For that the residual variable  $W(X^n, \alpha)$  is introduced, where  $J''(x_c, X^n) = \mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}(1+W(X^n, x_c))$ . Using the fact that  $W(X^n, \alpha)$  is bounded almost surely (see Remark 1 and Section 3.4.1),

$$\begin{aligned} \left| \frac{J'(x_c, X^n)}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} - \frac{J'(x_c, X^n)}{J''(x_c, X^n)} \right| &= \left| \frac{J'(x_c, X^n)}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} \cdot \left[ 1 - \frac{1}{1+W(X^n, x_c)} \right] \right| \\ &\leq \left| \frac{J'(x_c, X^n)}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} \right| \max_{w \in (-\delta, \delta)} \left| 1 - \frac{1}{1+w} \right| \\ &\leq \frac{|J'(x_c, X^n)|}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} \cdot \frac{\delta}{1-\delta}, \end{aligned} \quad (5.30)$$

the last step uses the fact that  $\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\} \geq 0$  (see Remark 1). On the other hand, by *Jensen's inequality* [65],

$$\begin{aligned} &\left| \frac{\mathbb{E}_{X^n \sim f_{x_c}}\{J'(x_c, X^n)\}}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} - \mathbb{E}_{X^n \sim f_{x_c}}\left\{ \frac{J'(x_c, X^n)}{J''(x_c, X^n)} \right\} \right| \\ &\leq \mathbb{E}_{X^n \sim f_{x_c}} \left\{ \left| \frac{J'(x_c, X^n)}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} - \frac{J'(x_c, X^n)}{J''(x_c, X^n)} \right| \right\} \\ &\leq \frac{\mathbb{E}_{X^n \sim f_{x_c}}\{|J'(x_c, X^n)|\}}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} \cdot \frac{\delta}{1-\delta}, \end{aligned} \quad (5.31)$$

where the last inequality comes from (5.30). Then it is used that  $J'(x_c, X^n) = 2 \sum_{i=1}^n (\lambda_i(x_c) - X_i) \frac{d\lambda_i(x_c)}{dx_c}$ , and consequently  $\mathbb{E}_{X^n \sim f_{x_c}}\{J'(x_c, X^n)\} = 0$ . Then from (5.31) and (5.29), we have that

$$|x_c - \mathbb{E}_{X^n \sim f_{x_c}}\{\tau_{LS}(X^n)\}| \leq \frac{\mathbb{E}_{X^n \sim f_{x_c}}\{|J'(x_c, X^n)|\}}{\mathbb{E}_{X^n \sim f_{x_c}}\{J''(x_c, X^n)\}} \cdot \frac{\delta}{1-\delta}, \quad (5.32)$$

which provides (3.14).

Concerning the MSE, from the hypothesis on  $W(X^n, x_c)$  we have that:

$$\begin{aligned} \frac{J'(x_c, X^n)^2}{\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\}^2 (1 + \delta)^2} &\leq \left( \frac{J'(x_c, X^n)}{J''(x_c, X^n)} \right)^2 \\ &\leq \frac{J'(x_c, X^n)^2}{\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\}^2 (1 - \delta)^2}, \end{aligned} \quad (5.33)$$

almost surely. Then taking the expected value in (5.33) and using (5.29) for the central term, it follows that

$$\begin{aligned} \frac{\mathbb{E}_{X^n \sim f_{x_c}} \{J'(x_c, X^n)^2\}}{\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\}^2 (1 + \delta)^2} &\leq \mathbb{E}_{X^n \sim f_{x_c}} \{(x_c - \tau_{LS}(X^n))^2\} \\ &\leq \frac{\mathbb{E}_{X^n \sim f_{x_c}} \{J'(x_c, X^n)^2\}}{\mathbb{E}_{X^n \sim f_{x_c}} \{J''(x_c, X^n)\}^2 (1 - \delta)^2}, \end{aligned} \quad (5.34)$$

