

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA  
PhD COURSE: ELECTRONICS, COMPUTER SCIENCE AND TELECOMMUNICATIONS  
XXIII CYCLE - SCIENTIFIC DISCIPLINARY SECTOR: ING-INF/01

---

# Analysis and Modeling Techniques for Ultrasonic Tissue Characterization

**Simona Maggio**

SUPERVISOR  
Professor Guido Masetti  
COORDINATOR  
Professor Paola Mello

---

DEIS - DEPARTMENT OF ELECTRONICS, COMPUTER SCIENCE AND SYSTEMS  
JANUARY 2008 - DECEMBER 2010



### Abstract

This thesis introduces new processing techniques for computer-aided interpretation of ultrasound images with the purpose of supporting medical diagnostic. In terms of practical application, the goal of this work is the improvement of current prostate biopsy protocols by providing physicians with a visual map overlaid over ultrasound images marking regions potentially affected by disease. As far as analysis techniques are concerned, the main contributions of this work to the state-of-the-art is the introduction of deconvolution as a pre-processing step in the standard ultrasonic tissue characterization procedure to improve the diagnostic significance of ultrasonic features.

This thesis also includes some innovations in ultrasound modeling, in particular the employment of a continuous-time autoregressive moving-average (CARMA) model for ultrasound signals, a new maximum-likelihood CARMA estimator based on exponential splines and the definition of CARMA parameters as new ultrasonic features able to capture scatterers concentration.

Finally, concerning the clinical usefulness of the developed techniques, the main contribution of this research is showing, through a study based on medical ground truth, that a reduction in the number of sampled cores in standard prostate biopsy is possible, preserving the same diagnostic power of the current clinical protocol.

**Keywords:** ultrasound, tissue characterization, deconvolution, prostate biopsy, continuous-time autoregressive moving average, exponential splines.

## Résumé

Cette thèse introduit de nouvelles techniques de traitement pour l'interprétation guidée d'images ultrasons, dans le but de soutenir le diagnostic médical. L'objectif pratique de ce travail est l'amélioration du protocole standard pour la biopsie de la prostate, en fournissant au médecin une carte visuelle sur l'échographie qui marque les régions potentiellement malignes. En ce qui concerne les techniques d'analyse, la contribution principale de cette thèse est l'introduction de la déconvolution comme étape de pré-traitement dans la procédure standard de caractérisation des tissus par ultrasons, afin d'améliorer la valeur diagnostique des caractéristiques du signal.

Cette thèse présente, en outre, des innovations dans la modélisation du signal ultrason, en particulier la proposition d'un modèle autorégressif à moyenne glissante en temps continu (CARMA), le développement d'un nouvel estimateur au maximum de vraisemblance de paramètres CARMA fondé sur les splines exponentielles et la définition des paramètres CARMA comme des nouvelles caractéristiques du signal ultrason capables de capturer l'information sur la concentration de scatterers.

Enfin, concernant l'utilité clinique des techniques développées, la contribution principale de cette recherche est dans la démonstration, au travers d'une étude médicale, de la possibilité de réduire concrètement le nombre d'échantillons prélevés pendant la biopsie de la prostate standard, en préservant le même pouvoir diagnostique du protocole biopsique standard.

### Sommario

Questa tesi introduce nuove tecniche di elaborazione per l'interpretazione guidata di immagini ultrasoniche, allo scopo di supportare la diagnosi medica. Dal punto di vista applicativo, l'obiettivo di questo lavoro è il miglioramento del protocollo standard per la biopsia prostatica, fornendo al radiologo una mappa visiva di regioni potenzialmente malate. Per quanto riguarda le tecniche di analisi, il contributo principale di questa tesi è l'introduzione della deconvoluzione come passo di pre-processing nella procedura standard di tissue characterization basata su ultrasuoni, al fine di migliorare il potere diagnostico di alcune caratteristiche estratte dal segnale.

Questa tesi include, inoltre, delle innovazioni nel modeling del segnale ultrasonico, in particolare la proposta di un modello tempo-continuo autoregressive moving-average (CARMA), lo sviluppo di un nuovo stimatore maximum-likelihood di parametri CARMA basato su exponential spline e la definizione dei parametri CARMA come nuove caratteristiche del segnale ultrasonico capaci di catturare l'informazione sulla concentrazione di scatterer.

Infine, per quanto riguarda l'utilità applicativa delle tecniche sviluppate, il contributo principale di questa ricerca sta nel mostrare attraverso uno studio su ground-truth medico che è possibile ridurre effettivamente il numero di campioni prelevati durante la biopsia prostatica standard, preservando le stesse capacità diagnostiche del protocollo clinico standard.



---

# CONTENTS

---

<b>Abstract</b>	<b>i</b>
<b>Résumé</b>	<b>ii</b>
<b>Sommario</b>	<b>iii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Background</b>	<b>5</b>
1.1 Diagnostic ultrasound . . . . .	6
1.2 Tissue characterization . . . . .	10
1.3 Application: prostate TRUS . . . . .	19
<b>2 Ultrasound image analysis and interpretation</b>	<b>27</b>
2.1 Dimensionality reduction . . . . .	29
2.2 Feature-based segmentation . . . . .	32
2.3 Deconvolution as pre-processing step . . . . .	36
2.4 Feature-based classification . . . . .	39
2.5 Ground truth database . . . . .	40
2.6 Experimental results . . . . .	43
2.7 Conclusion . . . . .	51
<b>3 A CARMA model for ultrasound signals</b>	<b>53</b>
3.1 CAR and CARMA models . . . . .	55
3.2 CARMA processes and exponential splines . . . . .	58
3.3 Insights on the exponential spline framework . . . . .	68
3.4 A new approach to CARMA model identification . . . . .	70
3.5 CARMA estimation algorithm . . . . .	79
3.6 Simulation results . . . . .	86
3.7 CARMA for ultrasound tissue characterization . . . . .	92
3.8 Comparative study on phantom images . . . . .	95
3.9 Conclusion . . . . .	99

---

<b>4</b>	<b>Improving standard prostate biopsy protocol</b>	<b>101</b>
4.1	Database collection . . . . .	102
4.2	Processing and learning scheme . . . . .	106
4.3	Classification results . . . . .	111
4.4	The future . . . . .	111
4.5	Conclusion . . . . .	113
	<b>Conclusions</b>	<b>115</b>
<b>A</b>	<b>Features for ultrasonic tissue typing</b>	<b>119</b>
<b>B</b>	<b>State of the art in tissue characterization</b>	<b>123</b>
<b>C</b>	<b>Initial conditions in CARMA estimation</b>	<b>125</b>
<b>D</b>	<b>Non symmetric exponential B-splines</b>	<b>127</b>
	<b>Notations</b>	<b>131</b>
	<b>Publications</b>	<b>137</b>
	<b>Bibliography</b>	<b>139</b>



---

# INTRODUCTION

---

*Research is what I'm doing when I don't know  
what I'm doing.*

---

Wernher von Braun (1912-1977)

ULTRASOUND imaging is one of the most used technologies for non invasive medical diagnostic. Resolution limits and speckle noise affecting ultrasonography prevent the employment of standard image processing techniques to perform tissue characterization supported by computer. On the contrary the interpretation of ultrasound image requires an approach based on features specific of the echo signal, able to capture the state of the imaged tissue.

The realization of computer-aided tissue characterization schemes in ultrasound imaging is constrained to the analysis of characteristics of the ultrasound signal and their correlation to the pathological state of tissues. In addition, the study of new parametric models for the ultrasonic echo allows the definition of new image attributes which may be likewise correlated to the state of the investigated tissue and useful for characterization. Finally, the design of a medical ground truth and a learning scheme to recognize pathological tissues is essential for an unbiased approach to tissue characterization.

In this thesis all the fundamental steps for the realization of ultrasonic tissue characterization schemes are taken into account and dealt with through the development of new processing methods. Although the developed techniques are applicable to the investigation of any biological tissue, the final goal of this work is the design of a computer-aided detection scheme to support physicians in the diagnosis of prostate cancer.

This thesis is subdivided in three stages. First, a preliminary study of the existing ultrasonic features is performed to select the most significant attributes for tissue representation. The ground truth used in this first stage is a collection of ultrasound images

of the prostate gland, where the pathological region is outlined by pathologists. Then, a non linear learning approach is validated on the collected medical ground truth. Within the first stage the novel contribution of this thesis is the introduction of shift-variant deconvolution as a pre-processing step in the traditional tissue characterization scheme. In this context deconvolution breaks away from its traditional goal of improving the visual quality of ultrasound images and is used to enhance the diagnostic power of ultrasonic features. Experimental results show that performing deconvolution before feature extraction yields to significantly increase the ability of some ultrasonic features to recognize pathological tissues.

In the second stage, a new model is proposed for the ultrasound echo signal, in order to enable the definition of novel attributes useful for tissue characterization. The ultrasound signal is modelled as a sampled continuous-time autoregressive moving average (CARMA) process. Two new algorithms are proposed for the estimation of continuous-time domain parameters of a CARMA process from its sampled data. The first is a maximum-likelihood estimator based on exponential splines suitable for any sampling condition, even when some aliasing is present. Simulations results show that this estimator outperforms current estimation methods based on polynomial splines, which are appropriate only in high sampling frequency conditions.

Additionally, another estimator is designed to provide good estimates with a lower computational cost when high sampling frequency is employed: this is usually the case in ultrasound images. Although not maximum-likelihood, this estimator is still based on exponential splines and can be used to compute the CARMA parameters of the ultrasound signals composing an image. CARMA parameters are considered as new ultrasonic features and their ability in tissue representation are analysed on a phantom ground-truth and compared to traditional ARMA parameters, often used to model the discrete ultrasound echo. The results of this study show that CARMA parameters are able to capture the changes in scatterers concentration better than traditional ARMA parameters. In the second stage the main contributions of this thesis are (a) the proposal of a new model for ultrasound signal, (b) the design of a CARMA estimator providing correct estimates also in aliasing conditions and (c) the definition of new features correlated to the scatterers concentration in ultrasound images.

The third stage of this work focuses on the application of the

studied techniques to the improvement of the current prostate biopsy protocol. A large ground truth database of imaged biopsy cores and the corresponding histo-pathological analysis are collected to validate the accuracy of processing and classification techniques. Improved statistical and textural features are used here in a specific learning scheme, able to use both completely known data and uncertain biopsy cores, where pathology covers just a percentage of the sample.

The developed algorithm for real-time computer-aided biopsy (rtCAB) includes feature extraction and learning techniques and is implemented exploiting CUDA™ parallel processing in order to enable real-time support to physicians during biopsy sessions. This technique is able to reduce the biopsy core number from 8-12 to 7, preserving the same diagnostic power of a standard protocol and working in real-time.

The structure of this thesis is formed by 5 chapters. The first chapter develops the background resuming the basic information about ultrasound imaging and tissue characterization. Chapter 2 describes the proposed characterization scheme with introduction of deconvolution as pre-processing step. Chapter 3 contains the description of the new continuous-time domain model for ultrasound signals; the derivation of CARMA parameters estimators and the application of CARMA parameters to the characterization of ultrasound scatterers concentration. The fourth chapter reports the realization of the ground truth and the developed learning algorithm for the improvement of the current prostate biopsy protocol. The conclusive chapter summarizes the main contributions of this thesis and provides some clues about future research in ultrasonic tissue characterization.



---

## BACKGROUND

---

*The essential is invisible to the eyes.*

---

Antoine de Saint-Exupéry (1900-1944)

THE main goal of biomedical image analysis is monitoring inner physiological systems to early diagnose potential diseases. A pathological process affecting the investigated system reflects on measured signals that, because of the disease, present characteristics different from the corresponding normal pattern. The essence of image analysis is to capture these characteristics, which are often not visible to human eye, and to describe the state of the imaged physiological system. Image modeling can also serve this purpose, providing some mathematical descriptors of the measured signals that result to be modified by the state of the imaged tissues and, thus, useful for their characterization.

The interpretation of biomedical images, traditionally performed by physicians, can be supported by automatic image analysis tools, making a quantitative study of physiological systems possible. This process is called Computer-aided Diagnosis (CAD) [1] and it is aimed at improving the reliability of physician judgement by reducing subjectivity, inter- and intra-observer variability and potential human errors due to fatigue, boredom or environmental factors.

CAD tools have always been thought as a support and not a replacement of medical specialists, who integrate their experience and intuition with objective clinical signs to realize the final diagnosis. Biomedical image analysis and enhancement techniques constitute the core of CAD systems and can be crucial in health-care quality.

Although the main target of medical image analysis and improvement techniques is early detection of potential diseases, these

post-processing steps are also important to boost the performance of cost-effective imaging devices. In fact the software enhancement of medical imagery is necessary both to compensate for quality reduction in video acquisitions of portable devices and to answer to the recent needs of world healthcare market, interested in producing affordable medical equipments for developing countries.

Among all imaging modalities, ultrasound presents some interesting advantages: ultrasound devices are cheap, they provide real-time feedbacks, they don't use ionizing radiations, they are a well established technology present in all health institutions. On the other hand, ultrasound video sequences are affected by low resolution and are characterised by speckle noise, preventing the employment of traditional image processing algorithms to perform computer-based analysis and enhancement.

Indeed, while some CAD tools are already clinically used in more expensive imaging modalities, like CT and MRI, the realization of effective image analysis techniques in ultrasound, matching the clinical standards and possibly real-time, is still a challenge.

## 1.1 Diagnostic ultrasound

Ultrasound echo imaging [2] is the fastest and simplest medical imaging technique, based on recording the reflected echo of ultrasound pulses. The main information provided by ultrasonography is the echo intensity value associated to some spatial coordinates.

Unlike CT and MRI, ultrasound does not provide directly any information about some tissue parameters, in fact the echo intensity depends in a complex way both on the imaged tissue and on the acquisition system. For this reason ultrasound are used with the only purpose of imaging the anatomy, or guiding some clinical operations, but not to characterize tissues.

Nevertheless the echo intensity is just a small part of the information the reflected pulse can provide. The raw radio-frequency (RF) signal returned by reflection is neglected by the traditional B-scan imaging modality, which directly computes its envelope (figure 1.1). Nevertheless the study of this signal can lead to the definition of significant features of the imaged tissues, as well as provide a way to reduce system dependent effects. The RF signal is, thus, the key to perform ultrasonic tissue characterization.

The reflection of an ultrasound pulse generates a single data

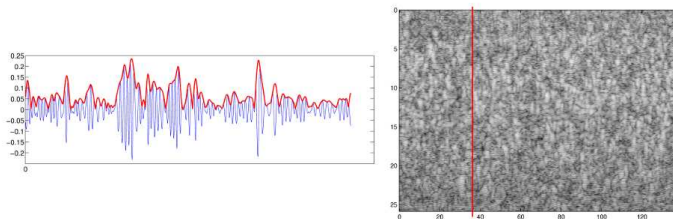


Figure 1.1: Example of an RF signal and its envelope shown as a scan line in a phantom ultrasound image.

line in ultrasound image along the propagation direction, called axial direction. The resolution along this direction depends on both the length of the ultrasonic pulse and its wavelength, still it can only be increased within strict limits, since ultrasound absorption in tissues is linearly proportional to frequency. The poor quality of ultrasound image is due to the intrinsic resolution limits of this technology.

The lateral resolution depends on the geometry of the probe generating the ultrasound beam. The beam divergence angle is smaller for larger diameters of the probe, making small transducer suitable for near focusing and large probes suitable for greater depth focusing. A curved probe surface, an acoustic lens or a delayed stimulation of the transducer elements gives the possibility to tune the beam divergence and radial direction.

The geometrical arrangement of several acquisition lines along different directions produces 2D and 3D ultrasound data. For this reason most algorithms in ultrasound imaging are developed as 1D processing techniques, and then applied to each image or volume scan line.

The fan scan is one of the most used ultrasound rays configurations and it is advantageous when the organ to be investigated is protected by structures unpenetrable by ultrasound and requires a small interface area of the transducer. A fan scan transducer is used, for example, to image the prostate by trans-rectal ultrasound (TRUS). The main disadvantage of the fan scan probe is the reduction of lateral resolution along the axial depth. Furthermore the echo data are recorded in polar  $(r, \theta)$ -coordinates and need a conversion into Cartesian coordinates and interpolation before dis-

playing on monitors.

From the point of view of image analysis, the fan scan configuration does not constitute an additional difficulty, because the line signals are processed directly in the  $(r, \theta)$ -space, where they appear as a standard rectangular image. The  $(r, \theta)$ -space is always the reference, unless the processing algorithms require the knowledge of the real anatomical shapes of the imaged scene or in case the processing results are forwarded to physicians as a map on the anatomical image.

The three phenomena contributing to the realization of an ultrasound echo are (a) the pure reflection from a smooth surface, (b) the diffuse reflection from a rough surface and (c) the refraction, this latter causing artefacts during image formation which considers possible only line transmission. A fourth factor characterizing the echo is (d) the diffusion caused by interaction with many rough surfaces smaller than the ultrasound wavelength. In this case the pulse is diffused in all directions, determining a random component of the ultrasound signal, called scattering. The presence of scattering in the recorded echo makes the signal non-stationary.

Nevertheless the RF signal can be considered stationary in a homogeneous area, where it is characterised by a slowly varying amplitude and a stationary random component. The RF signal can, thus, be decomposed into its local mean value, representing the echogenicity of the area, and a variable component representing the speckle noise, a complex type of disturbance usually modelled as multiplicative.

Although the speckle noise is dependent on the acquisition system properties, it also contains some information about the imaged area, intuitively linked to the texture appearance in ultrasound images. While image processing, aimed at improving the visual quality of ultrasonographies, are focused on suppressing speckle noise, image analysis techniques are often interested in capturing as much information as possible from this random component which can be useful for tissue characterization.

One more important characteristic of the RF signal is the frequency shift: as signal attenuation increases with frequency, during propagation the high frequency components of the echo signal are more attenuated than the low frequency ones, globally reducing the value of the central frequency of the received signal. The frequency shift may also deteriorate the radial resolution, but the estimation of these changes are useful to study the attenuation of different types of materials and thus perform tissue characteriza-



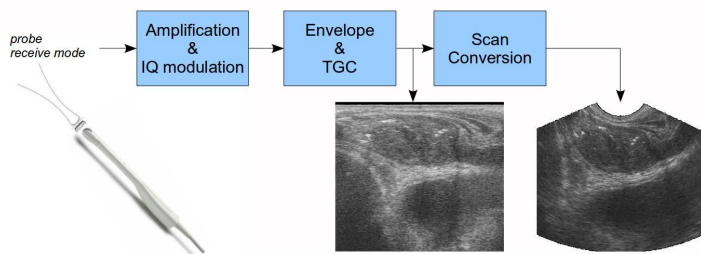


Figure 1.2: Standard signal processing chain in medical ultrasound acquisition systems. The signal recorded by the probe is the raw RF echo signal, while the final output displayed on monitor is the envelope of the base-band signal.

tion. The knowledge about the frequency shift can also be used to compensate the attenuation independently for each ray during time-gain compensation (TGC) procedures and restore the original echo spectrum.

Since the dynamic range of the RF signal is very large because of the different amplitude of strong reflection components and weak scattering, a logarithmic transform is employed to reduce echo dynamics before digitalization. The final digital raw signal is obtained by sampling the recorded signal, amplified and processed by TGC and logarithmic transform. The sampling frequency is chosen in the range of 20-60 MHz and the number of bits used for quantization is between 16 to 24.

In medical ultrasound imaging the standard signal processing chain performs a demodulation of the RF signal by Hilbert Transform, providing the in phase-quadrature signal (IQ), and computes the envelope as the module of the IQ analytic signal. Dealing with the envelope allows a decimation of the signal according to the reduction of its band and provides a better visualization of the ultrasound image. A scheme of the standard processing chain in ultrasound imaging is shown in figure 1.2.

The base-band signal or B-mode signal is the standard output of an ultrasound acquisition system, while the original RF signal is the unprocessed information acquired by the probe. The basic principle of most tissue characterization techniques in ultrasound is extracting as much information as possible from both the base-band signal and the RF signal.

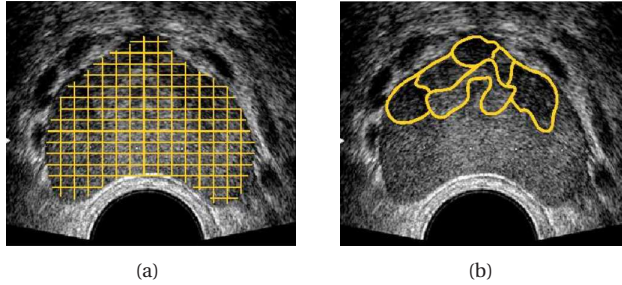


Figure 1.3: Rectangular ROI segmentation inside the prostate region is illustrated in figure 1.3(a), while figure 1.3(b) shows irregularly shaped ROIs.

## 1.2 Tissue characterization

The final purpose of biomedical image analysis is tissue characterization. In practice the analysis consists in segmenting the image into several regions of interest (ROIs), computing some characteristics, features or measures related to each ROI and classifying them into one of few known categories. Any category represents a particular state of the tissue outlined in a ROI; the simplest case is binary classification, aimed at discriminating a normal and pathological tissue, neglecting the presence of benign abnormalities.

ROIs can be defined as irregular shaped areas in the image, following some anatomical characteristics of the imaged tissues; alternatively they can be chosen as rectangular regions covering the whole image, with or without overlapping (figure 1.3). While rectangular tessellation can be easily performed by computers, segmentation of ultrasound images into irregular regions requires either a manual procedure or a complex automatic algorithm subject to high degree human supervision.

Ultrasound image segmentation presents some characteristic artefacts that make the segmentation task complicated, and almost impossible by means of standard image segmentation techniques. In fact, while in other areas of medical imaging (CT, MRI) the application of standard image processing methods is sufficient to obtain good segmentations, in ultrasound satisfactory results can only be achieved by specific methods. A review of segmentation techniques in medical imaging is provided by [3] and [4].

The ROIs consist of a two-dimensional arrangement of recorded RF signals, which are used to compute some features useful to clas-

sify the state of the outlined tissue. Let us consider the extraction of  $N$  different features from each ROI: they can then be represented as a point  $\mathbf{x} = x_1, \dots, x_{N_f}$  in an  $N_f$ -dimensional feature space. ROIs with similar characteristics will form clusters in the feature space.

The first step in tissue characterization is the selection of significant features which are able to capture the discriminant characteristics of normal and pathological tissues, ideally providing a representation of different tissues in the feature space as separate clusters. In the end, a pattern classification step in image analysis is aimed at defining the optimal decision boundary in the feature space to separate normal and pathological tissues.

The construction of the optimal decision boundary for tissue characterization is possible thanks to supervised pattern classification techniques based on an accurate medical ground truth. The collection of a large number of ultrasound images from biological tissues and their corresponding histological examination is essential to create a ground truth, which is used in the design of a classification algorithm and its validation.

Ground truth data are divided into a training set and a testing set. The training set consists of a set of ROI-label pairs, already classified by a chosen gold standard like histological analysis, which are used to perform the learning of a classification algorithm. The testing set instead is used to validate the algorithm. Test ROIs labels are used only to compute the performance criteria and do not provide any further information to the classifier. In some cases also a validation set is extracted from the ground truth in order to tune some parameters of the classification algorithm.

The main problem in ultrasonic tissue characterization is the absence of a public standard database, also for very common and highly diagnosed diseases, as prostate carcinoma or breast tumour. One fundamental contribution of this research is the realization of a large ground truth database for prostate cancer, collected in collaboration with the department of Urology of the University of Bologna.

In ultrasound imaging, system effects and tissue micro structures affect in a complex way the signal and the speckle noise, thus any conclusion about tissue characterization is valid under the same conditions and equipment settings. Nevertheless algorithms designed to make the signal less dependent on acquisition system adjustments can be employed to provide more efficient tissue characterization.

## Reducing system effects

Ultrasound imaging acquisition systems can affect and deteriorate tissue response. In order to reduce these undesirable effects and, hopefully, obtain a higher quality and more significant image, several processing techniques based on image formation linear models were developed.

The most common two-dimensional linear image formation model is defined through a convolution operator:

$$\begin{aligned} x &= \mathbf{H}(w) + \mu \\ x[n, m] &= \sum_k \sum_l h[n-k, m-l] w[k, l] + \mu[n, m] \end{aligned} \quad (1.1)$$

where  $x$  is the image,  $w$  is the tissue response to be recovered,  $\mathbf{H}$  is the linear operator accounting for system effects which defines the point spread function (PSF)  $h$ , and  $\mu$  is some additive term accounting for both experimental and model noises. The inverse problem of recovering  $w$  is also named as deconvolution [5]. In medical ultrasound imaging the PSF is usually not known and is estimated together with the true image  $w$ . In this case the problem is referred to as blind deconvolution [6].

The problem with model (1.1) is that the presence of tissues between the transducer and the target changes locally the PSF which cannot be considered shift-invariant in the whole *in vivo* image. The presence of phase aberration and dispersive attenuation in ultrasound imaging lead necessarily to a spatially variant model for the PSF which requires a non stationary deconvolution.

The most used technique to deal with shift-variant convolution kernel is to segment the image in several areas where the signal can be considered stationary and the local PSF shift-invariant. An alternative to this approach is the employment of adaptive algorithms based on models able to change their characteristics along with the signal.

Furthermore, although the image formation model is expressed in 2D, ultrasound imaging deconvolution algorithms are often developed in 1D and separately applied along the lateral and axial direction. Anyway it is common practice to perform deconvolution along the axial direction only, as the PSF is wider. In this case the convolutional model becomes:

$$x[n] = \sum_k h[n-k] w[k] + \mu[n] \quad (1.2)$$

The separability of the PSF allows one to reduce the problem of a 2D deconvolution to a sequence of 1D deconvolution problems,

applied along each direction independently. Specifically, it is assumed that the 2D convolution kernel can be expressed as product of two one dimensional convolution kernels along the two directions: the axial PSF includes the blurring effects due to the finite bandwidth of the transducer and the dispersive attenuation of the acoustical pressure pulse in tissue, the lateral PSF represents the convolutional components of lateral blurring due to the complex beam pattern.

One of the classical methods of blind deconvolution used in medical ultrasound imaging originates from the theory of system identification and is based on statistical modeling. An efficient statistical method for deconvolution assumes the RF signal to be an autoregressive moving average (ARMA) process and estimates model parameters to recover the tissue response. Hence, in this interpretation, the RF-signal  $x[n]$  is considered to be the output sequence of a causal filter  $h[n]$ , while the tissue response or reflectivity function,  $w[n]$ , is thereby considered to be an input driving sequence. The latter is commonly assumed to be a realization of an independent and identically distributed process of zero mean and fixed variance.

ARMA models present highly flexible modeling capability, because they integrate properties of AR models, useful for modeling random signals that possess peaky power spectral densities, and properties of MA models, more appropriate for signals that have broad peaks or sharp nulls in their spectra.

For the case of ultrasound imaging the choice of an ARMA model is well justified. This is because the spectra of the RF lines are normally wide-peaked, and, since the requirement of stationarity imposes limitations on the range of sampling intervals, stationary RF signals are usually short in duration. ARMA modeling is appropriate in this context, since it provides a well-posed estimation problem with fewer parameters than observations and is characterized by excellent modeling capability for relatively low orders and for short data segments.

While deconvolution techniques are generally used to improve the visual quality of ultrasound image, in this work it is introduced as a preprocessing step in the standard tissue characterization procedure, in order to enhance the diagnostic significance of ultrasound features. A scheme of the proposed tissue characterization protocol is shown in figure 1.4.

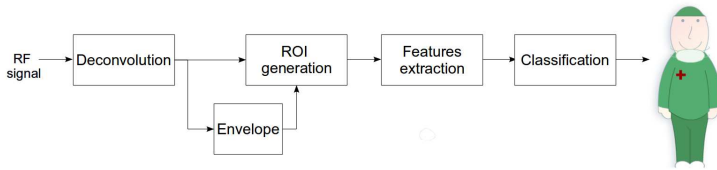


Figure 1.4: Proposed characterization procedure with the addition of deconvolution to improve diagnostic performance.

## Feature extraction and selection

A large number of features were designed to extract interesting information from the ultrasound RF signal. The B-mode signal envelope itself can be considered as a feature containing all the visual information available to the physician.

The features derived from the RF signal developed for tissue characterization usually belong to one of these three classes:

- *Spectral*: these features gather information about the spectral behaviour of the RF signal. This may include the study of the frequency shift and the related attenuation; the spectral decomposition of the coherent and random component of the signal; multi-resolution analysis by wavelets or the estimation of any mathematical model for the spectrum.
- *Statistical*: the RF image can be modeled as random samples from some statistical distribution. The parameters of such a distribution can be estimated locally and provide statistical features.
- *Textural*: these characteristics give information about the arrangement of grey levels and are computed from the B-mode signal. Textural features represent the closest attributes to what is processed by human brain when analysing an image.

The number of features can increase rapidly as new models for the RF signal or new attributes are proposed. The rise in dimensionality can only apparently make the discrimination task easier. In practice not all the features extracted are interesting for tissue characterization. Furthermore high dimensionality causes more complex classification models which tend more likely to overfit the

training data. For this reason it is essential to select few significant and non-redundant features.

Feature selection algorithms [7] focus on the realization of the optimal feature subset for a certain problem, according to a dataset containing feature values for any sample and the corresponding category. In the context of image feature selection, the dataset is constituted by feature values from any ROI and the information about the pathological state of the tissue imaged by the region. The selection approaches can be roughly divided into three categories: filters, wrappers and hybrid methods.

Filters sort features according to a score measure that summarizes the importance of a feature with respect to the state of the tissue. Typically, the ranking criterion can measure either distance, information, dependency, or consistency between features extracted from each ROI and the respective class. Filters measures are independent on any classification algorithm.

On the contrary, wrappers sort attributes from a feature set according to the performance of a classifier trained on that feature set; they are thus classifier-dependent and require a larger computational cost.

Hybrid methods take advantage both of filter and wrapper methods. The filter measure is used to decide the best subset for a given cardinality, while the wrapper mining algorithm selects the final best subset among the best subsets across different cardinalities.

In general a complete search procedure is never adopted in feature selection because of its huge computational cost; on the contrary, although they cannot guarantee that the actual best feature subset is selected, forward and backward sequential search or random search techniques are preferred because of their lower computational cost.

The above methods are obviously supervised since they take advantage of the information about tissue state, but also unsupervised methods such as principal component analysis (PCA) can be useful, above all in order to reduce redundancy in subsets where features have similar informative value. An unsupervised method can thus be used as a first step to discard redundant attributes and a second step of supervised feature selection can subsequently be applied to further reduce the feature set, eliminating attributes irrelevant for the particular classification problem.

A complete feature selection scheme can be tested by introducing artificially designed irrelevant and redundant attributes in a benchmark feature set; an efficient feature selection algorithm

must be able to detect and discard these test attributes.

### ROI classification

The goal of image tissue characterization is employing the feature set to build a classifier able to discriminate between different categories of imaged tissues. Among the several approaches to classification existing in literature [8], non linear discriminant analysis through kernels is used in this thesis to take advantage of the non linear mapping of the feature space and ease the classification task for real data.

In supervised machine learning the data set  $(\mathbf{x}_k, c_k)$  composed of ROI feature vector and label pairs is used to determine the predictive model representing the relationship  $c = f(\mathbf{x})$  between the input variables  $\mathbf{x}$  and the output variable  $c$ . It is the knowledge about class labels  $c_k$  to make the learning problem supervised.

The found predictive model is then used to separate in the feature space and normal and pathological ROIs. In this case the classification problem is binary and the label index  $c_k$  can only assume two values, for example 1 for benignant tissue and 2 for pathological tissue.

A linear model  $f(\mathbf{x})$  employs a predictive function represented by an hyperplane in the feature space, whose parameters  $\mathbf{w}$  and  $b$  are determined using classical criteria such as least squares or maximum likelihood:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (1.3)$$

In this case the predictive model is also called discriminant function. Traditional linear methods give the same importance to all samples in the dataset and this can easily provide non efficient linear separation, because of the excessive influence due to samples very far from the separation boundary.

A way to overcome this problem is focusing on those training examples which lie close to the class they do not belong and, thus, close to the boundary. These samples are called support vectors and in this approach they are the only data that explicitly define the model. A linear model in terms of support vector,  $sv$ , is expressed as in the following:

$$f(\mathbf{x}) = \sum_{i \in sv} \alpha_i c_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b \quad (1.4)$$



where the summation includes only the training examples  $\mathbf{x}_i$  that are support vectors, and  $\alpha_i$  are coefficients determined as Lagrange multipliers in the optimization procedure.

An advantage of this approach is that the classifier concentrates directly on examples  $sv$  that are difficult to classify and its computation scales with the number of support vectors rather than the dimension of the space which can be very large. In addition this approach is shown to balance training classification error and model complexity, thereby avoiding overfitting, a situation in which the model is too finely tune to the training examples and fails to perform well on new data.

When real ultrasound data are analysed, linear separation is generally not possible in the feature space, even if a large number of features is employed. In this condition it is important to exploit techniques providing a simple way to obtain a non linear model from any linear model based on inner products. Even classical techniques can be turned easily into flexible non linear techniques via the so called *kernel trick*.

According to this technique, a non linear function  $\phi$  mapping the data into a different higher dimensional feature space can be applied. After this transformation separability will be enhanced thanks to the higher dimension of the mapped space, which can also be infinite. The kernel trick recognises that the classification in this new space can be obtained without actually performing the transformation, because the transformation  $\phi$  applied to samples in predictive function only appears in the form of an inner product. Therefore non linear approaches can be performed efficiently by using kernel functions  $\mathbf{K}(.,.)$  which act like dot products in the re-mapped feature space [9].

It is never actually necessary to compute  $\phi$  or to define it explicitly. Instead it is sufficient to define the kernel function  $\mathbf{K}(.,.)$ , and it can be shown that any symmetric positive semidefinite function suffices.

For each ROI sample,  $\mathbf{x}$ , the kernel approach provides its projection onto a vector in  $F$ , expressed as an optimized linear combination of the inner products of the mapped sample and a sequence of mapped support vectors:

$$\sum_{i \in sv} a_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle = \sum_{i \in sv} a_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}). \quad (1.5)$$

The combination weights  $a_i$  are tuned to maximize the inter-class variance and minimize the intra-class variance of data in the mapped

space. Intuitively, the effect of the kernel is to measure the similarity between a test vector  $\mathbf{x}$  and each of the support vectors. In fact vectors belonging to one of the classes are presumably most similar to the support vectors belonging to that class, hence these similarity values provide us with the information to build the predictive function.

In non linear data transformations an appropriate kernel for representing the similarity of a mapped sample to the support vectors is the radial basis function (RBF) kernel:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}\|^2}{2\sigma_{rbf}^2}\right). \quad (1.6)$$

The kernel variance  $\sigma_{rbf}^2$  is tuned by cross-validation on a validation data set. In particular, a part of the dataset is reserved to perform several tests varying the value of  $\sigma_{rbf}^2$ , in order to find the optimal value for this parameter, which provides the best classification performances.

Cross-validation is one of the techniques for statistical resampling used to evaluate performance and improve robustness of machine learning models and to estimate statistical significance levels. These techniques are as important as the predictive model itself.

After non linear transformation, a linear classifier can be used in the mapped space to find the discrimination hyperplane defined by the following predictive function:

$$f(\mathbf{x}) = \sum_{i \in sv} a_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b. \quad (1.7)$$

Linear classifiers are suitable to classify kernel transformed data, which are characterized by a better separation in the new feature space. For example, the Fisher Linear Discriminant (FLD) approach tunes the parameters of the linear model in (1.3),  $\mathbf{w}$  and  $b$ , by maximizing the Mahalanobis distance and in such a way that different clusters have the same distance from the hyperplane. The Mahalanobis distance  $J$  between two clusters belonging to different classes is expressed through the intra class mean value  $\boldsymbol{\mu}_i$  and covariance matrix  $\boldsymbol{\Sigma}_i$  of class  $i$ :

$$J = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (1.8)$$

The performance indicators computed on the testing set offer a way to assess the diagnostic validity of a CAD scheme. In a

binary tissue characterization problem, where the two categories benignant/malignant tissue are available, it is common to define True Positives ( $TP$ ) the pathological samples classified as malignant, False Positives ( $FP$ ) the misclassified normal regions, True Negatives ( $TN$ ) the normal samples classified as benignant, False Negatives ( $FN$ ) the misclassified pathological regions.

According to these definitions, besides accuracy ( $ACC$ ), the standard performance criteria are sensitivity ( $SE$ ) and specificity ( $SP$ ), giving information about the ability in recognizing pathological and normal tissue respectively. The prediction capability of the method is also expressed by the probability that a malignant classified sample is actually pathological. This information is provided by the positive predictive value ( $PPV$ ). The performance indicators are resumed in the following:

$$\begin{aligned}
 ACC &= \frac{TP+TN}{TP+TN+FP+FN} \\
 SE &= \frac{TP}{TP+FN} \\
 SP &= \frac{TN}{TN+FP} \\
 PPV &= \frac{TP}{TP+FP}
 \end{aligned} \tag{1.9}$$

The above definitions of performance criteria consider a region as a sample of the dataset, but according to the focused scale of the problem, also pixel-based, core-based and patient-based performances can be defined, considering as sample of the dataset respectively the pixel, the core or the patient.

The variation of the threshold in a linear classification method allows the complete description of the classifier performances, which can be visualized in a receiver operating characteristic (ROC) curve plotting the ( $SE, 1 - SP$ ) graph. The area under the ROC curve is also specific of the method diagnostic accuracy which increases as the area approaches the unit value.

### 1.3 Application: prostate TRUS

One of the applications where Computer-Aided Detection (CAD) tools would be extremely valuable is the recognition of prostate cancer in trans-rectal ultrasound (TRUS) images. Some information about the physiology and pathology of prostate gland will be presented in the following paragraphs.

The main function of the prostate is to store and secrete a clear, slightly alkaline (pH 7.29) fluid that constitutes 10 – 30% of the volume of the seminal fluid that, along with spermatozoa, consti-

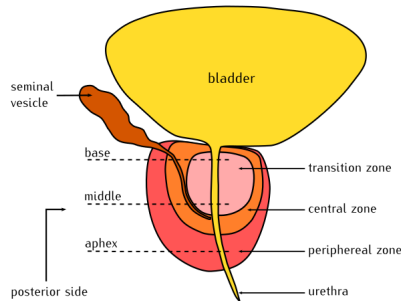


Figure 1.5: Schematic representation of the prostate gland.

tutes semen. The rest of the seminal fluid is produced by the two seminal vesicles.

A healthy adult human prostate is a chestnut shaped gland enveloped in a fibrous capsule. Its base is attached below the urinary bladder neck and the apex is fixed to the urogenital diaphragm. It borders on the posterior side with the rectum and on the anterior side with the fibromuscular stroma connected to the pubis through the puboprostatic ligaments. On the superior posterior side it is attached to the seminal vesicles, a pair of simple tubular glands that secrete a significant proportion of the fluid that ultimately becomes semen. The excretory ducts of seminal vesicles open into the vas deferens, as they enter the prostate gland, and they are lined with the epithelium of the transition zone. Within the prostate, the urethra coming from the bladder is called the prostatic urethra and merges with the two ejaculatory ducts. The prostate is finally sheathed in the muscles of the pelvic floor, which contract during the ejaculatory process.

The prostate gland is divided in three different zones: the transition zone (TZ), the central zone (CZ), and the peripheral zone (PZ) (figure 1.5). The transition zone surrounds the urethra and extends from the ejaculatory ducts proximally. The peripheral zone encompasses the urethra from the base to the apex. The central zone is composed of tissue immediately surrounding the ejaculatory ducts and it expands inferiorly. The significance of this architecture is based upon the relationship of these three zones to prostatic disease [10]. In the young males the peripheral zone comprises 75% of prostate volume, the transition zone 20% and the

central zone 5%. These ratios change and after the age of 40 years the transition zone may enlarge and occupy most of the gland as benign hyperplasia is almost inevitable.

The prostate gland is the male organ most often smitten by either benign or malignant lesions and the second leading cause of cancer death for men [11]. Prostate diseases can be distinguished in three main categories: prostatitis, benign hyperplasia and prostate cancer.

Prostatitis is an inflammation of the prostate gland. It is a benign pathology and in its acute case it is mainly treated with antibiotics.

Benign hyperplasia (BPH) consists in a prostate enlargement. It is fairly common among ageing men and it occurs mainly in the transition zone. BPH can be treated with medication, a minimally invasive procedure or, in extreme cases, with a surgery procedure that removes the prostate.

Prostate cancer, often referred as prostate carcinoma (or adenocarcinoma), is a malignant pathology and, as all cancers, it is characterized by an abnormal and uncontrolled cells mutation and replication. Like BPH, prostate cancer might cause pain, difficulty in urinating, problems during sexual intercourse, erectile dysfunction and other symptoms. If not detected on early stages and in presence of more aggressive forms, the disease can advance to stages characterized by local invasion of the surrounding tissues (seminal vesicles, bones, rectum), usually resulting in lethality.

The heterogeneous and multifocal nature of prostate cancer lesions poses significant difficulties in its detection. With regards to heterogeneity, prostate cancer tissue typically reveals a juxtaposition on benign cells, preneoplastic lesions and neoplastic lesions of varying severity [10].

The current clinical procedure to detect cancer is based on a combination of different diagnostic tools, because none of these tools is accurate enough to be used alone. Digital Rectal Examination (DRE), Prostate-Specific Antigen (PSA) evaluation, TRUS image analysis and biopsy are all part of the medical procedure for prostate analysis. Each of these tools presents important limitations that make accurate prostate cancer detection still an unsolved problem.

Historically, DRE has been the principal method of prostate analysis, but it is accurate only at detecting large and superficial lesions and it is strongly operator dependent.

The first step in public screening for prostate cancer is the mea-

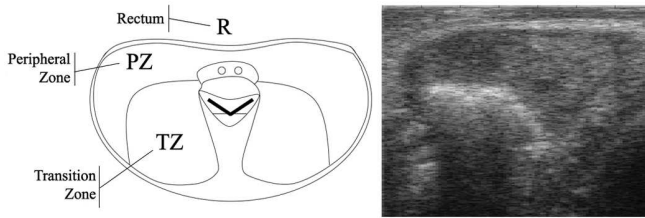


Figure 1.6: The picture on the left gives a schematic representation of prostate axial section with orientation inverted respect to its real anatomy. On the right a TRUS image shows the prostate gland in the same orientation.

surement of the blood PSA level, a glycoprotein produced almost exclusively in the epithelium of the prostate gland. Unfortunately, PSA is specific for prostate but not for cancer, since other factors such as BPH, prostate infection, urethral instrumentation and irritation can cause an increase in the PSA value.

At TRUS, the normal prostate gland has a homogeneous, uniform echo pattern. The appearance of carcinoma on ultrasound is variable and in its early stages a tumour can appear anechoic, hypoechoic or isoechoic with respect to the surrounding normal tissues. Potential hypoechoic regions could also include BPH or even normal biological structure, thus the specificity of the TRUS images visual inspection is low.

Standard TRUS images show the axial or the lateral section of the prostatic gland. The gland anatomical shape is inverted in the image due to the acquisition setting, as shown in Fig. 1.6, where the superior hyperechoic boundary marks the interface between the rectum and the peripheral zone of the prostate. The transition zone is visible as a darker region below the peripheral zone. The gland contours are generally well defined and visible in the images.

The histopathological analysis of biopsy samples is the standard for cancer detection confirmation.