which concludes the result.  $\square$

## 5.7 Proof of Proposition 3

PROOF. Recalling from Eq.(3.3) that  $g_i(x_c) = \int_{y_i - \frac{\Delta y}{2}}^{y_i + \frac{\Delta y}{2}} \phi(y - x_c, \sigma) dy$  and assuming a Gaussian PSF of the form  $\phi(y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-y^2}{2\sigma^2}\right)$  (see Section 3.4) by the mean value theorem and the hypothesis of small pixel ( $\Delta x/\sigma \ll 1$ ), it is possible to state that

$$g_i(x_c) \approx \phi(x_i - x_c, \sigma) \Delta x, \quad (5.35)$$

and then we have that

$$\begin{aligned} \frac{dg_i(x_c)}{dx_c} &\approx \frac{(x_i - x_c)}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(\frac{-(x_i - x_c)^2}{2\sigma^2}\right) \Delta x \\ &= \frac{(x_i - x_c)}{\sigma^2} \phi(x_i - x_c, \sigma) \Delta x. \end{aligned} \quad (5.36)$$

With the above approximation we have that

$$\begin{aligned} \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 g_i(x_c) &\approx \sum_{i=1}^n \left( \frac{d\phi(x_i - x_c, \sigma)}{dx_c} \Delta x \right)^2 \phi(x_i - x_c, \sigma) \Delta x \\ &= \sum_{i=1}^n \left( \frac{(x_i - x_c)}{\sigma^2} \phi(x_i - x_c, \sigma) \Delta x \right)^2 \phi(x_i - x_c, \sigma) \Delta x \\ &\approx \frac{\Delta x^2}{2\sqrt{3}\pi\sigma^6} \sum_{i=1}^n (x_i - x_c)^2 \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{3}}} \exp\left(\frac{-(x_i - x_c)^2}{2\left(\frac{\sigma}{\sqrt{3}}\right)^2}\right) \Delta x \end{aligned} \quad (5.37)$$

The term inside the summation in (5.37) can be approximated by an integral due to the small-pixel hypothesis and assuming that the source is well sampled by the detector (see Section 3.2.1, Eq.(3.4)) we can obtain that:

$$\sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 g_i(x_c) \approx \frac{\Delta x^2}{2\sqrt{3}\pi\sigma^6} \int_{-\infty}^{+\infty} (x - x_c)^2 \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{3}}} \exp\left( \frac{-(x - x_c)^2}{2 \left( \frac{\sigma}{\sqrt{3}} \right)^2} \right) dx \quad (5.38)$$

$$= \frac{\Delta x^2}{2\sqrt{3}\pi\sigma^6} \cdot \frac{\sigma^2}{3} = \frac{\Delta x^2}{6\sqrt{3}\pi\sigma^4}, \quad (5.39)$$

where (5.39) follows from the fact that the term inside the integral in (5.38) corresponds to the second moment of a normal random variable of mean  $x_c$  and variance  $\frac{\sigma^2}{3}$ .

By the same set of arguments used to approximate  $\sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 g_i(x_c)$  in (5.39), we have that:

$$\begin{aligned} \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 &\approx \sum_{i=1}^n \left( \frac{d\phi(x_i - x_c, \sigma)}{dx_c} \Delta x \right)^2 \\ &= \sum_{i=1}^n \left( \frac{(x_i - x_c)}{\sigma^2} \phi(x_i - x_c, \sigma) \Delta x \right)^2 \\ &\approx \frac{\Delta x}{2\sqrt{\pi}\sigma^5} \sum_{i=1}^n (x_i - x_c)^2 \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{2}}} \exp\left( \frac{-(x_i - x_c)^2}{2 \left( \frac{\sigma}{\sqrt{2}} \right)^2} \right) \Delta x \\ &\approx \frac{\Delta x}{2\sqrt{\pi}\sigma^5} \int_{-\infty}^{+\infty} (x - x_c)^2 \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{2}}} \exp\left( \frac{-(x - x_c)^2}{2 \left( \frac{\sigma}{\sqrt{2}} \right)^2} \right) dx \quad (5.40) \end{aligned}$$

$$\approx \frac{\Delta x}{2\sqrt{\pi}\sigma^5} \cdot \frac{\sigma^2}{2} = \frac{\Delta x}{4\sqrt{\pi}\sigma^3}, \quad (5.41)$$

where (5.41) follows from the fact that the term inside the integral in (5.40) is the second moment of a normal random variable of mean  $x_c$  and variance  $\frac{\sigma^2}{2}$ . Finally, for  $\sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 \frac{1}{g_i(x_c)}$  we proceed again in the same way.