The dominant and most reliable method for prostate carcinoma diagnosis and aggressiveness assessment, in research as well as in clinical procedure, is the Gleason grading [12]. This method is based entirely on the histological evaluation of carcinomas cells arrangement patterns in contrast agents stained prostatic tissue sections. Specifically, the method is a categorization of glandular differentiation and pattern of growth of the tumour, at relatively low magnification ( $\times 10 - \times 40$ ) in five basic grades patterns [12].

Due to the heterogeneous nature of prostate cancer lesions, histopathological inspection of prostate tissues often reveal an ensemble of benign lesions, preneoplastic lesions and neoplastic lesions with different aggressiveness. To take into account this heterogeneity, the five basic grade patterns are used to generate an histological score, which can range from 2 to 10, by adding the primary grade pattern and the secondary grade pattern. The primary pattern is the one that is predominant in area, by simple visual inspection. The secondary is the second most common pattern.

The Gleason grading score is therefore an indicator of cancer stage: the higher the grade is, the more advanced the cancer is. Typically, cancers with Gleason scores lower than 6 are considered well differentiated and associated with a good prognosis. Those with a Gleason score of 8-10 have the worst prognosis and the highest risk of recurrence. Gleason score is often combined with PSA level and clinical stage for risk assessment. The primary goal of staging is to distinguish patients with organ-confined, locally invasive, or metastatic diseases.

Ultrasounds are used for biopsy guidance, to enable the sampling of all relevant areas of the prostate by means of systematic sampling protocols [13]. However, the main limitations in this procedure are due to the multifocal nature of cancer and to the sampling process.

In fact, uniquely among image-guided biopsies, prostate biopsy is not lesion-directed but rather based on a systematic sampling of those areas where cancer incidence is higher. As the disease is often multi-focal, different areas throughout the whole gland are sampled. In particular, since most of the cancers arises in the peripheral zone, most of the protocols aim to maximize its sampling. The motivation behind this, as discussed before, is the weak diagnostic significance of imaging inspection techniques due to the high variability of prostate cancer patterns.

Ultrasound imaging is thus used to guide and monitor the biopsy needles insertions in the different areas of the prostate gland. Since metal, if compared to soft tissues, has an high acoustic impedance, biopsy needles appears as hyperechoic lines in TRUS scans, as shown in figure 1.7.

The first landmark sampling technique was the sextant protocol reported in 1989 [14]. As originally described, six biopsies were obtained in a parasagittal line drawn halfway between the lateral border and midlines bilaterally, from the base, midgland and apex, as shown on the left side of figure 1.7.

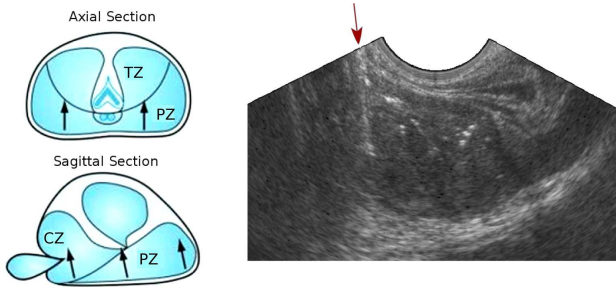


Figure 1.7: The picture on the left gives a schematic representation of double sextant biopsy protocol. On the right a TRUS image shows the insertion of biopsy needle in the prostate.

Although sextant biopsy protocol was a major advance, with a 20-25% of positive biopsy rate, with a wider experience it was found also inaccurate, principally because it under-samples the peripheral zone [13].

Modifications of the sextant protocol were introduced from the mid 1990s onwards. For example, in the modified sextant biopsy protocol a better sampling of the peripheral zone around the lateral margins is obtained moving laterally and angling anterolaterally the biopsy trajectories. This improved in some cases the detection rate from 80 to 89 % [13].

However, even the modified sextant protocol was found to miss some tumours with time and many alternatives were explored. Therefore, several extended protocols using more cores (8 or 10) directed to the peripheral zone were introduced. Although the available results show that the extended protocols can improve diagnosis accuracy it is still a matter of debate whether extended protocols are substantially better than the modified sextant protocol.

As regards patients perception of this examination, it was documented that 55% of men report physical discomfort during biopsy. Moreover, this procedure is not completely safe, since it carries the risk of bleeding infection or even urosepsis [15].

Research to improve the efficiency of the standard prostate biopsy protocol is proceeding both from an image processing point of view and from a purely clinical perspective. In fact, as discussed before, the positive predictive value  $ppv_0$  of this protocol, that is the probability of sampling a pathological core given a pathologi-



cal patient, is quite low, between 20 and 25%. As the probability to detect the disease in a pathological patient increases with the number of samples  $N_0$ , according to  $1 - (1 - ppv_0)^{N_0}$ , the detection rate of the protocol is about 89%. The inefficiency of this protocol stems from the fact that most of the sampled tissue is benign and, thus, represents an unnecessary biopsy.

In order to improve the efficiency of the prostate biopsy protocol it is essential to increase the  $PPV$ , in such a way that the clinical procedure can preserve the same detection rate of the standard protocol with a reduced number of cores.

Since the processing of TRUS images highlights important characteristics of the investigated tissues, tissue characterization techniques can be employed to increase the  $PPV$ , providing the radiologist with a malignancy map to perform lesion-directed biopsy. In this new scenario the number of unnecessary biopsies would be reduced, while preserving the same diagnostic power as the standard protocol.



---

## ULTRASOUND IMAGE ANALYSIS AND INTERPRETATION

---

*I imagine this is because Nature wants to ensure that the evils of wisdom shall not spread further throughout mankind.*

Desiderius Erasmus of Rotterdam (1466 - 1536)

COMPUTER-AIDED detection schemes are decision making support tools, useful to overcome limitations of problematic clinical procedures. The topic treated in this thesis is the study and the realization of CAD techniques based on trans-rectal ultrasound images, which may be extremely important to support prostate cancer diagnosis.

First works on TRUS-based CAD schemes for detection of prostate cancer consisted on biopsy ground truth and analysis of rectangular regions around needle insertion points. The main characteristic of these studies is the use of textural features extracted from TRUS images to discriminate different tissues. Basset *et al.* [16], Huynen *et al.* [17] and Houston *et al.* [18] realized clinical studies based on the extraction of first and second order statistics textural parameters and used simple decision trees to perform classification.

After these works, the main trend shifted to a multi-feature approach, extracting features of different nature from TRUS images and combining them to obtain higher classification performance. Schmitz *et al.* [19] and Scheipers *et al.* [20] employed textural features and spectral parameters extracted from RF data, while Feleppa *et al.* [21] introduced clinical data like PSA value and patient's age in the feature vector besides spectral parameters.

The classifiers used for these works (self organizing Kohonen map, neuro-fuzzy systems, Artificial Neural Network) are more

complex with respect to decision trees and define a new direction in this research field. Furthermore in these works the ground truth comes not from biopsy but from prostatectomy and histological analysis of prostate slices.

The work of Mohamed and Salama[22] represents an exception since the gold standard is based on radiology visual inspection. The proposed scheme utilizes a pure textural feature vector with a Support Vector Machine (SVM) classifier. In this case the values of sensitivity, specificity and area under ROC curve are high, but they are obtained on a small ground truth.

Similar performance was obtained in latest studies like the extension of Mohamed's work [23] which includes spectral features and the study performed by Han *et al.* [24], where morphological features and multi-resolution textural features are employed and SVM is used as classifier. In the latter work notable values of sensibility and specificity are reported, but the proposed method was only tested on malignant images, thus there is no information about its behaviour in completely healthy cases.

This chapter presents an effective approach to realize a CAD scheme for prostate cancer detection, employing a multi-feature nonlinear classification model and the predictive deconvolution of the acquired RF signals to reduce system-dependent effects.

The mutual information of feature values and tissue pathological state is used to select, among several analysed ultrasonic characteristics, the features essential for tissue characterization.

A clinical study, performed on ground truth images from biopsy findings, provides a comparison of the classification model applied before and after deconvolution, showing in the latter case a significant gain in accuracy and area under the ROC curve.

Here the aim is to investigate the ability of the features extracted from deconvolved US images at discriminating pathological tissues. This issue is analysed in terms of performance comparison of a nonlinear classification model trained on features extracted from US images, with and without deconvolution preprocessing.

In addition, while rectangular ROIs are usually employed in ultrasonic tissue characterization literature, we propose an algorithm for irregular shaped ROIs segmentation and perform analysis to test the usefulness of this segmentation for tissue characterization purposes, without including the deconvolution step.

A large number of features is extracted from US RF signal to characterize different aspects of biological tissues. The combined use of features of different nature results in more accurate tissue

characterization, but requires a critical operation of feature selection to identify features highly correlated to the pathological state of the tissue.

In our approach a hybrid feature selection algorithm based on the mutual information of feature set and ground truth class is used to prune unimportant features and achieve fast computation.

The ground truth used in this clinical study is based on biopsy findings and histological analysis of some regions in the images considered suspicious at visual inspection by expert radiologist.

The proposed CAD scheme was tested on a ground truth containing both benignant and malignant cases.

The analysis of the CAD scheme employing irregular shaped ROIs shows that non linear classification ensures in this case a sensitivity always greater than 80%, but provides a limited specificity.

Furthermore, the addition of predictive deconvolution in the proposed CAD scheme increases classification performances, providing a sensitivity of 90%, specificity of 93% and area under the ROC curve of 95%.

## 2.1 Dimensionality reduction

Results of recent studies show that combining features extracted from RF analysis of ultrasound signals and image-based texture parameters results in more effective classification procedures [25]. Literature about tissue characterization in ultrasound analysis provides a large amount of features of different nature which can be subdivided according to their contribution in highlighting some properties of the tissue.

Parameters of statistical distributions, such as Nakagami model [26], give information about scatterer density, regularity and amplitude.

Spectral features [27] describe fluctuations of physical properties as acoustic impedance, viscosity and elasticity resulting in backscattering signals. Typical spectral parameters capture the shifting of RF signal central frequency due to attenuation. Also the wavelet coefficients of RF signal, their polynomial fitting [28] and the coherent and diffuse components obtained by signal decomposition [29] belong to the spectral features group and provide important properties to type tissues.

Features extraction from B-mode images aims mainly at the detection of textural properties of speckle which represents the

macroscopic appearance of the scattering generated by tissue microstructures. Different kind of textural parameters are available in literature. Haralick [30] and Unser features [31] are both based on the gray levels distribution statistics, while Fractal features [32] rely on modeling and computation of fractal dimension. In our approach, the characteristic skills of different features are combined to define a feature set endowed with a high discriminating power between healthy and cancerous tissues.

A complete feature set of all parameters mentioned before would have a huge dimensionality of about 140 attributes and is described in appendix A. For this reason, a first selection step is performed keeping for each group of features only those correlated to the ground truth class, and discarding the other ones.

The feature set is built from the ground truth dataset that will be described in section 2.5, computing features on rectangular overlapping ROIs selected from the images. The feature data matrix is represented as  $S = [\mathbf{f}_1 | \dots | \mathbf{f}_{N_f}]$ , where  $\mathbf{f}_i$  is the  $i$ -th feature vector with a value for each ROI, such that  $S$  is a matrix of size  $N_{ROI} \times N_f$ : number of ROI  $\times$  number of features. The ground-truth class vector, containing the index representing the pathological state of the tissue outlined by a ROI, is expressed by  $\mathbf{c}$  of length  $N_{ROI}$ .

By selecting feature through correlation analysis, the dimensionality is reduced, but synergies between different features are saved. The defined feature set is constituted by 54 features, as shown in table 2.1.

Table 2.1: Feature Set

Feature	Origin/Type	#
Wavelet Transform (WT)	RF/Spectral	2
Polynomial Fit of WT	RF/Spectral	1
Wavelet Decomposition	RF/Spectral	1
Central Frequency	RF/Spectral	1
Attenuation	RF/Spectral	13
B-mode	B-mode/Envelope	1
Nakagami	RF/Statistic	4
Statistic	RF/Statistic	2
Haralick	B-mode/Textural	4
Unser	B-mode/Textural	9
Fractal	B-mode/Textural	16

The importance of CAD in diagnostic ultrasound imaging depends on its ability to perform nearly real-time classification in

order to give a second opinion to the physician. A large feature set dimensionality prevents the use of ultrasound images automatic characterization as a diagnostic decision support tool, because feature computation would be too demanding in computational terms. For this reason a further classification-oriented Feature Selection (FS) step is essential.

In particular, in our approach a Mutual Information Hybrid FS (MIHFS) algorithm is used to rank and prune the whole feature set. Ranking measures based on distance and dependency [33] were also tested and, although most of them succeed in discarding irrelevant features, only information based measures are able to recognize and discard redundant features. In the proposed FS technique the chosen classifier-independent measure is the min-Redundant Max-Relevance (mRMR) criterion proposed by Peng *et al.* [34]. The mRMR measure is based on mutual information between the current feature set and ROI class, corrected by the averaged mutual information between features in the feature set. Maximizing this measure allows to define a feature set with maximum relevance  $D(S, \mathbf{c})$  and minimum redundancy  $R(S)$ , by maximizing the mRMR measure:

$$\max_S D(S, \mathbf{c}) \quad D(S, \mathbf{c}) = \frac{1}{|S|} \sum_{\mathbf{f}_i \in S} I(\mathbf{f}_i, \mathbf{c}) \quad (2.1)$$

$$\min_S R(S) \quad R(S) = \frac{1}{|S|^2} \sum_{\mathbf{f}_i, \mathbf{f}_j \in S} I(\mathbf{f}_i, \mathbf{f}_j) \quad (2.2)$$

$$S = \operatorname{argmax} \Phi(D, R) \quad \Phi(D, R) = D - R \quad (2.3)$$

where  $I(X, Y)$  is the mutual information of two discrete random variables  $X$  and  $Y$ , representing two features or a feature and the ROI class:

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left( \frac{p(x, y)}{p_1(x) p_2(y)} \right) \quad (2.4)$$

The joint and marginal probability density functions,  $p(x, y)$ ,  $p_1(x)$  and  $p_2(y)$ , can be estimated through the ROI-based histograms of feature values.

MIHFS is an algorithm in two steps. The filter step consists in ranking all features according to the mRMR measure, shown in (2.3), following a sequential forward selection as search technique.

MIHFS wrapper step exploits a Fisher Linear Discriminant (FLD) [35] as classifier and evaluates the mining performance of the ranked

feature set at increasing set size. For different cardinalities, the subset which maximizes the mRMR measure is selected and the performances of FLD trained on this subset are computed. The best cardinality and consequently the best subset is chosen as the one performing the minimum FLD misclassification error. Typically the best cardinality is smaller than the maximum number of features because the classifier over-fits on training data given by a large number of features.

Since hybrid FS algorithms require both a time-consuming search for ranking and several training iterations, FLD was selected to speed up selection experiments. For the selected subset the MIHFS algorithm produces a ranked list of features which highlights their predictive skill. A previous analysis performed on ground truth samples has shown the reliability of MIHFS technique in discarding both irrelevant and redundant features. Obviously, any kind of preprocessing of ultrasound images (such as deconvolution) could modify the features ranking and, in general, can improve or worsen the discriminating power of a feature.

## **2.2 Feature-based segmentation**

CAD techniques relying on imaging are typically based on the segmentation of the given image into ROIs and this task is usually performed by an expert radiologist. In order to make image classification independent of radiologists skills and give them an additional tool for making diagnosis, ROI segmentation may be possibly automated.

The use of rectangular and possibly overlapping ROIs makes the automation of segmentation very easy. Nevertheless irregular shaped ROIs may be more interesting, since they can outline anatomical regions or abnormal structure present in ultrasound image. Segmentation into irregular shaped ROIs is less easy to automatise, but still some algorithms based on ultrasonic features are able to automatically highlight significant zones.

The ROI segmentation procedure proposed in this work consists of two phases: feature calculation and image segmentation. In the feature calculation phase the RF signal is used to compute spectral, statistical and textural features over a rough selection of the prostate gland that can be performed also by non trained personnel. These features are then used in the segmentation procedure, based on an unsupervised Bayesian learning technique, to



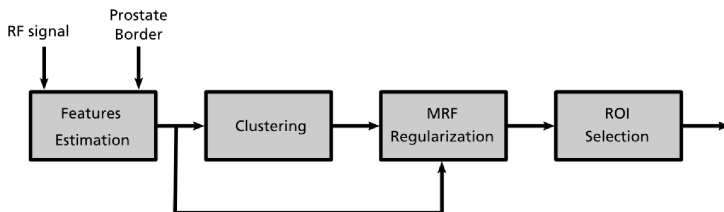


Figure 2.1: Schematic representation of the proposed procedure for computer aided classification of TRUS images.

assign each image pixel to a class. The assignment is realized by a K-means clustering followed by an iterative Bayesian regularization algorithm. Such algorithm imposes spatial smoothness constraints to the solution, minimizing the energy function derived by modeling the signal features with a multivariate Gaussian and the pixel segmentation class with a Markov random field (MRF). The contiguous pixels forming a closed region are then identified as a ROI.

Ultrasound image segmentation is strongly influenced by signal quality as well as by the tissue imaged. The characteristic artifacts of ultrasound images like attenuation, speckle, shadows and signal dropouts, missing boundaries due to the orientation dependence of acquisition, make the segmentation task complicated. While in other areas of medical imaging (CT, MRI) application of general image processing methods is sufficient, in the case of ultrasound signals more complex and specialized methods are needed in order to obtain satisfactory results.

In literature, there are a large number of papers describing segmentation procedures applied to ultrasound images. For a complete review of the state of the art of medical and ultrasound image segmentation methods see [3] and [4].

Most of the techniques presented in literature concern prostate boundary segmentation, while only few works about ROI segmentation have been published. Moreover, most of the methods depend on human interaction and their performances are user dependent. Therefore, to overcome the limitations of the existing procedures, new procedures should ideally be: user independent, independent on training images, focused on ROI segmentation.

According to the review on the published methods for prostate segmentation and tissue characterization, relying on the paradigm

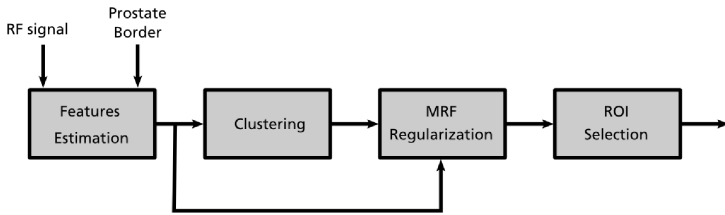


Figure 2.2: Schematic representation of ROI segmentation algorithm.

of fig.2.1 the procedure for ultrasound aided prostate cancer diagnosis is based on three steps: automatic segmentation of prostate into ROIs, feature estimation on the segmented ROIs, characterization of ROIs through a classifier.

The automatic ROIs segmentation procedure, proposed here, is composed of four sequential steps as summarized in fig.2.2: feature estimation, clustering, Markov random field based regularization and ROIs selection.

Since many boundary estimation algorithms are available in literature, a rough estimation of prostate boundary performed by automatic algorithm or done manually by non trained personnel is considered as given. Feature estimation is then performed pixel-wise only inside the prostate gland, by means of a sliding window. The estimation is done following a multi feature approach, i.e. different classes of features are estimated starting from the radio frequency signal before time gain compensation and envelope detection. B-mode features are estimated after envelope detection and log compression of the radio frequency signal. Among the features proposed in literature we employed texture parameters (Unser features [36]) and B-mode statistical modeling (Nakagami fitting [26]).

The estimated features are then used to segment the image by mean of a clustering procedure. Clustering techniques are unsupervised learning technique that easily allows the combination of multiple description parameters and data mining in a high dimension space, with a feasible computational cost [35]. In this work we adopted K-means algorithm, a based clustering method which evaluate the similarity of pixels by means of their distance in feature space. More precisely, pixel portioning is obtained by mini-

mizing the following energy function:

$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{C(i)=k} d(\mathbf{X}_i, y_k) \quad (2.5)$$

where  $C$  represents the clustering of all the data points  $i$  in  $K$  pre-determined number of clusters,  $C_i = k$  stands for the assignment of the  $i$ -th data point to the  $k$ -th cluster with center  $y_k$ , and  $d(.,.)$  is the distance.  $\mathbf{X}_i = \{X_{i,1}, \dots, X_{i,N_f}\}$  represents the  $N_f$ -dimensional feature vector of pixel  $i$ ; the whole feature data matrix is  $\mathbf{X}$ . The  $K$  means algorithm minimize the functional  $W(C)$ , with a two steps iterative procedure: for a given clustering  $C$ ,  $W(C)$  is minimized with respect to cluster centers  $\{y_1, \dots, y_k\}$  yielding the mean of the points currently assigned to each clusters; given a current set of means  $\{y_1, \dots, y_k\}$  the function  $W(C)$  is minimized assigning each data point to the nearest cluster center  $y_k$ , with respect to distance  $d(.,.)$ . Such procedure is repeated until convergence is reached.

The main drawback of clustering techniques is the lack of spatial information, which in the case of ultrasound images causes noisy segmentations, with irregular jagged-edges, wrong classified pixel and small isolated ROIs. Such issue can be mitigated applying a MRF based regularization algorithm on the segmentation obtained with the clustering procedure, as proposed in [37].

Image segmentation is formulated as a *maximum a posteriori* (MAP) estimation problem, involving the estimation of the estimated image  $\Sigma$  by maximization of  $P(\Sigma|\mathbf{X}) \propto P(\Sigma)P(\mathbf{X}|\Sigma)$ . The data to be estimated is the assignment of each pixel  $i$  to a class  $k$ , indicated as  $\Sigma_i = k$  and is modelled through the *a priori* probability  $P(\Sigma)$  as a Gaussian Markov random field on a 2D rectangular grid. The observed feature  $j$  in pixel  $i$ ,  $X_{i,j}$ , belonging to class  $k$  is modelled through the conditioned probability  $P(\mathbf{X}|\Sigma)$  as a Gaussian distribution with mean value  $\mu_{i,j}^k$  and standard deviation  $\sigma_{i,j}^k$ .

Since the MAP estimation by maximization of the true  $P(\Sigma|\mathbf{X})$  is computationally infeasible, a complexity reduction is mandatory. One of the most used suboptimal minimization algorithm is the Iterated Conditional Modes (ICM), first proposed in [37] and applied to different medical image segmentation problems. The ICM minimizes the MAP estimator no more on the entire grid, but pixel-by-pixel. In other words, we have to minimize for every pixel with

respect to the assignment  $\Sigma_i = k$  the functional:

$$U(\Sigma_i|\mathbf{X}) = \sum_{j=1}^{N_f} \left( \log(\sigma_{i,j}^k) + \frac{(X_{i,j} - \mu_{i,j}^k)^2}{2(\sigma_{i,j}^k)^2} \right) + \sum_{\delta_i} V(\Sigma_i) \quad (2.6)$$

where  $\delta_i$  is the neighbourhood system centered on the pixel  $i$  and  $V$  is the MRF potential function defined over a second order neighbourhood system. Intuitively the first term of (2.6) imposes that class intensity is close to observed data, while the second term is a regularization constrain suggesting the algorithm to associate the same class to close pixels. This technique requires the minimization of several simple functions, rather than the optimization of a single, much more complex, functional. This results in an iterative algorithm which drastically reduces computational requirements, and provides a local optimum very close to the global one, if initialized with  $K$ -means segmentation.

The algorithm was implemented following the adaptive methodology proposed in [37]: conditional mean and conditional variance are estimated over windows whose size is reduced after a certain number of iteration until algorithm convergence is reached, i.e. either the number of modified pixels is lower than a fixed threshold or the maximum number of iteration has been reached.

Finally, the different clusters are examined and further divided according their spatial distribution: all the pixel that forms a connected and closed area are identified as a single region of interest. If ROI of a minimum size are required, all the regions smaller than a fixed number of pixel are reassigned to the adjacent regions applying a morphological operations on them.

### 2.3 Deconvolution as pre-processing step

An important step in data preparation for analysis is the compensation for system-dependent effects. In this work, this problem is faced through the use of deconvolution, which is able to reduce noise and increase both contrast and quality of US images. Although deconvolution is a well known topic, its usefulness in CAD schemes was rarely explored.

One of the limitations on image quality is due to the blurring effect on the back-scattered echoes produced by the transducer's point spread function. The acquired RF-signal  $x[n]$  can be considered to be the result of the transducer's impulse response  $h[n]$  con-

volution with the tissue reflectivity function (or tissue response)  $w[n]$ :

$$x[n] = w[n] * h[n] \quad (2.7)$$

Given the imaging system's output  $x[n]$ , the deconvolution algorithm produces an estimate of the tissue response  $w[n]$ .

Standard techniques for *in vivo* US signal deconvolution require the knowledge of system PSF and are based on regularized inverse solution of (2.7), obtained imposing  $L_1$  or  $L_2$  norm constraints on the solution [38]. The system impulsive response  $h[n]$  *in vivo* can not be properly estimated on phantoms because tissues cause unpredictable phase aberration, non linearity and dispersive attenuation. A minimum phase version of system PSF can be estimated from the observed signal through homomorphic deconvolution techniques [39]. The main drawbacks of these techniques are the high sensitivity of the obtained solution from the estimated PSF, which often results in image artifacts [40], loss of information and high computational cost.

Predictive deconvolution techniques that are used as standards in astronomy and geophysics are less popular in medical ultrasound. Although blind deconvolution techniques are intrinsically more limited, their choice for tissue characterization is due to several aspects: they do not require PSF estimation, they are simpler to implement and have low processing time. Furthermore, blind techniques based on linear prediction perform a local whitening of the signal, taking into account time-varying PSF, which is a realistic situation in ultrasound due to the variable focusing and pulse attenuation.

In this work, techniques like Wiener deconvolution with minimum-phase PSF estimation [41] have been compared to the linear prediction technique. We verified that Wiener technique provides over-smoothed solution and features extracted from deconvolved images are not significant for tissue characterization. On the contrary, predictive deconvolution [42] based on linear prediction is able to highlight diagnostic significance of extracted features.

Predictive deconvolution uses a predictive filter to discard any deterministic aspects from  $x[n]$  and consider the unpredictable part, that is the error respect to real values of  $x[n]$ , as the restored tissue response. Thus it is necessary to estimate the coefficients of a filter  $\epsilon_k$  to predict future values of  $x[n]$ :

$$\sum_k \epsilon_k x[n-k] = \hat{x}[n+\alpha] \quad (2.8)$$

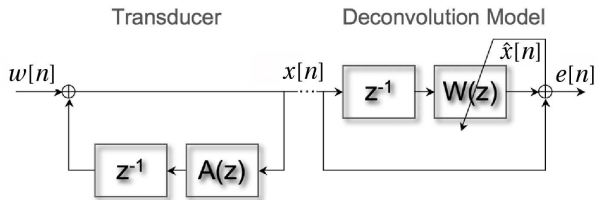


Figure 2.3: Autoregressive convolutional model and predictive deconvolution iterative algorithm.

with  $\hat{x}[n + \alpha]$  being an approximation of  $x[n]$  after  $\alpha$  samples. The non predictable part of the input signal is represented by the error:

$$\begin{aligned} e[n + \alpha] &= x[n + \alpha] - \hat{x}[n + \alpha] \\ &= x[n + \alpha] - \sum_k \epsilon_k x[n - k] \end{aligned} \quad (2.9)$$

The Recursive Least Squares (RLS) algorithm [43] provides a recursive way to compute the filter which minimizes the weighted least squares error function:

$$E(\boldsymbol{\epsilon}, n) = \sum_{i=1}^n \lambda^{n-i} |e[i]|^2 \quad (2.10)$$

The information of new data updates the old estimate and the forgetting factor  $\lambda$  decreases the weight of data in the distant past to follow the statistical variations of the observable data. The imaging system model and corresponding deconvolution model are visualized in figure 2.3. It can be shown that the output of the deconvolution process, converges to the tissue response  $w[n]$  if the following hypotheses are fulfilled: *i) Feedback hypothesis*:  $h[n]$  has an all pole transfer function, *ii) Random hypothesis*:  $w[n]$  is a white Gaussian noise process, i.e. the RF signal  $x[n]$  is modeled as an autoregressive process [44].

Other deconvolution algorithms such as inverse filters as in [45], or wavelet-based approaches as in [46] and [47] were considered and tested but the RLS prediction resulted to be the best compromise in terms of classification performance and computational cost.

## 2.4 Feature-based classification

The last stage of the proposed method is a supervised classifier. In this phase, the selected ROIs are classified by means of a set of features computed from the regions outlined by these ROIs. The result of the classification will be displayed over the standard B-mode image: position and extension of all cancerous ROIs are marked with false colors. Such information can be used by radiologist in guiding biopsy protocols.

Previous studies reported that features of different nature hardly are linearly inter-correlated and that a nonlinear classification model can be able to extract valuable information from a mixed feature set and to reach higher level of accuracy [25]. On the other hand, complex model classifiers fail in preserving physical significance of features and their true dependence on pathology.

As regards tissue characterization based on irregular ROIs, among the possible supervised machine learning techniques a Support Vector Machine (SVM) classifier was adopted [8]. As reported in the literature, nonlinear classifier seems to be preferable for prostate tissue characterization and SVM have been proved to be a good choice. In order to obtain a nonlinear SVM classifier a Gaussian radial basis function was adopted as kernel. All the simulations were performed using the publicly available software implementation of SVM classifiers (svmlight [48]).

Likewise, in the analysis of deconvolution in tissue characterization, a nonlinear classification model was designed and trained to discriminate between healthy and cancerous zones. The studied classifier is based on a first nonlinear feature extraction step and a second linear classification step. Empirical results on the analysed dataset lead to choose the Generalized Discriminant Analysis (GDA) as feature extraction algorithm. This technique [49] uses kernel transformation to obtain a new feature space  $F$  that is related to the former by a nonlinear mapping  $\phi$ , performed on feature vector  $\mathbf{x}$ :

$$\phi : \mathbb{R}_f^N \rightarrow F \quad \mathbf{x} \rightarrow \phi(\mathbf{x}) \quad (2.11)$$

The subsequent classifier is a FLD that finds the best hyperplane separating healthy and unhealthy samples of the training set. The resulting new GDA features,  $\mathbf{z}$ , are projected in the direction which maximizes the between-class distance and minimizes the within-class distance of samples of the mapped training set. This is a linear operation realized by simply combining mapped

feature vectors:

$$d(\mathbf{z}) = \langle \mathbf{w}, \mathbf{z} \rangle + b \quad (2.12)$$

where  $b$  is an offset chosen to impose the same distance between centroids of the two classes of samples and hyperplane, while the weight coefficients  $\mathbf{w}$  are found by maximizing the following Fisher criterion:

$$D_F(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \quad (2.13)$$

In the previous expression  $S_B$  is the between-class scatter matrix, while  $S_W$  is the within-class scatter matrix, referred to the remapped samples of the training set.

This maximization involves the computation of dot products  $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$  in  $F$ : such calculation can be performed efficiently by choosing kernel functions  $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$  which act like dot products in the re-mapped feature space [9]. The choice of these kernel functions is crucial. In this study different kind of kernels (namely polynomial, gaussian and sigmoidal kernels) were tested. The best results were obtained when a Gaussian kernel is applied.

The effect of GDA feature projection is compared to the results of a simpler linear (i.e. without nonlinear mapping) discriminant analysis (LDA) in figure 2.4. In particular, the histograms in figure 2.4(a) show the distributions of healthy (in dark) and unhealthy (in white) samples after LDA projection obtained combining 5 features, while figure 2.4(b) shows the GDA projection of the same samples: the improvement in separation is evident.

This classification system was compared to different mining techniques, both linear and non linear (regularized least squares, elastic net, SVM), and was selected because the computational time required for its training is significantly smaller with respect to the other tested classifiers, while providing the same accuracy. A scheme of the whole classification procedure of an ultrasound image is provided in figure 2.5.

## 2.5 Ground truth database

In this study the analyzed TRUS images of prostate sagittal section are acquired by a standard US System Esaote Megas<sup>TM</sup> with trans-rectal transducer EC123 at sampling frequency of 50 MHz and central frequency of 7.5 MHz. Through this commercial US equipment (combined with FEMMINA [50]) it is possible to have direct access to RF echo signal before time gain compensation sampled



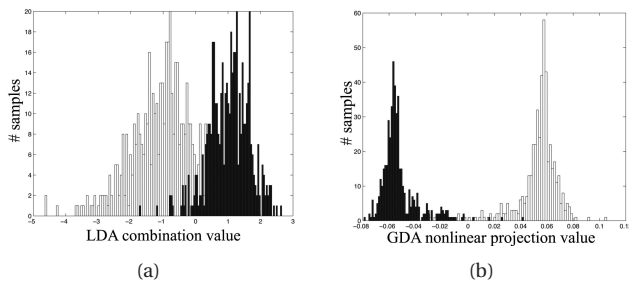


Figure 2.4: Comparison between linear and non linear projection on real data. figure 2.4(a) and figure 2.4(b) show the histograms of healthy (in dark) and unhealthy (in white) samples (5-feature vectors) after LDA and GDA projection respectively

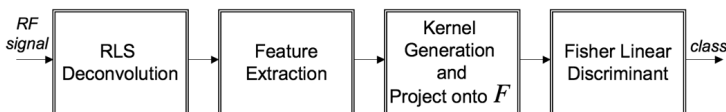


Figure 2.5: Classification Scheme: processing steps from RF signal measurements to the derivation of final classification label.

with a resolution of 12 bits. The RF signal is essential to compute some spectral parameters, useful to characterize prostate tissues.

The clinical study involves 37 patients with a high level of PSA and undergoing their first biopsy. The age of patients ranges from 53 to 75 years. For each patient 10 US image frames were produced with probe held in a fixed position. Since the frame rate is about 20 fps, all the frames provide the same information unless for the measurement noise and micro-movements of radiologist's hand driving the probe. Thus in this analysis only the first frame is considered, although the other frames can be employed in the study of stability of features and possibly to extract features averaging on several frames. The available dataset for this investigation consists of 15 benignant cases (normal and hyperplastic tissue) and 22 malignant cases (presence of adenocarcinoma).

The ground truth comes from histopathological analysis from tissue samples acquired through biopsy. The matching of biopsy results on TRUS images is performed by a single expert radiologist, using the exact information of needles location in the images. For

each case the radiologist has matched biopsy results with precise regions on TRUS images: both pixels found to be part of cancerous zones and pixels belonging to healthy regions are selected and labeled in every zone examined through biopsy.

Cancer tissue which has not been sampled during this procedure will be labeled as healthy, thus the accuracy of this process depends on the detection rate of biopsy, gold standard in this study.

In reality, unless radical prostatectomy is performed, with high probability the portion of tissue sampled during biopsy is smaller than the true cancer extension, the available ground truth must be considered as an incomplete labeling. The learning problem is thus not fully supervised and specificity of the classification can't be evaluated in a reliable way.

However, since the main goal of computer methods for prostate cancer diagnosis is the identification of all the cancer foci for guiding biopsy, possible overestimations of the size of cancerous region does not affect the clinical performance of the computer aided diagnosis procedure. Thus, the main goal is the development of a high sensitivity classification method which is able to identify all the cancer foci with an average level of specificity.

As regards the first study about segmentation, the whole images database is segmented using the method described in section 2.2, obtaining about 300 ROI of irregular size per image.

On the contrary, in the analysis of the proposed CAD scheme involving predictive deconvolution, described in section 2.3, the images of  $2500 \times 96$  pixels are segmented in rectangular windows.

The choice of window size is a tradeoff between statistical significance of samples in a ROI used to compute features and the resolution of the output feature image. In this study ROIs are rectangular windows of  $101 \times 7$  pixels and of about  $9\text{mm}^2$  area.

From each image about 3000 ROIs are generated from the prostate zone and used for feature extraction.

The pairs ROI-label constitute the ground truth for both segmentation algorithm validation and classifier training. The evaluation of features is performed over ROIs lying inside the zones of different class; ROIs centered on the border of two different zones are not considered in the analysis.

All the examples of benignant ROI are taken only from the benign cases, because there is no sure knowledge about benign regions from malign glands not examined by histology. Furthermore, only the ROI marked by radiologist as pathological are used as ex-

amples of cancerous regions in the learning phase.

## 2.6 Experimental results

### Feature selection

Beside Unser textural features, we adopted also Haralick and Fractal textural features [36], radio frequency signal statistical modeling (generalized Gaussian [26]) and spectral parameters (central frequency, midband and slope [51], polynomial fitting of wavelet spectrum [52]). As described in section 2.1, since the complete feature set of all parameters mentioned before would have a huge dimensionality of about 140 attributes, a first selection step is performed, keeping for each group of features only those highly correlated to the ground truth class, and so to the pathology, and discarding the other ones. In this application a hybrid feature selection algorithm, MIHFS, is used to rank and prune the initial feature set, as explained in section 2.1.

In the study based on the dataset described in section 2.5, MIHFS algorithm selects a feature subset whose cardinality is equal to 10.

Feature extraction performed after deconvolution yields to different distributions of feature values. As a consequence, MIHFS performed on the post-deconvolution feature set provides a new ranked list and a different optimal subset.

FS output for the case without preprocessing and the case with predictive deconvolution are given in table 2.2, presenting the features from the most to the less important. Listed attributes in the selected feature sets are referred to the groups described in table 2.1. A brief description of used features is provided in appendix A.

Table 2.2: Ranked Feature Sets

Rank	No Preprocessing	RLS Deconvolution
1	Variance-Fractal(2) Alfa1	Mean-Intercept2 AR(3) lmsd
2	Variance-WDES Diff. Proj. n.8	Variance-Fractal(2) Alfa2
3	Mean-Intercept AR(2) burg	Mean-Nakagami logm
4	Mean-correlation 135	Mean-correlation 45
5	Mean-Nakagami logm	Variance-Fractal(2) Beta 8
6	Variance-Fractal(1) Beta 10	Mean-Haralick Sum of Squares
7	Variance-Fractal(2) Beta 10	Mean-contrast 90
8	Mean-Naka w Diff. Proj. n.8	Variance-Fractal(1) Beta 10
9	Mean-Nakagami logOmega	Mean-Nakagami logOmega
10	Mean-homogeneity 90	Mean-Naka w Diff. Proj. n.8

MIHFS results show that RLS deconvolution improves the predictive skills of textural features and worsens the abilities of some spectral parameters. As a matter of fact: *i)* deconvolution causes good speckle reduction aimed at decreasing textures dependence on the acquisition system and allows an easier edge detection, rising textural pattern significance; *ii)* some spectral features like the high frequency diffuse component extracted from RF signals partly lose their diagnostic value because deconvolution acts as a whitening filter, therefore these features must be computed before PSF compensation; *iii)* the diagnostic significance of Nakagami features is assured in both cases, with and without predictive deconvolution.

### **Irregular ROI-based tissue characterization**

Concerning the study about irregular ROI segmentation, as ROI have different sizes, the classification results must be normalized in order to take into account the size of each ROI. Therefore, classification performance is evaluated in terms of correctly classified pixels inside the prostate border. As stated in section 1.2, typical measures to assess classifier performances are sensitivity, specificity, accuracy defined here pixel-wise: TP are true positive pixels, TN are true negative pixels, FP are false positive pixels and FN are true negative pixels. Pixel-based criteria instead of ROI-based criteria are necessary when irregular regions are involved, since ROI can contain at the same time healthy and unhealthy ground-truth pixels, thus, unless a threshold is defined, the class of the whole ROI is ambiguous.

The average classification performances obtained by the proposed procedure on the dataset are resumed in table 2.3. In this study we employed the ranked feature list of 54 attributes provided from MIHFS algorithm.

The classification results show that the first 20 features are the most relevant: while increasing the number of features from 10 to 20 a significant performance improvement is noted, only slightly improvement are obtained when the complete set of feature is used. Results confirm that a multi-feature approach is fundamental.

The high value of sensitivity obtained (always > 80%) suggests that the proposed method is able to identify most of the known cancer locations. As said before, since only partial information about the real extension of cancerous areas is available and since classification was performed on the whole image, the specificity

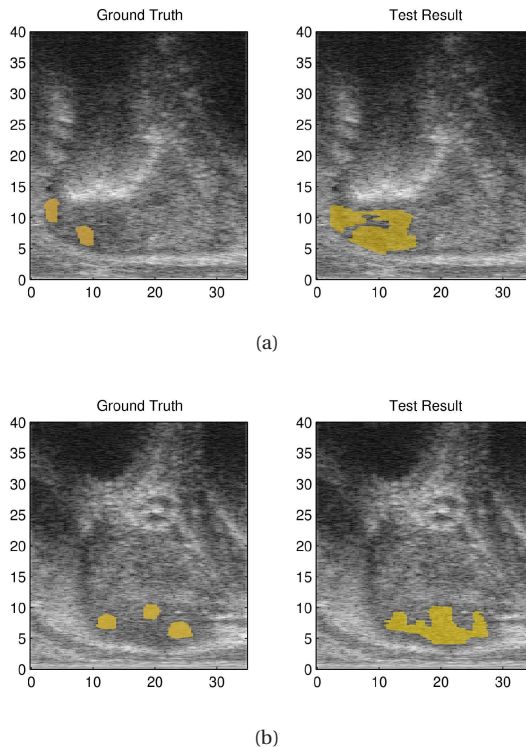


Figure 2.6: Examples ground truth reference images (left panel) and classification results displayed above B-mode prostate sagittal scans as visual guidance for biopsy (right panel).

value obtained is not fully reliable and it gives just quantitative information about the method performance. However, analysing the results obtained on the benign cases we observed that the number of false positive is extremely low and we can therefore conclude that the low specificity value is due to the over estimation of cancer size in malignant images. Since the classification results are meant to be used for guiding biopsy, an over estimation of the tumour size is less problematic than false positive on benign cases, which would causes unnecessary additional biopsies. Finally, in figures 2.6(a) and 2.6(b) examples of the classification results displayed above the B-mode scans as visual guidance for biopsy are given.

Table 2.3: Classification performance of the proposed tissue characterization method

Settings	SE	SP	Acc
10 features	0.80	0.71	0.75
20 features	0.93	0.73	0.87
54 features	0.95	0.74	0.89

### Deconvolution in tissue characterization

In this study the classification model is trained on 1000 ROIs, sampled randomly from 18 TRUS images (7 benign and 11 malignant). Training is performed through stratified 10-fold cross validation, so the classification model is generated from 900 samples and validated on the remaining 100 samples. The testing set consists of the other 19 images, completely unknown to the classification model.

Classification performance is computed through ROI-based metrics on a total of 58.602 testing ROIs, where 58.286 ROIs are healthy and 316 are unhealthy. Examples of classified images and their corresponding ground truth are shown in figure 2.7. These TRUS images show the axial section of a prostatic gland without scan conversion. According to the gland anatomical shape shown in figure 1.6 of section 1.3, the images in figure 2.7 show a prostate in the same orientation, where the superior hyperechoic boundary marks the interface between the rectum (indicated with “R” in figure 2.7) and the peripheral zone (PZ) of the prostate.