$$\begin{aligned} \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 \frac{1}{g_i(x_c)} &\approx \sum_{i=1}^n \left( \frac{d\phi(x_i - x_c, \sigma)}{dx_c} \Delta x \right)^2 \frac{1}{\phi(x_i - x_c, \sigma) \Delta x} \\ &= \sum_{i=1}^n \left( \frac{(x_i - x_c)}{\sigma^2} \phi(x_i - x_c, \sigma) \Delta x \right)^2 \frac{1}{\phi(x_i - x_c, \sigma) \Delta x} \\ &\approx \frac{1}{\sigma^4} \sum_{i=1}^n (x_i - x_c)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left( \frac{-(x_i - x_c)^2}{2\sigma^2} \right) \Delta x \\ &\approx \frac{1}{\sigma^4} \int_{-\infty}^{+\infty} \frac{(x - x_c)^2}{\sqrt{2\pi}\sigma} \exp\left( \frac{-(x - x_c)^2}{2\sigma^2} \right) dx \quad (5.42) \end{aligned}$$

$$\approx \frac{1}{\sigma^4} \cdot \sigma^2 = \frac{1}{\sigma^2}, \quad (5.43)$$

where (5.43) follows from the fact that the term inside the integral in (5.42) corresponds to the second moment of a normal random variable of mean  $x_c$  and variance  $\sigma^2$ .

Then adopting (5.39), (5.41) and (5.43) in (3.20) and (3.11), respectively, we have that:

$$\begin{aligned}
\sigma_{LS}^2 &= \frac{\sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 g_i(x_c)}{\tilde{F} \left( \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 \right)^2} + \frac{\tilde{B}}{\tilde{F}^2 \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2} \\
&\approx \frac{\frac{\Delta x^2}{6\sqrt{3}\pi\sigma^4}}{\tilde{F} \left( \frac{\Delta x}{4\sqrt{\pi}\sigma^3} \right)^2} + \frac{\tilde{B}}{\tilde{F}^2 \frac{\Delta x}{4\sqrt{\pi}\sigma^3}} \\
&= \frac{\sigma^2}{\tilde{F}} \left[ \frac{8}{3\sqrt{3}} + 4\sqrt{\pi} \frac{1}{\sigma} \frac{\tilde{B}}{\tilde{F}} \right] \approx \frac{\sigma^2}{\tilde{F}} \frac{8}{3\sqrt{3}} \tag{5.44}
\end{aligned}$$

and

$$\sigma_{CR}^2 = \frac{1}{\sum_{i=1}^n \frac{\left( \tilde{F} \frac{dg_i(x_c)}{dx_c} \right)^2}{\tilde{F} g_i(x_c) + \tilde{B}}} = \frac{1}{\tilde{F} \sum_{i=1}^n \left( \frac{dg_i(x_c)}{dx_c} \right)^2 \frac{1}{g_i(x_c)}} \approx \frac{\sigma^2}{\tilde{F}}. \tag{5.45}$$

In the last step in (5.44) and the last step in (5.45), it is used the assumption that  $\tilde{F} \gg \tilde{B}$ .  $\square$