Classification output is visualized directly on the original TRUS image, by means of transparent colored patches located over the ROIs. Ground truth images are visualized on the left side of each image pair in figure 2.7, where healthy regions are marked in yellow and cancerous regions in red. The classification output is shown on the right side of each image pair in figures 2.7: unhealthy classified ROIs are covered in red, while healthy classified regions are marked in blue. Red and blue colors move toward violet and cyan respectively when the discriminant function value is near zero, pointing out zones where classification is more difficult. A color bar is added to the side of each images set to support an easier reading of classification results. figure 2.7(a) represents a benign case, while the others pictures show glandes affected by car-

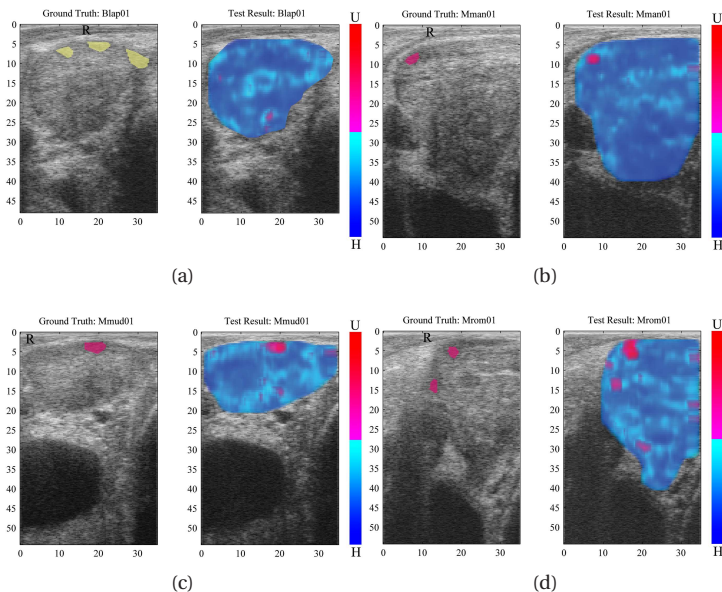


Figure 2.7: Examples of ground truth and classified images. The letter “R” marks the location of rectum, while the color bar shows the colors used to mark zones classified as healthy (H) and unhealthy (U). The gradients indicate the discriminating function values and, thus, the colors next to the threshold between the two classes mark regions where classification is more difficult. figure 2.7(a) represents a benignant case, while figure 2.7(b), figure 2.7(c) and figure 2.7(d) show malignant cases.

cinoma. Since some false positives are always present, also in the best case and always outside biopsy region, the positive predictive value is low, but this is due to the low prevalence of unhealthy regions respect to healthy regions in a ROI-based metrics. Criteria independent on disease prevalence are employed to evaluate this clinical test.

In order to assess classifier performances, mean value and standard deviation of sensitivity, specificity, accuracy and area under the ROC curve ( $A_z$ ) were estimated over 10 different experiments, where images in the training and test set are randomly selected. In table 2.4 performance measures related to the GDA-FLD classifier with and without RLS deconvolution preprocessing step are collected (mean value  $\pm$  standard deviation). Each table row refers to a classifier trained on a different number of features from 5 to 54.

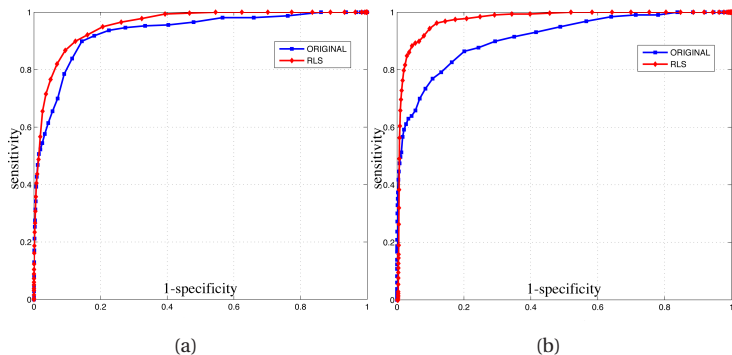


Figure 2.8: ROC of classification with and without RLS predictive deconvolution. figure 2.8(a) and figure 2.8(b) show the performance for a 5 features and 10 features classification model respectively.

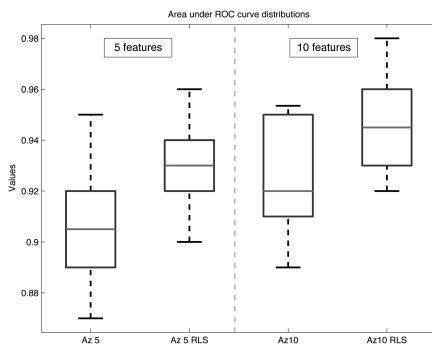


Figure 2.9: Boxplots of distributions of ROC area for classification without preprocessing (In figure: *Az 5* and *Az 10* when using 5 and 10 features respectively) and classification after RLS preprocessing (In figure: *Az 5 RLS* and *Az 10 RLS* when using 5 and 10 features respectively).

Classification without deconvolution shows sensitivity ranging between 0.51 and 0.72, specificity and accuracy ranging between 0.89 and 0.97, and ROC area ranging between 0.87 and 0.92. Classification performances decrease as number of features increase, when more than 10 parameters are employed. This indicates an over-fitting on training data when a high number of attributes is used. The addition of predictive deconvolution improves perfor-



mances of 3% in terms of ROC area when 5 or 10 features are used, 1% and 4% for the case with 20 and 54 features respectively, where the effect of overfitting is evident. In the best case classification performances after RLS preprocessing provide a value of the ROC area of 98%. These results are better visualized in figure 2.8, where ROC curves of the two classifiers in the best case are compared when involving 5 and 10 features.

The improvement is visible in the shifting of Az distributions mean value toward higher values from the case without preprocessing to the case with RLS preprocessing. Az distributions (5 and 10 features) are shown in figure 2.9 for the cases with and without RLS preprocessing.

Table 2.4: Classifier Performance

#	No Preprocessing			
	SE	SP	Acc	Az
5	$0.72 \pm 0.08$	$0.89 \pm 0.02$	$0.89 \pm 0.02$	$0.91 \pm 0.02$
10	$0.69 \pm 0.06$	$0.94 \pm 0.02$	$0.93 \pm 0.02$	$0.92 \pm 0.02$
20	$0.64 \pm 0.07$	$0.95 \pm 0.01$	$0.95 \pm 0.01$	$0.92 \pm 0.02$
54	$0.51 \pm 0.06$	$0.97 \pm 0.01$	$0.96 \pm 0.01$	$0.87 \pm 0.03$
#	RLS Deconvolution			
	SE	SP	Acc	Az
5	$0.79 \pm 0.05$	$0.89 \pm 0.02$	$0.89 \pm 0.02$	$0.93 \pm 0.01$
10	$0.75 \pm 0.09$	$0.93 \pm 0.01$	$0.93 \pm 0.02$	$0.95 \pm 0.02$
20	$0.72 \pm 0.07$	$0.94 \pm 0.01$	$0.94 \pm 0.01$	$0.93 \pm 0.03$
54	$0.62 \pm 0.10$	$0.96 \pm 0.01$	$0.96 \pm 0.01$	$0.91 \pm 0.02$

The overall result of the introduction of predictive deconvolution is an improvement of classification performances with an addition of complexity that can be negligible with respect to feature computation complexity.

These processing times are still far from being real-time, but further work can be done to speed up the whole classification procedure:

- use C or CUDA<sup>TM</sup> implementation of spectral and textural features
- define larger “natural” regions of interest, based on textural characteristics of the prostate, for example with the algorithm proposed in section 2.2. The additional step for irregular ROI definition does not bring extra computational complexity, as the cost saved in feature computation due

to the analysis of a smaller number of ROIs is more important than the supplementary cost of segmentation. Thus this step would provide a reduction of both feature generation and classification computational time.

In table 2.5 a summary of previously published methods for prostate tissue characterization (discussed at the beginning of this chapter) is visualized in order to show research developments in computer-aided cancer detection and as a comparison with the study presented here. A more detailed descriptions of the state of the art methods is presented in appendix B.

Table 2.5: Published methods for ultrasound-based prostate tissue characterization

Work	Ground Truth	Technique	Results %			
	# patients		Features	SE	SP	Acc
Basset [16]	16	Textural	83	71	-	-
Huynen [17]	51	Textural	80	88.20	-	-
Houston [18]	25	Textural	73	86	80	-
Schmitz [19]	33	Multi	82	88	-	-
Scheipers [20]	100	Multi	-	-	75	86
Feleppa [21]	200	Spectral	-	-	80	85
Mohamed [22]	20	Textural	83.3	100	93.75	-
Llobet [53]	300	Textural	68	53	61.6	60.1
Mohamed [23]	20	Multi	83.3	100	94.4	-
Han [24]	51	Multi	92	95.9	-	-
HistoScanning <sup>TM</sup> [54]	29	Statistical	95	-	-	-

## 2.7 Conclusion

Prostate cancer is a common disease among men and early stage detection is crucial for a successful pathology treatment. The current clinical procedure for diagnosis and grading of the disease, i.e. biopsy and histopathological images analysis, is still matter of controversy because it is invasive, relatively dangerous and affected by errors due to the sampling process and the multi-focal nature of prostate cancer. In this context the availability of a CAD tool would be very useful to physicians to support their decision and detect cancer even in its early stages.

TRUS images based CAD systems can be efficient tools to perform more objective prostate cancer diagnosis reducing inter operator variability. The use of ultrasound technology for prostate cancer detection is motivated by its minimal invasiveness and cost. Furthermore, TRUS technique is already integrated in the standard procedure for prostate examination.

This study reported tissue characterization based on a multi feature approach, where spectral, statistic and textural attributes were extracted from TRUS images in the prostate zone and used for ROI classification. A hybrid feature selection algorithm exploits mutual information between features and pathological state of ground truth tissue to reduce feature set dimensionality, discarding irrelevant or redundant attributes. In the proposed CAD scheme the resized feature set is processed by a Gaussian re-mapping, shifting the problem in a sub dimensional space, where linear classification is more effective.

Interesting results were obtained by employing both rectangular regions of interest and irregular zones, selected by automatic segmentation algorithms discussed in this thesis. This work was also meant to explore the utility of deconvolution for classification purposes and to provide a comparison between non linear classification performance with and without deconvolution pre-processing. To this aim, a predictive deconvolution is applied to ultrasound data before feature computation, producing an average increase in classification performance of about 3% in terms of area under the ROC curve.

The proposed classification procedure was designed and tested on a 37 images ground truth and provides a CAD system performing a highly accurate detection inside the prostate zone.



# Three

---

## A CARMA MODEL FOR ULTRASOUND SIGNALS

---

*For my part I have sought liberty more than  
power, and power only because it can lead to  
freedom.*

---

Memoirs of Hadrian (1951)

IN ultrasonic tissue characterization several kinds of information may be inferred from the RF signal by computing some specific characteristics as textural or spectral features. One more way to extract information is to find a suitable model for the echo signal and to estimate its parameters.

Modelling the ultrasound RF signal allows to define new features, the model parameters, which may be dependent on the acquisition setting and/or on the tissue crossed during radiation, thus providing interesting clues to reduce acquisition system effects or to type tissues.

The autoregressive moving-average (ARMA) model, which is classical for time-series modelling, has been employed in a shift-variant fashion to model ultrasound signal and to perform blind deconvolution [6], as also proposed in section 2.3, where the signal is supposed to be shift-variant auto-regressive.

The ARMA model is suitable to model a process depending on a series of unobserved shocks as well as on its past behaviour. Since in ultrasound images the PSF is spatially variant, ultrasound signals can be modelled as ARMA processes, provided that ARMA parameters are allowed to change over time, i.e. over the axial direction.

An easy way to deal with nonstationary models in ultrasound is to assume that the spatial variability of the PSF is slowly varying, hence well approximable by a piecewise constant function. In this

case, the image can be divided into segments, whose size is chosen to be small enough to guarantee that each of the segments is formed by a stationary convolution with a different PSF. Each segment can thus be modelled as a different standard ARMA model.

An ARMA process  $x[n]$ , driven by white noise  $w[n]$  with variance  $\lambda^2$ , is defined as:

$$x[n] + \sum_{i=1}^n c_i x[n-i] = w[n] + \sum_{j=1}^m d_j w[n-j] \quad (3.1)$$

The signal  $x[n]$  can thus be seen as the output of the linear time-invariant casual system  $H(z)$ , excited by a white noise input:

$$H(z) = \frac{\sum_{j=0}^m d_j z^{-j}}{\sum_{i=0}^n c_i z^{-i}} \quad (3.2)$$

An Autoregressive (AR) process is a special case of an ARMA process, not depending on its own past samples and, thus, corresponding to an all-pole transfer function. There are several estimation techniques for AR and ARMA parameters in literature, mostly based on prediction error [55].

In this chapter, a continuous-time ARMA (CARMA) model is proposed for ultrasound signals. Since all the available information is discrete, estimation techniques for CARMA parameters are based on sampled data, constituted in this case by the RF signal. Using a continuous-time model for ultrasound signals means employing some function to perform interpolation of the available sampled RF signal and move this way to the continuous-domain.

In the next section, continuous-time AR (CAR) and CARMA models will be presented, then a new approach for CARMA identification from sampled data, based on exponential splines [56, 57] is proposed. The presented identification technique is very efficient also in some aliasing conditions. A comparison study on simulation data is described to show the accuracy of the proposed approach with respect to an algorithm based on the interpolation of discrete data by polynomial splines. CARMA parameters can hence be estimated from RF signal and used for its characterization.

A second, simpler algorithm is further proposed to actually extract CARMA parameters from ultrasound signals, assuming high frequency sampling, which is usually the case in ultrasounds acquisitions. Finally an analysis of these parameters is performed to assess their skill in capturing information about the concentration of scatterers in ultrasound signals.

### 3.1 CAR and CARMA models

CARMA processes are widely used in control theory and in signal/image processing and analysis. Typical examples of applications are system identification and adaptive filtering [55, 43], speech analysis and synthesis [58], stochastic differential equations and image modeling [59, 60, 61]. In practice, the available data is discrete and one is usually required to estimate the underlying continuous domain parameters from sample values. Potential examples are continuous domain structure modeling of physical phenomena; linear time invariant (LTI) system identification as well as numerical analysis of differential operators.

The sampled version of a CARMA process is a discrete-domain ARMA process whose zeros and poles are coupled in a non-trivial way [62, 63, 64, 65]. To overcome this difficulty, currently available methods use various types of approximations. Recent works on this subject are classified into two broad categories: direct and indirect [66, 67]. Direct methods consist first of parametrizing a discrete-domain model by the continuous domain parameters. The discrete-domain model is then used to minimize a cost function that involves the available data. In this way, the required continuous domain parameters are directly estimated by the minimization process. An example of such a method would be replacing derivative operators by finite difference operations and replacing the continuous domain white noise by a discrete-domain white noise [68, 69, 70, 71, 72].

In [73], a continuous-time AR model was recast into a discrete-time linear regression formulation rather than into a discrete-time ARMA process. The discretization of the continuous-time white noise was considered then to be part of the regression error. This error also includes contributions from the weighted finite difference approximation. As the least square solution of the regression problem might result into a biased estimation, the authors of [73] suggest two methods for reducing the bias effect. This was done by imposing constraints on the finite difference weights or, alternatively, by compensating for the bias at the final stage of the estimation process. The advantage of these two methods is that they require no shifting of the data when approximating the derivative values, giving rise to a reduced computational complexity over other least square methods.

In [74] it is suggested to parametrize the autocorrelation sequence of the sampled process by applying the Schur decompo-

sition numerical method to obtain a state-space representation of the discrete-domain process. The cost function that has been proposed there minimizes the  $\ell_2$  norm of the difference between a sampled version of the autocorrelation model and the autocorrelation sample values. Another example of a direct method consists of power spectrum parametrization [75, 76]. Indirect methods, on the other hand, rely on standard discrete-domain system identification methods such as prediction error minimization [55]. The discrete-domain system is then mapped into a continuous domain one while preserving some required properties, e.g. the bi-linear transform. Such mappings are common practice for digital filter designers, but alternative transformations are also available [77]. Current methods are sensitive to aliasing artifacts, meaning that they require the signals to be sampled at a sufficient rate.

Some mathematical conventions and notations that will be used throughout this work are provided in the next paragraph.

### Notations

The bilateral Laplace transform of a scalar function  $\varphi(t)$  is

$$\Phi(s) = \mathcal{L} \{ \varphi(t) \} (s) = \int_{-\infty}^{\infty} \varphi(t) e^{-st} dt, \quad (3.3)$$

where  $s$  takes complex values that satisfy  $\varphi(t) e^{-st} \in L_1$ . The inverse Laplace transform is denoted  $\mathcal{L}^{-1}$ . The Fourier transform of this function is  $\hat{\varphi}(\omega) = \mathcal{F} \{ \varphi(t) \} (\omega) = \Phi(j\omega)$ . The bilateral  $z$ -transform of the sequence  $\{ \varphi[n] \}_{n \in \mathcal{Z}}$  is

$$\Phi_d(z) = \mathcal{Z} \{ \varphi[n] \} (z) = \sum_{n=-\infty}^{\infty} \varphi[n] z^{-n}, \quad (3.4)$$

where  $z$  takes complex values for which the infinite sum converges. The Discrete Time Fourier Transform (DTFT) of the same sequence is  $\hat{\varphi}_d(\omega) = \Phi_d(e^{j\omega})$  and for sequences of length  $N$ , the Discrete Fourier Transform (DFT) is given by  $\hat{\varphi}_d[k] = \Phi_d(e^{2\pi jk/N})$ . The argument  $\omega$  denotes either radial frequency in units of [rad/time-unit] when used with the Fourier transform, or normalized frequency in units of [rad/sample] when used with the DTFT. continuous domain and discrete-domain convolution operations are denoted as  $*$ . The transpose of a matrix  $\Sigma$  is  $\Sigma^T$ , its inverse is  $\Sigma^{-1}$  and its determinant is  $|\Sigma|$ .  $\mathbb{E}$  denotes expected value.

The dependence of a function  $\varphi(t)$  or a sequence  $\varphi[n]$  on a set of parameters  $\theta$  is represented by a second input:  $\varphi(t; \theta)$ ,  $\varphi[n; \theta]$ .



The same notation is employed when the function or the sequence are expressed in the Fourier, Laplace or  $z$  domain:  $\hat{\varphi}(\omega; \theta)$ ,  $\Phi(s; \theta)$ ,  $\hat{\varphi}_d(\omega; \theta)$  and  $\Phi_d(z; \theta)$ .

### Model Description

A CARMA model that is driven by white Gaussian noise is specified in the time domain by the following expression:

$$A(D)x(t) = B(D)w(t), \quad (3.5)$$

where  $x(t)$  is the CARMA output signal and  $w(t)$  is white Gaussian noise with zero mean and variance  $\sigma^2$ ;  $A(D) = D^p + a_1 D^{p-1} + \dots + a_p I$  is the  $p$ -th order constant coefficient differential operator that acts on the output of the system (AR part), while  $B(D) = D^q + b_1 D^{q-1} + \dots + b_q I$  is the  $q$ -th order differential operator that acts on the Gaussian noise (MA part). In the special case of CAR processes  $B(D) = 1$ . In order to identify this stochastic system, we parametrize its power spectrum by introducing a vector  $\theta$  that consists of variance, poles and zeros of the system:

$$\theta = \{\sigma^2, r_1, \dots, r_q, s_1, \dots, s_p\} \quad (3.6)$$

Therefore, the CARMA process  $x(t)$  of order  $p$  is characterized by the power spectrum:

$$\Phi(j\omega; \theta) = \sigma^2 \left| \frac{\prod_{k=1}^q j\omega - r_k}{\prod_{k=1}^p j\omega - s_k} \right|^2 \quad (3.7)$$

where  $\mathbf{s} = (s_1, \dots, s_p)$  are the poles of the system and  $\mathbf{r} = (r_1, \dots, r_q)$  its zeroes. Stability of the system (3.7) is ensured if  $\Re\{s_k\} < 0$ . The one-to-one relation between the poles, the zeroes and the coefficients of the differential operators  $A(D)$  and  $B(D)$  allows one to express (3.7) as a function of real variables that can be used in the optimization process and  $\theta_o$  as  $\{\sigma^2, a_1, \dots, a_p, b_1, \dots, b_q\}$ :

$$\Phi(j\omega; \theta_o) = \sigma^2 \left| \frac{(j\omega)^q + b_1 (j\omega)^{q-1} + \dots + b_q}{(j\omega)^p + a_1 (j\omega)^{p-1} + \dots + a_p} \right|^2 \quad (3.8)$$

Typically, CARMA identification is performed in the discrete domain where the corresponding model is an ARMA process, derived by some discretization scheme. The main contribution of this work

is the use of exponential B-splines (E-B-splines) to establish an exact link between the discrete- and the continuous- time domain models. Interpolating the discrete autocorrelation sequence with E-B-splines seems to be a proper choice since the autocorrelation function of a stochastic model (CAR or CARMA) is a linear combination of exponentials, which admits an exact representation in an E-B-spline basis.

### 3.2 CARMA processes and exponential splines

Motivated by the deterministic theory of LTI systems, we exploit the exponential spline mathematical formulation in this work. Exponential splines provide a formal link between continuous domain convolution operators and their discrete-domain counterparts [56, 57] and they will be shown suitable for describing sampled ARMA processes, too.

Considering an ideal sampling procedure, the autocorrelation sequence of the sampled process corresponds to sample values of the autocorrelation function of the original continuous domain process. It then follows that both autocorrelation measures are of an exponential type, suggesting an exponential spline framework for describing the relation between an ARMA process and its sampled version.

While many of the currently available estimation algorithms are focused on base-band power spectra, it seems possible to derive an estimator that overcomes aliasing. Such an estimator could be found useful to the area of optical imaging for which the acquisition device has limited resolution, and to the compressed sensing setup which uses a reduced number of measurements. Another potential application is resolution conversion in which low resolution digital images are displayed on a high resolution device [78]. From a numerical perspective, Whittle's likelihood function plays an important role in deriving frequency-based estimation algorithms. This function is often approximated by means of discrete Fourier transform values and by Riemann sums. Such an approximation is not necessarily optimal and there may exist other numerical schemes that better approximate it.

This work provides a rigorous derivation of a maximum likelihood estimator of continuous domain ARMA parameters from sampled data. It utilizes the exponential B-spline framework while introducing an exact zero-pole coupling for the sampled process.

For that purpose, the relation between the autocorrelation function and the autocorrelation sequence is investigated in both time- and frequency domains. Based on this relation, it is shown that the Cramér-Rao bound can be made arbitrarily small by considering more sample values where the sampling interval can take an arbitrary value. The likelihood function of the sampled process is investigated, too. In particular, it is shown that this function possesses local minima that originate from aliasing. The global minimum, however, corresponds to the maximum likelihood value. The only assumption that is made throughout this study is that the number of available samples is relatively large, allowing one to replace the whitening matrix with a digital filter. This approximation is shown to be valid when considering expected values of the likelihood function. Experimental results indicate that the proposed exponential-based approach is a preferable choice over currently available methods that are restricted to relatively high sampling rates.

A differential LTI system of order  $p$  is fully described by a rational transfer function ( $p > q$ )

$$\hat{h}(\omega) = \frac{\prod_{k=1}^q (j\omega - r_k)}{\prod_{k=1}^p (j\omega - s_k)}, \quad (3.9)$$

If such a system is driven by continuous domain white Gaussian noise, its output is a CARMA process. The spectral density function of such a process is  $\hat{\varphi}(\omega) = \sigma^2 |\hat{h}(\omega)|^2$  and the autocorrelation function is  $\varphi(t) = \sigma^2 \{h(\tau) * h(-\tau)\}(t)$ , where  $\sigma^2$  is the variance of the noise and where  $h(t) = \mathcal{F}^{-1}\{\hat{h}(\omega)\}(t)$  is the impulse response of the system. Exponential splines provide a mathematical framework for relating continuous domain LTI systems with their discrete-domain counterparts. The dependence of  $\varphi(t)$  on  $h(t)$  suggests that these splines may be equally helpful for relating continuous domain and discrete-domain ARMA processes, too.

### Motivating Example: First Order AR Process

The autocorrelation function of a continuous domain AR(1) process that has a pole at  $s = s_1$  and a unit variance innovation is the symmetric exponential

$$\varphi(t; s_1) = -\frac{1}{2s_1} e^{s_1|t|}. \quad (3.10)$$

The spectral density function is then

$$\Phi(j\omega; s_1) = \mathcal{F} \{ \varphi(t; s_1) \} (\omega) = \frac{1}{(j\omega - s_1)(-j\omega - s_1)}. \quad (3.11)$$

Upon sampling, the autocorrelation sequence of the corresponding discrete-domain process is

$$\varphi[n; s_1] = -\frac{1}{2s_1} e^{s_1|n|}, \quad (3.12)$$

where a unit-time sampling interval was assumed for simplicity. The spectral density function of the sampled process is then

$$\Phi_d(e^{j\omega}; s_1) = \mathcal{F}_d \{ \varphi[n; s_1] \} (e^{j\omega}) = \frac{e^{2s_1} - 1}{2s_1} \cdot \frac{1}{(1 - e^{s_1} e^{-j\omega})(1 - e^{s_1} e^{j\omega})}. \quad (3.13)$$

The important observation is that one is able to link the continuous domain and the discrete-domain autocorrelations via the Shannon-like interpolation formula (see also figure 3.1)

$$\varphi(t; s_1) = \sum_{n=-\infty}^{\infty} \varphi[n; s_1] \beta(t - n; s_1), \quad (3.14)$$

where  $\beta(t; s_1)$  is an interpolating basis function whose Fourier expression is

$$\hat{\beta}(\omega; s_1) = \frac{2s_1}{e^{2s_1} - 1} \cdot \frac{(1 - e^{s_1} e^{-j\omega})(1 - e^{s_1} e^{j\omega})}{(j\omega - s_1)(-j\omega - s_1)}. \quad (3.15)$$

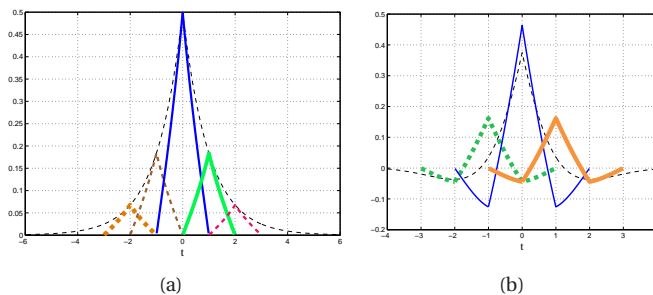
Observe that the latter expression is also equal to the ratio of (3.11) and (3.13). The corresponding time-domain expression is

$$\beta(t; s_1) = \begin{cases} \frac{\sinh[s_1(1-|t|)]}{\sinh s_1} & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}. \quad (3.16)$$

The key property that will be exploited in this work is that  $\beta(t; s_1)$  is compactly supported, which is not obvious a priori from the Fourier-domain expression (3.15). In fact,  $\beta(t; s_1)$  is an exponential B-spline and the above method generalizes for higher order systems.

### General Case

As stated before, a continuous domain ARMA process is fully characterized by its parameters vector,  $\theta$ .



**Figure 3.1:** Exponential B-spline decomposition of an ARMA model. Shown is the autocorrelation function of an ARMA process (black, dashed line). This function can be decomposed by a weighted sum of exponential B-spline functions (various colors, various widths, dashed or solid lines). Shown are only few exponential B-spline functions. The parameters of the ARMA process are 3.1(a)  $\sigma = 1, s = -1$  and 3.1(b)  $\sigma = 1, r = -1, s = -1 \pm i$ .

The Laplace transform of the corresponding autocorrelation function is given by

$$\Phi(s; \theta) = \sigma^2 \frac{\prod_{k=1}^q (s - r_k)(-s - r_k)}{\prod_{k=1}^p (s - s_k)(-s - s_k)}. \quad (3.17)$$

By performing the partial fraction decomposition of  $\Phi(s; \theta)$  (we are assuming for simplicity that the poles are simple), we find that

$$\Phi(s; \theta) = \sigma^2 \sum_{k=1}^p \alpha_k(\theta) \left\{ \frac{1}{s - s_k} - \frac{1}{s + s_k} \right\}, \quad (3.18)$$

and we deduce that the autocorrelation function is a sum of exponentials,

$$\varphi(t; \theta) = \sigma^2 \sum_{k=1}^p \alpha_k(\theta) \cdot e^{s_k |t|}. \quad (3.19)$$

In cases where  $\Phi(s; \theta)$  introduces pole multiplicity,  $\varphi(t; \theta)$  would involve polynomial multiplications as given in table 3.1.

Ideally sampling a continuous domain ARMA process yields a discrete-domain ARMA process. The autocorrelation sequence of the discrete-domain process is then given by the ideal samples of the autocorrelation function. For the partial decomposition of (3.19) and for a unit-time sampling interval, this sequence is given

by

$$\varphi [n; \theta] = \sigma^2 \sum_{k=1}^p \alpha_k(\theta) \cdot e^{s_k |n|}. \quad (3.20)$$

Autocorrelation functions that consist of multiple poles can be discretized in a similar manner as given in table 3.1.

Table 3.1: continuous-to-discrete domain mapping of multiple poles.

Continuous Domain		Discrete Domain	
Laplace	Time	index	z-transform
$\frac{1}{s-s_k} + \frac{1}{-s-s_k}$	$e^{s_k t }$	$e^{s_k n }$	$-\frac{e^{-s_k}}{z^{-1}-e^{-s_k}} + \frac{e^{s_k}}{z^{-1}-e^{s_k}}$
$\frac{1}{(s-s_k)^2} + \frac{1}{(-s-s_k)^2}$	$ t  \cdot e^{s_k t }$	$ n  \cdot e^{s_k n }$	$\left[ \frac{e^{-s_k}}{z^{-1}-e^{-s_k}} + \frac{e^{-2s_k}}{(z^{-1}-e^{-s_k})^2} \right] + \left[ \frac{e^{s_k}}{z^{-1}-e^{s_k}} + \frac{e^{2s_k}}{(z^{-1}-e^{s_k})^2} \right]$
$\frac{1}{(s-s_k)^3} + \frac{1}{(-s-s_k)^3}$	$\frac{1}{2} t^2 \cdot e^{s_k t }$	$\frac{1}{2} n^2 \cdot e^{s_k n }$	$-\left[ \frac{1}{2} \frac{e^{-s_k}}{z^{-1}-e^{-s_k}} + \frac{3}{2} \frac{e^{-2s_k}}{(z^{-1}-e^{-s_k})^2} + \frac{e^{-3s_k}}{(z^{-1}-e^{-s_k})^3} \right] + \left[ \frac{1}{2} \frac{e^{s_k}}{z^{-1}-e^{s_k}} + \frac{3}{2} \frac{e^{2s_k}}{(z^{-1}-e^{s_k})^2} + \frac{e^{3s_k}}{(z^{-1}-e^{s_k})^3} \right]$

ARMA models are closely related to generalized exponential B-splines. These finite-support functions stem from Green's functions of rational operators [57], and in our case the Green's function is the autocorrelation function itself. Unlike [57], however, this work introduces non-causal symmetric B-splines.

**Definition 1** (*Symmetric Exponential B-spline*). *The exponential B-spline  $\beta(t; \theta)$  with parameters  $\theta$  is specified by the following inverse Fourier transform*

$$\beta(t; \theta) = \mathcal{F}^{-1} \left\{ \prod_{k=1}^q (j\omega - r_k)(-j\omega - r_k) \cdot \prod_{k=1}^p \frac{(1 - e^{j\omega + s_k})(1 - e^{-j\omega + s_k})}{(j\omega - s_k)(-j\omega - s_k)} \right\} (t) \quad (3.21)$$

Observe that this function corresponds to the ratio of the ARMA power spectrum and of a discrete-domain AR power spectrum with poles at  $e^{s_k}$ . The exponential B-splines kernels have the following properties:

- Compactly supported function within the interval  $[-p, p]$ .
- Bounded and symmetric functions.
- Smooth functions: the first  $2(p - q) - 1$  derivatives are  $L_2$  functions.
- Their integer shifts form a Riesz basis.
- Reproduce exponential functions of the type (3.19).
- The convolution of two exponential B-spline kernels yields another B-spline kernel of an augmented order. This property allows one to iteratively construct an exponential B-splines of any order.

**Definition 2** (*Localization filter*). *The localization filter with parameters  $\theta$  is specified by the Laplace transform*

$$\Delta(s; \theta) = \prod_{k=1}^p (1 - e^{s + s_k})(1 - e^{-s + s_k}). \quad (3.22)$$

**Proposition 1.** *The autocorrelation function of an ARMA process with parameters  $\theta$  can be written as*

$$\varphi(t; \theta) = \sigma^2 \sum p[m; \theta] \cdot \beta(t - m; \theta), \quad (3.23)$$



where  $\beta(t; \theta)$  is the exponential B-spline with parameters  $\theta$  and the sequence  $p$  is given in the z-domain by

$$P_d(z; \theta) = \frac{1}{\prod_{k=1}^p (1 - e^{s_k} z)(1 - e^{s_k} z^{-1})} = \frac{1}{\Delta(\ln z; \theta)}. \quad (3.24)$$

*Proof.* Take the Fourier transform of (3.23) and substitute (3.21) and (3.17).  $\square$

**Definition 3.** The discrete exponential B-spline kernel with parameters  $\theta$  is given by  $\mathcal{B}_d(z; \theta) = \sum_{n=-p}^p \beta[n; \theta] z^{-n}$  and  $\hat{\beta}_d(\omega) = \mathcal{B}_d(e^{j\omega})$ , where the sequence  $\beta[n; \theta]$  contains the samples of  $\beta(t; \theta)$ .

We rely on [65] and assume that ideal sampling preserves the autocorrelation values of the continuous domain ARMA process. The power spectrum of the sampled process is then related to the continuous domain power spectrum through aliasing and it can be described by means of a rational transfer function in the z-domain [65]. The following Theorem suggests a computationally efficient way for extracting the discrete-domain power spectrum from the continuous domain parameters.

**Theorem 1.** The discrete-domain ARMA process that is given by the ideal unit-interval samples of the continuous domain process (3.17) has the following parameters:

$$H_d(z; \theta) = \sigma_d(\theta) \frac{\prod_{k=1}^{p-1} (1 - \nu_k(\theta) z^{-1})}{\prod_{k=1}^p (1 - \rho_k(\theta) z^{-1})}, \quad (3.25)$$

where

$$\rho_k(\theta) = e^{s_k} \quad (3.26)$$

$$\nu_k(\theta) = \text{roots of } \mathcal{B}_d(z; \theta) \text{ inside the unit circle} \quad (3.27)$$

and where

$$\sigma_d^2(\theta) = \sigma^2 \frac{\mathcal{B}_d(1; \theta)}{\prod_{k=1}^{p-1} (1 - \nu_k(\theta))^2} \quad (3.28)$$

is the variance of the discrete-domain innovation process.

*Proof.* As  $P_d(z; \theta)$  is an all-pole filter, the zeros of

$$\Phi_d(z; \theta) = \sigma^2 P_d(z; \theta) \cdot \mathcal{B}_d(z; \theta) \quad (3.29)$$

originate from the zeros of  $\mathcal{B}_d(z; \theta)$  only. Those zeros appear in reciprocal pairs due to the symmetry property of  $\beta(t; \theta)$ . Also, the DFT of the sampled version of the exponential B-spline satisfies

$$\hat{\beta}_d(\omega; \theta) = \sum_{k=-\infty}^{\infty} \hat{\beta}(\omega + 2\pi k; \theta) = \mathcal{B}_d(e^{j\omega}), \quad (3.30)$$

and it does not vanish on  $\omega \in [0, 2\pi)$  due to (3.21).  $\beta(t; \theta)$  is of finite support so that  $\mathcal{B}_d(z; \theta)$  has no poles. The symmetry property of  $\Delta(s; \theta)$  indicates that the poles of  $P_d(z; \theta)$  appear in reciprocal pairs and it then follows that  $\Phi_d(z; \theta)$  can be described by a minimum-phase filter. The finite support property of  $\beta(t; \theta)$  and the structure of  $\Delta(s; \theta)$  also guarantee that this minimum-phase filter has a rational transfer function. We further observe that,

$$\mathcal{B}_d(z; \theta) = \mathcal{B}_d^-(z; \theta) \cdot \mathcal{B}_d^+(z; \theta) \quad (3.31)$$

$$P_d(z; \theta) = P_d^-(z; \theta) \cdot P_d^+(z; \theta) \quad (3.32)$$

where ‘-’ denotes a causal filter and where  $\mathcal{B}_d^+(z; \theta) = \mathcal{B}_d^-(z^{-1}; \theta)$  and  $P_d^+(z; \theta) = P_d^-(z^{-1}; \theta)$ . In particular,  $P_d^-(z; \theta) = 1 / \prod_{k=1}^p (1 - e^{s_k} z^{-1})$  and  $\mathcal{B}_d^-(z; \theta)$  has zeros only; those zeros originate from the roots of  $\mathcal{B}_d(z; \theta)$  inside the unit circle. It then follows that,

$$\Phi_d(z; \theta) = H_d(z; \theta) \cdot H_d(z^{-1}; \theta), \quad (3.33)$$

where  $H_d(z; \theta) = \sigma \cdot P_d^-(z; \theta) \cdot \mathcal{B}_d^-(z; \theta)$  is the required filter. Imposing the structure of (3.25) results in the expression for  $\sigma_d^2(\theta)$ .  $\square$

Beside equation (3.28) there is another way of computing the analytical expression of the variance of the discrete innovation signal based on the classical definition of exponential B-spline [56]. The difference with the previous formulation is a multiplicative term (in the frequency domain) in the definition of  $P_d(z; \theta)$  (3.24) and of the E-B-spline  $\beta(t; \theta)$  (3.21). The new definition of  $P_d(z; \theta)$  differs from (3.24) by  $[z^p \prod_{k=1}^p e^{s_k}]$ , likewise the new definition of  $\beta(t; \theta)$  differs from (3.21) by  $[(-1)^q e^{-j\omega p} \prod_{k=1}^p e^{-s_k}]$ . These differences are clarified in appendix D. Provided these changes, we can derive the following corollary.

**Corollary 1.** *The variance of the innovation signal of the sampled CARMA process has the following form,*

$$\sigma_d^2(\theta) = \sigma^2 \beta[1; \theta] \prod_{k=1}^{p-1} (-1/v_k(\theta)) \prod_{k=1}^p e^{s_k} \quad (3.34)$$

*Proof.* From equation (3.29) we derive

$$\Phi_d(z; \theta) = \sigma^2 z^p \prod_{k=1}^p e^{s_k} \cdot \frac{\mathcal{B}_d(z; \theta)}{\prod_{k=1}^p (1 - e^{s_k} z^{-1})(1 - e^{s_k} z)} \quad (3.35)$$

$\mathcal{B}_d(z; \theta)$  is a reciprocal polynomial of order  $2p - 2$  due to its symmetry and, in order to make its reciprocal roots explicit, it can be expressed in the following manner

$$\begin{aligned} \mathcal{B}_d(z; \theta) &= \sum_{k=0}^{2p} \beta[k, \theta] z^{-k} & (3.36) \\ &= \beta[1, \theta] \cdot z^{-1} \prod_{k=1}^{p-1} (1 - v_k(\theta) z^{-1})(1 - \frac{1}{v_k(\theta)} z^{-1}) \\ &= \beta[1, \theta] \cdot z^{-p} \prod_{k=1}^{p-1} \frac{-1}{v_k(\theta)} \prod_{k=1}^{p-1} (1 - v_k(\theta) z^{-1})(1 - v_k(\theta) z) \end{aligned}$$

Recalling equation (3.33) and replacing  $\mathcal{B}_d(z; \theta)$  in (3.35) results in the expression for  $\sigma_d^2(\theta)$ .  $\square$

**Corollary 2.** *Let  $\theta$  be known. Then the autocorrelation function of a continuous domain ARMA process is uniquely defined by its samples. Further,*

$$\varphi(t; \theta) = \sum_{n=-\infty}^{\infty} \varphi(n; \theta) \cdot \eta(t - n; \theta), \quad (3.37)$$

where the interpolation kernel  $\eta(t; \theta)$  is specified by its Fourier transform,

$$\hat{\eta}(\omega; \theta) = \frac{\hat{\beta}(\omega; \theta)}{\hat{\beta}_d(\omega; \theta)}. \quad (3.38)$$

This is the generalization of (3.14) for arbitrary ARMA( $p, q$ ) processes.

*Proof.* It was shown that  $\Phi_d(z; \theta) = \sigma^2 P_d(z; \theta) \cdot \mathcal{B}_d(z; \theta)$  and that  $\Phi(s; \theta) = \sigma^2 \mathcal{B}(s; \theta) / \Delta(s; \theta)$ . Dividing the two equations while recalling that  $P_d(z; \theta) = \frac{1}{\Delta(\ln z; \theta)}$  yields the required result.  $\square$

The exponential interpolation function  $\eta(t; \theta)$  provides a mean for interpolating continuous domain ARMA models. It can also be interpreted as a spectral weighting function that relates the power spectrum of the discrete-domain sampled process with the power spectrum of the continuous domain process it originates from. Unlike the polynomial-based weighting function of [76], this weighting function is parameterized, allowing one to describe band-pass power spectrum, too.

### 3.3 Insights on the exponential spline framework

The motivation for using an exponential spline parametrization for CARMA models stems from the fact that the ideal sampling model preserves autocorrelation properties. The autocorrelation functions of continuous-time ARMA processes are exponential splines and the proposed framework has several advantages in this regard:

1. The spline framework allows one to map continuous domain parameters to discrete-domain parameters by using simple algebraic operations. The simple mapping of Theorem 1 can then be used in estimation algorithms that are based on numerical optimization and that require repeated mapping operations. This mapping involves (a) calculating the autocorrelation function by a partial fraction decomposition, (b) expressing the B-spline of Definition 1 by the FIR localization filter of Definition 2, and (c) calculating sample values of the B-spline. As the B-spline is compactly supported in  $[-p, p]$ , the z-transform of its sampled version is a polynomial of a relatively low degree. This in turn allows one to calculate the zeros of Theorem 1 relatively fast.
2. The proposed framework does not depend on sampling interval values when mapping continuous domain parameters to their discrete-domain counterparts. It provides a discrete-domain process whose autocorrelation sequence is in a perfect match with the point-wise values of the autocorrelation function for any sampling interval value.
3. The proposed framework provides a closed form expression for the interpolation kernel of Corollary 2. Considering the minimum-mean-square-error criterion, this kernel can be used for interpolating the sampled process in a straightforward manner. Moreover, the interpolation task can be carried out using the finite support B-spline of Definition 1. Finite support functions are extremely beneficial for data interpolation in terms of computational complexity.
4. The proposed spline framework can naturally incorporate non-ideal sampling procedures such as averaged sampling. For example the zero-order-hold rectangle function is a spline, too. In such cases, where the sampling function is

by itself a spline, the proposed framework allows one to derive the whitening digital filter in a similar way to Theorem 1. The key point is that there always exists a B-spline that admits the relation (3.23) with a filter  $p[m; \theta]$  that depends also on the sampling model.

An alternative mapping to Theorem 1 was given in [63, 64, 77]. Motivated by a state-space representation, these works introduce closed-form expressions for extracting the discrete-time power spectrum from a given continuous-time model. Nevertheless, these expressions introduce additional computational overload compared to the proposed approach as they rely on matrix operations. In particular, [77] involves exponentials of a matrix and integration of matrix elements, and involves polynomial factorization that requires one to represent an inverse of a matrix in a parametrized form. The proposed spline framework, on the other hand, does not rely on a matrix-based representation. Rather, it evaluates the autocorrelation function at a finite number of points only. Owing to the spline property of the autocorrelation function, the continuous-to-discrete mapping of Theorem 1 involves the partial fraction decomposition of 3.17 and the evaluation of the B-spline function at only  $2p - 1$  points. These points give rise to a polynomial in the  $z$  domain, whose roots are readily computed. Each point-wise value of the B-spline function corresponds to a weighted sum of  $2p + 1$  values of the autocorrelation function.