# Bibliography

- [1] A. Mock and D. Jerram, “Crystal size distributions (csd) in three dimensions: insights from the 3d reconstruction of a highly porphyritic rhyolite,” *Journal of Petrology*, vol. 46, no. 8, pp. 1525–1541, 2005.
- [2] M. D. Higgins, *Quantitative textural measurements in igneous and metamorphic petrology*, vol. 276. Cambridge University Press Cambridge, 2006.
- [3] M. D. Higgins and D. Chandrasekharam, “Nature of sub-volcanic magma chambers, deccan province, india: Evidence from quantitative textural analysis of plagioclase megacrysts in the giant plagioclase basalts,” *Journal of Petrology*, vol. 48, no. 5, pp. 885–900, 2007.
- [4] L. Vink, *Textures of the Hilton North deposit, Queensland, Australia, and their relationship to liberation*. PhD thesis, University of Queensland, 1997.
- [5] D. Bojcevski, “Metallurgical characterisation of george fisher meso-textures and micro-textures,” Master’s thesis, University of Queensland, 2004.
- [6] A. Croop and W. Goodall, “The influence of rock texture on mineral processing,” *MinAssist Australia*, 2013.
- [7] S. Dunham and J. Vann Proceedings Project Evaluation 2007.
- [8] S. Dunham, J. Vann, and S. Coward, “Beyond geometallurgy - gaining competitive advantage by exploiting the broad view of geometallurgy,” 1rst Ausimm International Geometallurgy Conference, September 2011.
- [9] N. K. Bonnici, *The mineralogical and textural characteristics of copper-gold deposits linked to mineral processing attributes*. PhD thesis, University of Tasmania, 2012.
- [10] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized gaussian densities and Kullback-Leibler distance,” *IEEE Transactions on Image Processing*, vol. 11, pp. 146–158, February 2002.
- [11] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

- [12] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval," *Image Processing, IEEE Transactions on*, vol. 22, no. 7, pp. 2545–2558, 2013.
- [13] M. Petrou and P. G. Sevilla, *Image processing: dealing with texture*, vol. 10. Wiley Chichester, 2006.
- [14] N. Vasconcelos, "Minimum probability of error image retrieval," *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, pp. 2322–2336, 2004.
- [15] E. J. Tarbuck, F. K. Lutgens, and D. Tasa, *Earth science*. Prentice Hall, 2000.
- [16] M. Jungmann, M. Kopal, C. Clauser, and T. Berlage, "Multi-class supervised classification of electrical borehole wall images using texture features," *Computers & Geosciences*, vol. 37, no. 4, pp. 541–553, 2011.
- [17] E. Donskoi, S. Suthers, S. Fradd, J. Young, J. Campbell, T. Raynlyn, and J. Clout, "Utilization of optical image analysis and automatic texture classification for iron ore particle characterisation," *Minerals Engineering*, vol. 20, no. 5, pp. 461–471, 2007.
- [18] S. Chatterjee, "Vision-based rock-type classification of limestone using multi-class support vector machine," *Applied intelligence*, vol. 39, no. 1, pp. 14–27, 2013.
- [19] L. Lepistö, I. Kunttu, and A. Visa, "Color-based classification of natural rock images using classifier combinations," in *Image Analysis*, pp. 901–909, Springer, 2005.
- [20] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [21] G. Kutyniok *et al.*, *Shearlets: Multiscale analysis for multivariate data*. Springer, 2012.
- [22] D. Szolgay and T. Szirányi, "Adaptive image decomposition into cartoon and texture parts optimized by the orthogonality criterion," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3405–3415, 2012.
- [23] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *Image Processing, IEEE Transactions on*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [24] J. F. Silva and S. S. Narayanan, "On signal representations within the bayes decision framework," *Pattern Recognition*, vol. 45, no. 5, pp. 1853–1865, 2012.
- [25] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [26] J. M. Fadili, J.-L. Starck, M. Elad, D. Donoho, *et al.*, "Mcalab: Reproducible research in signal and image decomposition and inpainting," *IEEE Computing in Science and Engineering*, vol. 12, no. 1, pp. 44–63, 2010.
- [27] G. Kutyniok and W.-Q. Lim, "Image separation using wavelets and shearlets," in *Curves*

*and Surfaces*, pp. 416–430, Springer, 2012.