An additional continuous-to-discrete parameter mapping was also introduced in [79]. This work derives explicit formulae for the first terms of the Maclaurin series of the zeros and the variance of the sampled process as a function of the sampling interval variable  $T$ . This, in turn, allows one to analyze convergence for the limiting case of  $T \rightarrow 0$ . The Maclaurin weights are given by means of the continuous-time parameters and such results can be used for mapping continuous domain parameters to their discrete-domain counterparts. The advantage of the results of [79] is in its extremely low computational complexity. Unlike the proposed spline framework, however, it does not provide an exact continuous-to-discrete mapping; and according to Example 2 of [79], such an approximation would better be used for sampling interval values that are considerably smaller than the characteristic time constants of the system. As we consider prominent aliasing conditions in our work, the approach of [79] is less applicable here.

### 3.4 A new approach to CARMA model identification

Let a continuous domain ARMA( $p, q$ ) process be given by its uniform samples only. Considering a large number of data points  $N \gg 1$ , the Fisher information matrix is given by [80],

$$I_{k,l}(\theta) \cong \frac{N}{4\pi} \int_0^{2\pi} \left( \frac{1}{\hat{\varphi}_d(\omega; \theta)} \frac{\partial \hat{\varphi}_d(\omega; \theta)}{\partial \theta_k} \right) \cdot \left( \frac{1}{\hat{\varphi}_d(\omega; \theta)} \frac{\partial \hat{\varphi}_d(\omega; \theta)}{\partial \theta_l} \right) d\omega, \quad (3.39)$$

where  $k, l = 1, \dots, p + q + 1$  and  $\hat{\varphi}_d(\omega; \theta)$  is the DTFT of the sampled ARMA process autocorrelation sequence. The Cramér-Rao Bound (CRB) is then given by the inverse of  $I(\theta)$ . Sampling interval dependency can be incorporated in the localization filter (3.22) by

$$\Delta(s; \theta, T) = \prod_{k=1}^p (1 - e^{s+s_k T})(1 - e^{-s+s_k T}), \quad (3.40)$$

where  $T$  is the sampling interval. It then follows that  $\hat{\varphi}_d(\omega; \theta)$  is not dependent on  $N$  nor does the integrand of (3.39). It then follows that, as the number of available samples becomes larger, the CRB becomes smaller regardless of  $T$ . This implies that aliasing effects can be compensated for by taking more measurements. It further suggests that there exists at least an ML (Maximum Likelihood) estimator that overcomes aliasing. Such an estimator has to take into account the fact that the discrete-domain poles and the zeros of  $\Phi_d(z; \theta)$  are coupled, ensuring that the discrete process spectrum corresponds to  $z$ -transform of the samples of the autocorrelation function  $\varphi(t; \theta)$ . Indeed, one can estimate the digital data by a discrete-domain ARMA model while introducing no zeros-poles coupling. However, in such a case the estimated parameters will not necessarily correspond to a continuous domain model. For example, a possible discrete-domain ARMA(1,0) estimation algorithm may yield a pole at  $z = -0.5$ ; the autocorrelation sequence is then  $\varphi[n] = (-0.5)^{|n|}$  which has no continuous domain ARMA(1,0) counterpart. Furthermore, non-coupled estimators result in baseband power spectrum, making it difficult to cope with aliasing effects.

Considering this matter further, there are at least two possible ARMA models that can be associated with the sampled data. One of them is a discrete-domain ARMA model that introduces no

zero-pole coupling. The ML estimator in this case corresponds to minimizing the variance of the prediction error. The other is a constrained discrete-domain ARMA model which takes the sampling process into account and which introduces the required zero-pole coupling. The ML estimator of such a constrained model does not necessarily minimize the variance of the prediction error. The constrained digital whitening filter (3.25) is guaranteed to yield a discrete-domain innovation process that is independent and identically distributed. This does not hold true for the non-constrained model as the innovation is not guaranteed to be independent, e.g. when the estimated power-spectrum does not comply with the required zero-pole coupling, although it may yield lower variance values. The continuous domain whitening filter (3.25) is always guaranteed to exist and to whiten the continuous domain process whereas the continuous domain whitening filter of the non constrained model, if exists [77], is not guaranteed to yield white noise.

The purpose of (3.39) is to provide motivation for deriving an estimation algorithm that overcomes aliasing. We assume that the number of samples is very large, allowing one to represent the discrete domain power spectrum by means of a rational transfer function. We then rely on the results of [80] which characterize the Cramér-Rao bound by means of the power spectrum function. Our parametrization of the discrete-domain power spectrum allows us, in turn, to demonstrate that the Cramér-Rao bound converges to zero for sampled ARMA processes and that it is inversely proportional to  $N$ . In this regard, the work [81] derives closed form expressions for the Cramér-Rao bound using the state space representation. Also, it assumes data sets of an arbitrary size. The closed form expressions, however, involve matrix inversion and eigenvalue decomposition, and it is difficult to implement such computations for large data sets as indicated in section 3 of [81]. It is also indicated there that, under certain circumstances, the Cramér-Rao bound becomes inversely proportional to  $N$ . We demonstrated such proportionality for every sampled ARMA process.

Motivated by the CRB, we approximate the log-likelihood function by means of a digital filter. The proposed approximation relies on the discrete-domain parameters of (3.25), namely  $\{\sigma_a^2, \nu_1, \dots, \nu_q, \rho_1, \dots, \rho_p\}$ . These parameters can be numerically calculated from  $\theta$  in a straightforward manner as was shown in section 3.2. We, however, do not allow the continuous domain poles of  $\theta$  to differ by  $2k\pi j$  with  $k \in \mathbb{Z}$  as such poles yield the same discrete-domain model upon sampling.

**Definition 4.** Let  $\theta$  be known and let  $\mathbf{x}$  be  $N$  uniform ideal samples of the continuous domain process (3.17) taken on a unit-interval grid. The probability density function of  $\mathbf{x}$  is

$$pdf(\mathbf{x}; \theta) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \right\}, \quad (3.41)$$

where  $\Sigma[m, n] = \varphi[m - n; \theta]$  is the autocorrelation matrix that corresponds to  $\theta$ .

**Definition 5.** The corresponding log-likelihood function, including a sign inversion, is

$$l(\theta; \mathbf{x}) = \ln |\Sigma| + \mathbf{x}^T \Sigma^{-1} \mathbf{x}. \quad (3.42)$$

**Definition 6.** Let  $\theta$  be known. The digital filter  $g_\theta$  is given by its  $Z$ -transform  $G_d(z) = 1/H_d(z)$  where  $H_d(z)$  is given as in (3.25).

**Definition 7.** Let  $\theta$  be known. The parameter only function  $\kappa(\theta)$  is given by

$$\kappa(\theta) = \sum_{n=1}^{\infty} n \cdot c[n; \theta]^2, \quad (3.43)$$

where

$$c[n; \theta] = \frac{1}{n} \left\{ v_1^n(\theta) + \dots + v_{p-1}^n(\theta) - \rho_1^n(\theta) - \dots - \rho_p^n(\theta) \right\}. \quad (3.44)$$

The constant  $\kappa(\theta)$  can be interpreted by means of an inner product operation. Considering a discrete-domain ARMA power spectrum, we define the Fourier coefficients of its logarithm as:

$$\ln \hat{\phi}_d(\omega) = \sum_{n=-\infty}^{\infty} c[n; \theta] e^{-j\omega n} \quad (3.45)$$

Recalling (3.25):

$$\ln H_d(e^{-j\omega}; \theta) = \frac{c[0; \theta]}{2} + \sum_{n=1}^{\infty} c[n] e^{-j\omega n} \quad (3.46)$$

it then follows that

$$\kappa(\theta) = \left| \left\langle \frac{d}{d\omega} (\ln H_d(e^{j\omega})), \ln H_d(e^{j\omega}; \theta) \right\rangle_{L_2} \right| \quad (3.47)$$



**Theorem 2.** *Let  $\theta$  be known and let  $\mathbf{x}$  be  $N$  uniform ideal samples of the continuous domain process (3.17) taken on a unit-interval grid. Then,*

$$\lim_{N \rightarrow \infty} \mathbb{E} \left( l(\theta; \mathbf{x}) - \bar{l}(\theta; \mathbf{x}) \right) = 0, \quad (3.48)$$

where

$$\bar{l}(\theta; \mathbf{x}) = N \ln \sigma_d^2(\theta) + \kappa(\theta) + \|\mathbf{x} * g_\theta\|_{\ell_2}^2 \quad (3.49)$$

and  $*$  denotes discrete-domain convolution of an  $N$ -length output sequence.

*Proof.* According to Szegö theorem for infinite Toeplitz matrices [82, 83, 84],

$$\lim_{N \rightarrow \infty} \left\{ \ln |\Sigma| - N \cdot c[0; \theta] \right\} = \sum_{n=1}^{\infty} n c[n; \theta] c[-n; \theta], \quad (3.50)$$

where  $c$  corresponds to the Fourier coefficients of  $\ln \hat{\phi}_d(\omega; \theta)$ . In our case, these coefficients are given by (3.44) and the right-hand-side of the equation can be replaced by  $\kappa(\theta)$ . Also, by the Residue theorem

$$c[0; \theta] = \frac{1}{2\pi} \int_0^{2\pi} \ln \hat{\phi}_d(\omega; \theta) d\omega = \ln \sigma_d^2(\theta). \quad (3.51)$$

It then holds that

$$\lim_{N \rightarrow \infty} \left\{ \ln |\Sigma| - N \ln \sigma_d^2(\theta) - \kappa(\theta) \right\} = 0. \quad (3.52)$$

The constant  $\kappa(\theta)$  also provides a mean for describing the limiting behaviour of the determinant. Writing (3.52) in a different form [84],

$$|\Sigma(\theta)| \propto e^{\kappa(\theta)} \cdot (e^{c[0; \theta]})^N \quad (3.53)$$

we observe that the right-hand-side of the equation is equal to the determinant up to a multiplicative constant; the important thing here is that this constant does not depend on  $N$  or on  $\theta$ .

As for the term  $\mathbf{x} \cdot \Sigma^{-1} \cdot \mathbf{x}$  that appears in  $l(\theta; \mathbf{x})$ , it may be approximated by digitally filtering  $\mathbf{x}$  and by calculating the energy of the output. A possible filter one can use is  $g_\theta$ . This filter is guaranteed by Theorem 1 to exist, to be stable and to be causal. The  $N$ -length output of such a filter can be described by means of the lower triangular matrix

$$L[m, n] = g_\theta[m - n], \quad m, n = 0, \dots, N - 1, \quad (3.54)$$

and it holds that  $\|\mathbf{x} * g_\theta\|_{\ell_2}^2 = \mathbf{x}^T L L^T \mathbf{x}$ . Following [85], the two matrices  $\Sigma^{-1}$  and  $L L^T$  are asymptotically equivalent as  $N$  becomes larger. This stems from the fact that they both originate from the same power spectrum  $\hat{\varphi}_d(\omega; \theta)$ . Such a power spectrum gives rise to an infinite Toeplitz matrix  $T_M$  and the  $N \times N$  matrix  $\Sigma^{-1}$  is defined by 1) truncating  $T_M$  and 2) taking the inverse. The matrix  $L$  is defined by 1) inverting  $T_M$ , 2) finding its Cholesky decomposition and 3) truncating the lower triangle matrix.

The asymptotically equivalent matrices are covariance matrices that describe asymptotically equivalent random processes. In particular, let  $\sigma_1^2 = \mathbb{E}(\mathbf{x}^T \Sigma^{-1} \mathbf{x})$  and  $\sigma_2^2 = \mathbb{E}(\mathbf{x}^T L L^T \mathbf{x})$  correspond to the variance values of asymptotically equivalent stochastic processes. Then,  $\lim_{N \rightarrow \infty} (\sigma_1^2 - \sigma_2^2) = 0$ , which proves the theorem.

Asymptotic equivalence implies that the norm of the difference between the two matrices converges to zero with increasing values of  $N$ . Focusing on a Toeplitz matrix  $T_M[m, n] = t_M[m - n]$ , one can associate a discrete-time Fourier transform to the sequence  $t_M[k]$  and, for autocorrelation matrices, this transform corresponds to the power spectrum function. Asymptotic equivalence of two covariance matrices implies that the power spectrum of one matrix converges uniformly to the power spectrum of the other. In the context of this proof, the matrix  $L$  describes a truncated version of the filter  $g_\theta$ . As  $g_\theta \in l_1$ , its discrete-time Fourier transform uniformly converges to  $\frac{1}{H_d(e^{j\omega})}$ . This means that with increasing numbers of samples, the expression  $\mathbf{x}^T L$  converges in the  $l_2$  sense to  $\mathbf{x} * g_\theta$ . This fact stems from Parseval's property of the discrete-time Fourier transform. It then follows that the output is a stochastic process with a power spectrum that converges to a constant function. This constant is  $\sigma_d(\theta)$ . The term  $\mathbf{x}^T L L^T \mathbf{x}$  is then an estimation for this value. The expression  $\mathbf{x}^T \Sigma^{-1} \mathbf{x}$  amounts to decorrelating the process  $\mathbf{x}$  and to estimating the variance of the uncorrelated process. The value of this variance is  $\sigma_d(\theta)$ , too. It then follows that  $\mathbb{E}\{\mathbf{x}^T L L^T \mathbf{x}\} = \mathbb{E}\{\mathbf{x}^T \Sigma^{-1} \mathbf{x}\} = N \sigma_d(\theta)$ . □

The proposed approach is related to Whittle's approximation of the log-likelihood function of (3.42). Whittle's approximation is given by

$$\tilde{l}_W(\theta; \mathbf{x}) = \frac{N}{2\pi} \int_0^{2\pi} \ln \hat{\varphi}_d(\omega; \theta) d\omega + \frac{1}{2\pi} \int_0^{2\pi} \frac{|\hat{\mathbf{x}}_d(\omega)|^2}{\hat{\varphi}_d(\omega; \theta)} d\omega. \quad (3.55)$$

It holds that

$$N \ln \sigma_d^2 = \frac{N}{2\pi} \int_0^{2\pi} \ln \hat{\varphi}_d(\omega; \theta) d\omega \quad (3.56)$$

$$\lim_{N \rightarrow \infty} \|\mathbf{x} * g_\theta\|_{\ell_2}^2 = \frac{1}{2\pi} \int_0^{2\pi} \frac{|\hat{\mathbf{x}}_d(\omega)|^2}{\hat{\varphi}_d(\omega; \theta)} d\omega, \quad (3.57)$$

where the limit is required due to the use of finite-length convolution in  $\tilde{l}(\theta; \mathbf{x})$ . It then follows that  $\lim_{N \rightarrow \infty} \tilde{l}(\theta; \mathbf{x}) = \tilde{l}_W(\theta; \mathbf{x}) + \kappa(\theta)$ . The integrals of Whittle's likelihood function are often approximated by Riemann sums that involve DFT values. DFT values of a finite length signal, in this case  $\mathbf{x}$ , indeed coincide with the samples of its  $Z$ -transform on the unit circle. Nevertheless, this is not the case for infinite sequences that are described by  $\Phi_d(z; \theta)$  and by  $1/\Phi_d(z; \theta)$ . The proposed approach, on the other hand, does provide an accurate evaluation of these integrals and the truncation error of the convolution operation is exponentially decreasing, providing a higher convergence rate than the Riemann sum method.

The proposed ML approach differs from currently available methods in several aspects:

- It considers exponential autocorrelation models for both the continuous- and the discrete-domain processes whereas several previous works considered polynomial models.
- It establishes a link between the continuous- and the discrete-domain models by incorporating the sampling process into the problem formulation. This, in turn, allows for the discrete model to stem naturally from the continuous domain formulation while no a priori assumptions are made on the digital data.
- The log-likelihood function suggested here holds true for any value of sampling interval, rather than describing the limiting case of  $T \rightarrow 0$ .
- The log-likelihood function considers discrete-domain data for determining continuous domain statistics while no approximation of continuous domain frequency spectrum or impulse responses is required.

The log-likelihood function (3.49) has several local minima as demonstrated by figure 3.2 where simulation results are shown for

a continuous domain AR(2) process having poles  $s_{1,2} = -1 \pm 5i$ . Given a sampled version of such a process, the log-likelihood function (3.49) was repeatedly minimized using different initial conditions; the initial conditions are the real and imaginary part of the poles one starts with, and they are indicated by the  $x$  and  $y$  axes respectively. A detailed description of the five ring-like regions is given in table 3.2. These local minima originate from aliasing and there exist several continuous domain processes that result in similar discrete-domain power spectrum upon sampling. These very processes generate the local minima. Shown in figure 3.3 is  $\Phi_d(e^{j\omega})$  for the AR processes of table 3.2 while considering a unit sampling interval. These processes correspond to the local minima of figure 3.2. The aliased spectrum shown here resembles each other in terms of their bandpass nature and in terms of the frequency of maximum response. Region 1, however, exhibits a low-pass signal, emphasizing the difference between the proposed ML estimation approach and other base-band methods. The peak response of these power spectra are distributed along the frequency axis in distinct bands as shown in figure 3.4; these frequency bands are  $\pi$  [rad/time-unit] wide. This property suggests that every local minimum can be obtained by minimizing (3.49) while allocating initial conditions that correspond to a peak response at the required band, e.g. at  $(k + 1/2)\pi$  where  $k$  is the band index [86]. Another way of determining initial conditions will be described later on. Following Theorem 2, the global minimum of (3.49) corresponds to a ML estimator and it is suggested here to minimize the likelihood function using several initial conditions.

The work of [87] investigates the maximum likelihood estimator of a discrete-domain ARMA process, too. In particular, it characterizes the local minima values of the likelihood function. The analysis of [87] relies on the equivalence of the maximum likelihood estimator with the prediction error variance minimization estimator. According to Theorem 2 of [87], there is a unique global minimum, provided that enough parameters are used in the fitted model. The ARMA model that originates from sampled data, however, introduces zero-pole coupling which is not considered in [87]. Also, the maximum-likelihood estimator for such cases does not necessarily correspond to minimizing the variance of the prediction error. For these reasons, the results of [87] can be partially applied here. Similar to [87], we also assume a large number of samples, allowing one to formulate the power spectrum by means of a ratio of two polynomials. We also demonstrate that

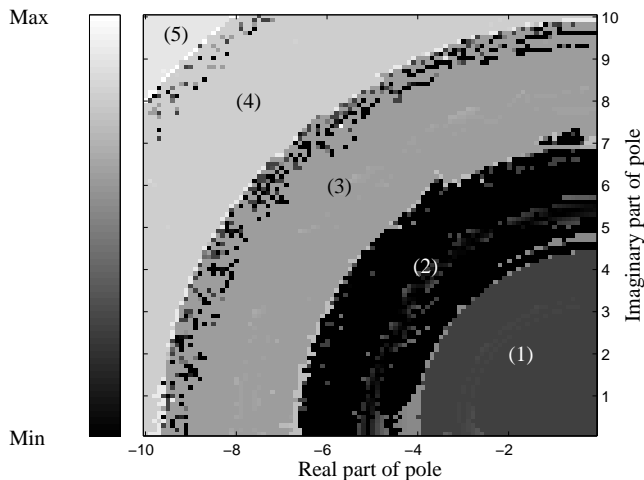


Figure 3.2: Local minima of the likelihood function.

the likelihood function of the sampled data has a global minimum. Nevertheless, due to the zero-pole coupling that is introduced by the sampling process, this global minimum differs from the global minimum of [87]. Furthermore, the sampling process introduces local minima that do not coincide with the global minimum and this property allows one to select one set of parameters which is the most probable.

Table 3.2: A detailed description of figure 3.2.

Region	Log-likelihood Value	Estimated Poles	Frequency <sup>b</sup> [rad/time-unit]
1	-3656	$-2.4 \pm 2.1i$	0
<b>2<sup>a</sup></b>	<b>-3675</b>	<b><math>-0.96 \pm 5.0i</math></b>	<b>4.9</b>
3	-3557	$-0.46 \pm 7.8i$	7.8
4	-3067	$-0.25 \pm 11.1i$	11.1
5	-2212	$-0.16 \pm 14.0i$	14.0

<sup>a</sup> Global minimum of the log-likelihood function.<sup>b</sup> Frequency of maximum response of  $\Phi_L(j\omega)$ .

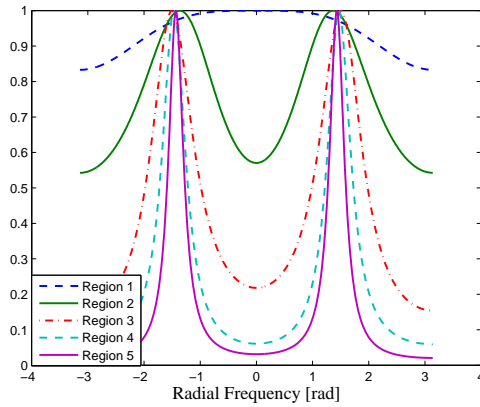


Figure 3.3: Discrete-domain spectrum of sampled AR signals.

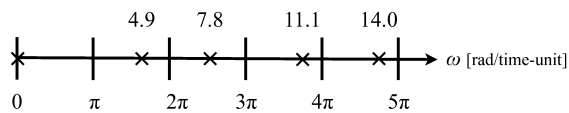


Figure 3.4: Frequencies of maximum response. Shown here by 'x' are the frequencies of maximum response of each local minimum of figure 3.2. These frequencies are located within bands of width  $\pi$  [rad/time-unit]. Additional information is given in table 3.2.