- [28] G. Easley, D. Labate, and W.-Q. Lim, “Sparse directional image representations using the discrete shearlet transform,” *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.
- [29] X. Gibert, V. M. Patel, D. Labate, and R. Chellappa, “Discrete shearlet transform on gpu with applications in anomaly detection and denoising,” *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–14, 2014.
- [30] S. Yi, D. Labate, G. R. Easley, and H. Krim, “A shearlet approach to edge analysis and detection,” 2009.
- [31] E. Høg, “400 years of astrometry: from tycho brahe to hipparcos,” *Experimental Astronomy*, vol. 25, no. 1-3, pp. 225–240, 2009.
- [32] S. Reffert, “Astrometric measurement techniques,” *New Astronomy Reviews*, vol. 53, no. 11, pp. 329–335, 2009.
- [33] E. Høg, “Astrometry lost and regained,” *Baltic Astronomy*, vol. 20, pp. 221–230, 2011.
- [34] W. F. van Altena, *Astrometry for Astrophysics: Methods, Models, and Applications*. Cambridge University Press, 2013.
- [35] M. Lattanzi, “Astrometric cosmology,” *Memorie della Societa Astronomica Italiana*, vol. 83, p. 1033, 2012.
- [36] S. B. Howell, *Handbook of CCD astronomy*, vol. 5. Cambridge University Press, 2006.
- [37] P. B. Stetson, “Daophot: A computer program for crowded-field stellar photometry,” *Publications of the Astronomical Society of the Pacific*, pp. 191–222, 1987.
- [38] I. R. King, “Accuracy of measurement of star images on a pixel array,” *Publications of the Astronomical Society of the Pacific*, pp. 163–168, 1983.
- [39] C. Alard and R. H. Lupton, “A method for optimal image subtraction,” *The Astrophysical Journal*, vol. 503, no. 1, p. 325, 1998.
- [40] R. K. Honeycutt, “Ccd ensemble photometry on an inhomogeneous set of exposures,” *Publications of the Astronomical Society of the Pacific*, pp. 435–440, 1992.
- [41] A. C. Cameron, D. Pollacco, R. Street, T. Lister, R. West, D. Wilson, F. Pont, D. Christian, W. Clarkson, B. Enoch, *et al.*, “A fast hybrid algorithm for exoplanetary transit searches,” *Monthly Notices of the Royal Astronomical Society*, vol. 373, no. 2, pp. 799–810, 2006.
- [42] J.-F. Lee and W. Van Altena, “Theoretical studies of the effects of grain noise on photographic stellar astrometry and photometry,” *The Astronomical Journal*, vol. 88, pp. 1683–1689, 1983.

- [43] R. C. Stone, “A comparison of digital centering algorithms,” *The Astronomical Journal*, vol. 97, pp. 1227–1237, Apr. 1989.
- [44] W. Van Altena and L. Auer, “Digital image centering, i,” in *Image Processing Techniques in Astronomy*, pp. 411–418, Springer, 1975.
- [45] L. Auer and W. Van Altena, “Digital image centering, ii,” *The Astronomical Journal*, vol. 83, pp. 531–537, 1978.
- [46] L. Lindegren, “Photoelectric astrometry—a comparison of methods for precise image location,” in *IAU Colloq. 48: Modern Astrometry*, vol. 1, pp. 197–217, 1978.
- [47] P. Jakobsen, P. Greenfield, and R. Jędrzejewski, “The cramer-rao lower bound and stellar photometry with aberrated hst images,” *Astronomy and astrophysics*, vol. 253, no. 1, pp. 329–332, 1992.
- [48] T. Zucchetto, R. Gonsalves, S. Ebstein, and P. Nisenson, “Estimating the cramer-rao bound for restored astronomical observations,” *The Astrophysical Journal*, vol. 439, pp. L43–L45, 1995.
- [49] H.-M. Adorf, “Limits to the precision of joint flux and position measurements on array data,” in *Astronomical Data Analysis Software and Systems V*, vol. 101, p. 13, 1996.
- [50] U. Bastian, “The maximum reachable astrometric precision - The Cramer-Rao Limit.” 2004BASNOCODE, April 2004.
- [51] C. Radhakrishna Rao, “Information and accuracy attainable in the estimation of statistical parameters,” *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.
- [52] H. Cramér, “A contribution to the theory of statistical estimation,” *Scandinavian Actuarial Journal*, vol. 1946, no. 1, pp. 85–94, 1946.
- [53] L. Lindegren, “High-accuracy positioning: astrometry,” *ISSI Scientific Reports Series*, vol. 9, pp. 279–291, 2010.
- [54] R. A. Méndez, J. F. Silva, and R. Lobos, “Analysis and interpretation of the cramér-rao lower-bound in astrometry: One-dimensional case,” *Publications of the Astronomical Society of the Pacific*, vol. 125, no. 927, pp. pp. 580–594, 2013.
- [55] R. A. Méndez, J. F. Silva, R. Orostica, and R. Lobos, “Analysis of the cramér-rao bound in the joint estimation of astrometry and photometry,” *Publications of the Astronomical Society of the Pacific*, vol. 126, no. 942, pp. 798–810, 2014.
- [56] M. Perryman, P. Jakobsen, L. Colina, G. Lelievre, F. Macchetto, J. Nieto, and S. Serego Alighieri, “An improved technique for the search for optical emission from radio pulsars, and its application to psr 0301+ 19, 1919+ 21 and 2303+ 30,” *Astronomy and Astrophysics*, vol. 215, pp. 195–210, 1989.

- [57] K. A. Winick, “Cramér-rao lower bounds on the performance of charge-coupled-device optical position estimators,” *J. Opt. Soc. Am. A*, vol. 3, pp. 1809–1815, Nov 1986.
- [58] H. So, Y. Chan, K. Ho, and Y. Chen, “Simple formulae for bias and mean square error computation [dsp tips and tricks],” *Signal Processing Magazine, IEEE*, vol. 30, no. 4, pp. 162–165, 2013.
- [59] J. A. Fessler, “Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography,” *Image Processing, IEEE Transactions on*, vol. 5, no. 3, pp. 493–506, 1996.
- [60] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, “Position calibration of microphones and loudspeakers in distributed computing platforms,” *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 1, pp. 70–83, 2005.
- [61] R. L. Gilliland, “Details of Noise Sources and Reduction Processes,” in *Astronomical CCD Observing and Reduction Techniques* (S. B. Howell, ed.), vol. 23 of *Astronomical Society of the Pacific Conference Series*, p. 68, 1992.
- [62] J. A. Tyson, “Low-light-level charge-coupled device imaging in astronomy,” *Journal of the Optical Society of America A*, vol. 3, pp. 2131–2138, Dec. 1986.
- [63] R. Gray and L. D. Davisson, *Introduction to Statistical Signal Processing*. Cambridge Univ Press, 2004.
- [64] K. J. Mighell, “Stellar photometry and astrometry with discrete point spread functions,” *Monthly Notices of the Royal Astronomical Society*, vol. 361, no. 3, pp. 861–878, 2005.
- [65] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Interscience, New York, second ed., 2006.
- [66] I. R. King, “The profile of a star image,” *Publications of the Astronomical Society of the Pacific*, pp. 199–201, 1971.
- [67] O. Bendinelli, G. Parmeggiani, A. Piccioni, and F. Zavatti, “The newton-gauss regularized method-application to point-spread-function determination in ccd frames,” *The Astronomical Journal*, vol. 94, pp. 1095–1100, 1987.
- [68] R. A. Méndez, E. Costa, M. H. Pedreros, M. Moyano, M. Altmann, and C. Gallart, “Proper motions of local group dwarf spheroidal galaxies i: First ground-based results for fornax,” *Publications of the Astronomical Society of the Pacific*, vol. 122, no. 893, pp. 853–875, 2010.
- [69] F. R. Chromey, *To measure the sky: an introduction to observational astronomy*. Cambridge University Press, 2010.
- [70] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.

- [71] G. Strang, “Wavelets and dilation equations: A brief introduction,” *SIAM review*, vol. 31, no. 4, pp. 614–627, 1989.
- [72] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 7, pp. 629–639, 1990.
- [73] M. Kendall, A. Stuart, J. Ord, and S. Arnold, *Vol. 2A: Classical inference and the linear model*. London [etc.]: Arnold [etc.], 1999.