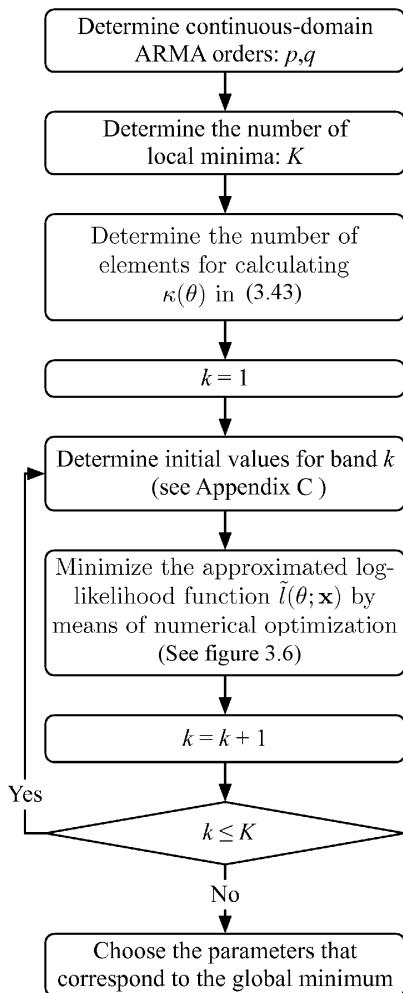


Figure 3.5: The proposed estimation algorithm. Complementary information is given in figure 3.6 and in the appendix C.

### 3.5 CARMA estimation algorithm

The proposed algorithm for estimation of CARMA coefficients exploits the multi-band minimization of the log-likelihood function parametrized by  $\theta$ . The vector of parameters  $\theta$ , which consists of

the zeros and the poles of (3.17), can also be represented by the polynomial coefficients of its numerator and of its denominator. These coefficients are real numbers and the numerical optimization was carried out by using this type of parameterization. For example, the power spectrum of an AR(2) process having two poles  $s_1, s_2$  is given by

$$\Phi(s; \theta) = \frac{\sigma^2}{(s^2 - a_1 s + a_0)(s^2 + a_1 s + a_0)}, \quad (3.58)$$

where  $a_0 = s_1 \cdot s_2$  and  $a_1 = s_1 + s_2$ . The proposed estimation algorithm aims at finding the coefficients  $a_0, a_1$  and the variance of the innovation process  $\sigma^2$ . The dominant coefficient in this example, i.e. the one that changes substantially as the values of the poles change, is  $a_0$ . It has the most prominent effect on the geometric distance between the coefficients one starts with (i.e. the initial conditions for extracting every local minimum) and the coefficients of the local minimum one ends up with. In cases where  $s_1, s_2$  are complex conjugate, the coefficient  $a_0$  corresponds to their module value as demonstrated in figure 3.2. Every location in the image corresponds to a different set of initial conditions. This set was then used for minimizing the proposed approximation of the likelihood function and the color at that point indicates the value it converged to upon minimization. Initial conditions were determined by choosing various complex values for the pole  $s_1$  and the radial appearance of the image stems from its module. For the general case of an ARMA( $p, q$ ) process, the parameters that are being optimized are  $\theta = \{\sigma^2, a_0, a_1, \dots, a_{p-1}, b_0, b_1, \dots, b_{q-1}\}$ , which gives rise to the following power spectrum

$$\Phi(s; \theta) = \sigma^2 \frac{[s^q + \sum_{n=0}^{q-1} b_n s^n] [(-s)^q + \sum_{n=0}^{q-1} b_n (-s)^n]}{[s^p + \sum_{m=0}^{p-1} a_m s^m] [(-s)^p + \sum_{m=0}^{p-1} a_m (-s)^m]}. \quad (3.59)$$

The proposed algorithm finds  $K$  local minima of the approximated likelihood function at consecutive frequency bands and chooses the minimum value among them. Obtaining  $K$  different local minima requires  $K$  sets of initial conditions. It is suggested here to choose bandpass power spectra that have peak responses at  $\omega_0 + \pi k$  [rad/time-unit] where  $k$  is the band index and where  $\omega_0$  is determined by fitting the available data with a discrete-domain AR( $p$ ) process and by extracting the frequency that corresponds to the maximum value of the power spectrum. For arbitrary sampling interval values, one should use  $\frac{\pi k}{T}$  instead of  $\pi k$ . While there are



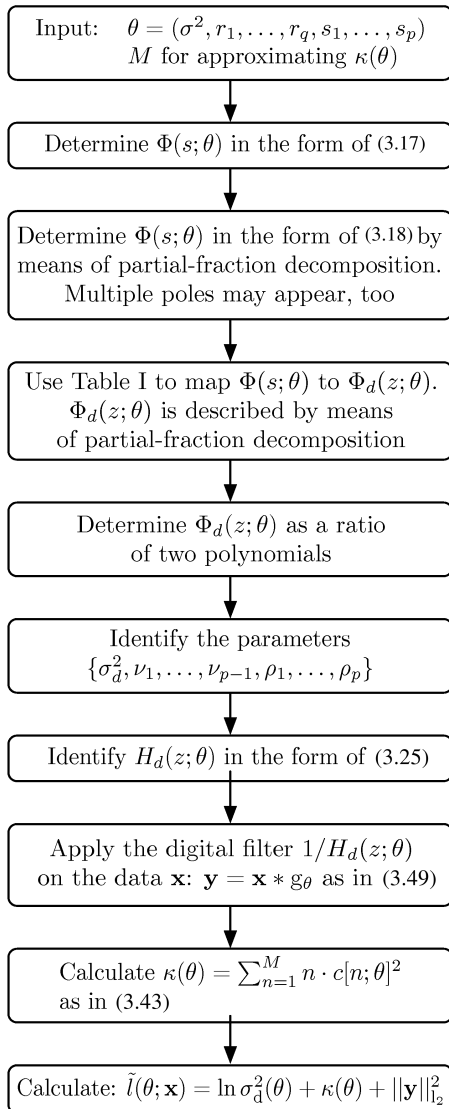


Figure 3.6: A detailed description of the log-likelihood function  $\tilde{l}(\theta; \mathbf{x})$  calculation. This workflow is part of the estimation algorithm of figure 3.5.

many ways of obtaining such a band pass spectrum, we provide a constructive way of doing so in appendix C. Initial values for  $\sigma$  can then be obtained by evaluating (3.19) at  $t = 0$  and equating it to the variance of the available data. The value of  $K$  may be derived from known physical constraints of the problem at hand. When there is no such knowledge, this value can be determined during the execution of the estimation algorithm: following table 3.2, likelihood values of local minima exhibit monotonicity; they decrease towards the global minimum and then start to monotonically increase as a function of the band index  $k$ . Such monotonicity can then provide a decision rule for setting a value for  $K$ . It is noted that this value does not depend on the size of the data or on the ARMA order. It is related to the sampling rate value. A single sampling interval value partitions the frequency axis into non-overlapping segments which are  $\pi/T$  wide, and, for different sampling interval values, a given continuous domain model can be associated with different segments. Associating a segment with a power spectrum amounts to allocating the peak response frequency to that segment. Every local minimum corresponds to a single segment and the choice of  $K$ , which is the number of segment the algorithm examines, should take this partitioning into account. The algorithm is described in figure 3.5 while complementary information is given in figures 3.6 and C.1.

The work of [77] suggests three algorithms for computing the continuous-time counterpart of a known discrete-domain model. These algorithms can be utilized for indirect estimation approaches, for which the sampled data is used for estimating the discrete-domain model first. Nevertheless, [77] points out two safeguards: existence of a continuous domain model is not guaranteed for all discrete-domain models, and uniqueness of the continuous domain poles is not guaranteed, either. If one, however, restricts the imaginary part of the continuous domain poles to be in the interval  $(-\frac{\pi}{T}, \frac{\pi}{T})$ , as suggested in [77], the poles can then be uniquely resolved. This uniqueness condition imposes an upper bound on the sampling interval  $T$ .

In our work, the discrete-domain model of Theorem 1 is always guaranteed to have a continuous domain counterpart, as it originates from the sampled version of the autocorrelation function. Also, our approach to the uniqueness problem is different than in [77]. Instead of imposing restrictions on sampling interval values, we are allowing for arbitrary values to be considered while exploiting the local minima properties of the likelihood function.

In particular, we find several sets of estimated parameters, each corresponding to a local minimum, and choose the one that corresponds to the global minimum rather than the one that resides in the baseband  $(-\frac{\pi}{T}, \frac{\pi}{T})$ . The work of [88] introduces two estimation algorithms. The first algorithm relies on the results [77] and it consists of two steps: 1) identifying parameters of a discrete-domain ARMA process by minimizing the variance of the prediction error, and 2) associating the discrete-domain parameters with a continuous domain ARMA process by means state space representation. Similar to the comparison we made with [77], when fitting a discrete-domain ARMA model to sampled data one does not take the zero-pole coupling of the discrete-model model into account. Such a fitting may result then in ARMA models that have no continuous domain counterpart. Uniqueness of the continuous domain poles is not guaranteed either. Furthermore, the maximum-likelihood estimator in the sampled case does not correspond to minimizing the variance of the the prediction error any more. The second algorithm in [88] consists of two steps: 1) replacing the differentiation operator by a difference operation and approximating the continuous domain innovation by a discrete-domain white noise with a variance that is proportional to the sampling rate, 2) optimizing the discrete-domain model to minimize the variance of the discrete-domain innovation process. The discrete-domain process that results in using the difference operator of [88] provides an approximated model to the sampled process. In particular, the autocorrelation sequence of this model does not necessarily correspond to the autocorrelation values of the continuous domain model one started with. Similar to the [88], our approach to discretizing the continuous domain model is simple too, as it involves partial fraction decomposition and finite weighted sums only. Nevertheless, our spline formulation does not depend on sampling interval values and it provides a discrete-domain process whose autocorrelation sequence is in a perfect match with the point-wise values of the autocorrelation function for any sampling interval value.

The work of [89] minimizes a likelihood function for parameters estimation. This function is expressed in terms of the discrete-time innovation process, allowing one to write the likelihood function as a product of independent Gaussian variables. The log likelihood function of [89] is then given by a sum. The innovation process corresponds to prediction error values and the work in [89] relies on a linear predictor. This predictor is recursively defined and

depends on point-wise evaluations of the autocorrelation function. The likelihood formulation we propose in our work is different. We take advantage of the fact that the samples are taken at fixed intervals and find a digital whitening filter that can be directly applied to the sampled data rather than using a recursion. Also, our suggested whitening filter is given in the  $z$ -domain, allowing one to use a Direct Form realization rather than a discrete-time convolution operation which is computationally more demanding. The recursion of [89] involves weighted sums of the sampled data, which is even more computationally demanding. Both approaches minimize the likelihood function by means of numerical optimization. Our formulation, however, allows one to consider large data sets of more than 100.000 sample values compared to less than 1.000 that are presented in [89]. Additionally, increasing the number of data points in our approach has no effect on the digital filter, whereas the approach of [89] requires one to calculate new additional weights. Another difference is the point-wise evaluation of the autocorrelation function. Our work relies on partial fraction decomposition of the Laplace transform (3.18). The time-domain expression for the autocorrelation function is then analytically derived, as shown in table 3.1. The work of [89] on the other hand, relies on a state-space representation of the continuous domain process. The point-wise evaluation requires a matrix exponential operation and several integration operations for every set of parameters. Such a method is computationally expensive when using numerical optimization. Furthermore, according to [90], these matrix operations are approximated by Riemann sums rather than defined by closed form expressions. The reference [90] is pointed out in [89]. Our likelihood formulation requires  $2p + 1$  evaluations of the autocorrelation function and there is no dependency on the size of the data,  $N$ . This stems from the fact that we characterize the whitening filter in the  $z$ -domain rather than by its time-domain values. The formulation of [89], on the other hand, requires one to evaluate the autocorrelation function at  $N$  distinct points.

The work of [91] relies on the Gaussian property of the sampled process and the likelihood function in [91] is expressed in terms of the autocorrelation matrix. The approximation that is suggested in [91] has the following properties: following the profile likelihood method, the variance of the innovation process is not included in the likelihood function formulation of [91]. This amounts to reducing the number of the estimated parameters. The innovation

variance is estimated instead by the sample variance of the data in [91]. In our work, the innovation variance appears in the likelihood formulation and we use the sample variance for initializing the numerical optimization only, as described in figure C.1 of appendix C. The continuous domain power spectrum model of [91] is assumed to have distinct roots. Our formulation does not include such an assumption. Power spectra of multiple poles are likely to arise during the numerical optimization process, and our approach is better designed to cope with such instances of power spectra. The likelihood function of [91] involves matrix inversion, which can be calculated by means of Cholesky factorization, as described in [91]. When the data size is too large, it is suggested in [91] to use a nearest-neighbor method, which corresponds to de-correlating the innovation process with respect to neighbouring sample values only. This means that for a fixed sampling interval value, one has to find the inverse of a relatively small matrix and repeatedly multiply it by small segments of data. Our work, on the other hand, uses a digital filter for de-correlating all of the data at once. The underlying assumption in our work is that the data size is very large, allowing one to approximate the inverted covariance matrix by a Toeplitz matrix. Theorem 2 guarantees convergence of the latter to the former with increasing number of samples. Furthermore, our digital filtering approach is computationally less complex than the nearest neighbour method, as it does not require repeated matrix multiplications. The proposed method takes advantage of the existence of local minima, rather than considering it as a difficulty. Our algorithm is designed to find several local minima and to choose the minimum value among them, as demonstrated in figure 3.2 and table 3.2. These local minima stem from aliasing and they can be found by properly allocating the initial conditions of the numerical optimization. The proposed algorithm minimizes the likelihood function using several sets of initial conditions and chooses the solution that corresponds to the minimum value among all the local minima that were found.

Numerical optimization was also used in [76, 89, 91] and our algorithm repeats the optimization procedures several times. The number of repetitions, denoted  $K$  in figure 3.5, is determined by the number of the local minima one wishes to find and it is a user-dependent parameter. Although our algorithm uses several minimizations of the likelihood function, our spline formulation allows for fast implementation as it involves the simple parameter mapping of Theorem 1 and a Direct Form implementation of a digital

filter.

### 3.6 Simulation results

The proposed approach was implemented in Matlab™ using non-constrained numerical optimization. It was compared with the polynomial B-spline estimator of [76]. Sampled signals were generated by filtering a discrete-domain Gaussian white noise with the digital filter (3.25). Several sets of the parameters  $\theta$  were considered. For each set, several Monte Carlo simulations were conducted; each simulation corresponds to a different sampling interval. Sampling interval values were chosen so as to capture different aliasing constellations. A single Monte Carlo simulation involved 500 experiments and every experiment was carried out using  $N = 10,000$  sample values. The variance parameter was set to  $\sigma^2 = 1$  and was unknown to the estimation algorithm. The value of  $\kappa(\theta)$  (3.43) was calculated using the first 500 terms in the infinite sum and the number of local minima that were examined was  $K = 2$ . The estimation error for a Monte Carlo simulation is the relative MSE (Mean Square Error) between the estimated parameter and the correct one. For example, the estimation error of a single parameter, e.g.  $a_0$ , is given by

$$\epsilon(a_0) = -10 \log \left( \frac{\frac{1}{500} \sum_{n=1}^{500} (\hat{a}_{0,n} - a_0)^2}{(a_0)^2} \right), \quad (3.60)$$

where  $\hat{a}_{0,n}$  is the estimation of  $a_0$  at the  $n$ -th experiment.

Experimental results for an AR(2) process are shown in figure 3.7 and in table 3.3. In figure 3.7, Monte Carlo simulation results for a continuous domain AR(2) process are shown. Sampling interval values were chosen so as to capture different aliasing constellations. The proposed exponential-based approach was compared with the polynomial model of [76]. The poles of the process are  $s_{1,2} = -0.2 \pm 7i$  and the variance is  $\sigma^2 = 1$ . The number of samples is  $N = 10,000$ . The  $\times$  mark that appears on the  $x$ -axis corresponds to the sampling interval  $T_{max}$  for which  $w_{max} = \pi / T_{max}$  is the frequency that corresponds to the value of the continuous domain power spectrum at  $-10$  [dB] respect to the peak value. The CRB is depicted for comparison purposes while other lines are depicted for presentation purposes.

Let us consider a generic unit measuring time or spatial extension of a simulated CARMA process: [time-unit]. The maxi-

imum value of the continuous power spectrum of the AR model of figure 3.7 is located at 5.05 [rad/time-unit] and the 3 [dB] frequencies are 4.09 and 6.10 [rad/time-unit]. A sampling interval value of  $T = 0.21$  [time-unit] corresponds to a sampling frequency of 29.92 [rad/time-unit] and it introduces minor aliasing effects. The value of  $T = 1.24$  [time-unit] corresponds to a sampling frequency of 5.07 [rad/time-unit] and it introduces substantial aliasing effects. The various aliasing configurations are depicted in figure 3.8

The CRB values were calculated according to [92] by using  $N = 1000$  and approximating its down shift for higher dimensional signals. Because of the approximated CRB computations and the limited number of Monte Carlo simulations, the analysed estimators may provide errors lower than the CRB. Figure 3.9 also describes the proposed exponential-based approach from a frequency domain point of view, emphasizing the flexibility of the exponential B-spline model to adapt to band-pass power spectra. Shown here is a periodogram of sample values (absolute square value of the DFT), periodized in the range  $[0, 4\pi]$ . The power spectrum of the AR(2) process is depicted as a solid red line. While the polynomial based estimator captures a base-band signal (black dashed line), the proposed approach adequately identifies the proper frequency band (magenta dashed-dotted line). The proposed windowing function corresponds to the Fourier transform of the interpolator  $\eta(t; \theta)$ . The poles of the process are  $s_{1,2} = -1 \pm 10i$  and the sampling interval is of a unit value, introducing prominent aliasing conditions.

An ARMA(2,1) estimation comparison is given in figure 3.10 and in table 3.4. We note that more sample values are required for the proposed estimation algorithm when the number of parameters increases. Our results indicate that the proposed approach outperforms the polynomial-based direct method while following the CRB at various aliasing constellations. It also guarantees that the estimated parameters correspond to a valid continuous domain model whereas this property does not necessarily hold true for other discrete-domain-based methods, such as the minimization of the prediction error variance. Unlike the proposed approach, the prediction error method might also result into a continuous domain power spectrum that occupies lower frequencies as it does not take aliasing into account.

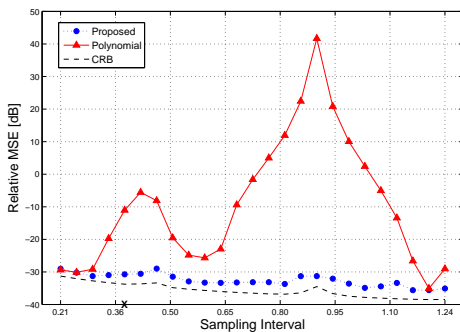
Table 3.3: Estimation comparison for AR(2) processes.

Power Spectrum $\Phi(s; \theta)$	Sampling Interval [time-unit]	Estimation Error [dB]					
		$\epsilon(a_0)$		$\epsilon(a_1)$		$\epsilon(\sigma^2)$	
		[76]	Proposed	[76]	Proposed	[76]	Proposed
$\frac{1}{(s^2+0.4s+49.04)(s^2-0.4s+49.04)}$ $(s_{1,2} = -0.2 \pm 7i)$	0.3370	-19.76	-30.98	-42.03	-55.90	-31.89	-39.60
	0.6827	-9.40	-33.25	-1.09	-58.61	-1.46	-39.20
	1.1150	-13.42	-33.39	-0.40	-59.63	-0.48	-37.81
$\frac{1}{(s^2+s+9.25)(s^2-s+9.25)}$ $(s_{1,2} = -0.5 \pm 3i)$	0.6059	-34.93	-34.97	-34.94	-45.89	-27.92	-36.99
	1.3050	-16.68	-35.79	-4.75	-43.77	-3.73	-35.81
	2.2370	6.93	-32.64	-1.86	-35.63	-8.16	-33.92
$\frac{1}{(s^2+6s+5)(s^2-6s+5)}$ $(s_{1,2} = -1, -5)$	0.9753	-21.09	-22.38	-17.95	-23.70	-11.73	-17.05
	2.3507	-3.65	-4.17	-3.44	-3.68	-1.19	-1.23
	3.6011	-1.83	-2.28	-1.31	-1.60	-0.27	-0.34

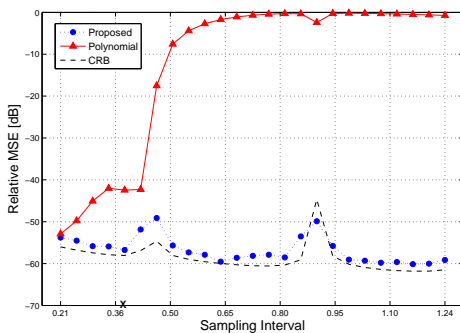
Table 3.4: Estimation comparison for ARMA(2,1) processes.

Power Spectrum $\Phi(s; \theta)$	Sampling Interval [time-unit]	Estimation Error [dB]							
		$\epsilon(a_0)$		$\epsilon(a_1)$		$\epsilon(b_0)$		$\epsilon(\sigma^2)$	
		[76]	Proposed	[76]	Proposed	[76]	Proposed	[76]	Proposed
$\frac{(s-3)(s+3)}{(s^2+2s+26)(s^2-2s+26)}$ $(r_1 = -3, s_{1,2} = -1 \pm 5i)$	0.3009	-18.52	-45.07	-32.18	-51.37	-12.76	-39.02	-23.37	-42.17
	0.6481	-8.44	-36.85	-18.96	-33.29	-1.65	-17.42	-10.43	-22.50
	1.1111	-11.60	-29.69	-0.05	-41.59	-0.26	-12.61	-11.83	-15.16
$\frac{(s-4)(s+4)}{(s^2+2s+101)(s^2-2s+101)}$ $(r_1 = -4, s_{1,2} = -1 \pm 10i)$	0.1911	-19.51	-33.71	-32.18	-51.37	-12.76	-39.02	-23.37	-42.17
	0.4115	-14.48	-34.78	-18.96	-33.29	-1.65	-17.42	-10.43	-22.50
	0.7055	-19.96	-29.00	-0.05	-41.59	-0.26	-12.61	-11.83	-15.16
$\frac{(s-3)(s+3)}{(s^2+8s+7)(s^2-8s+7)}$ $(r_1 = -3, s_{1,2} = -1, -7i)$	0.3250	-16.42	-21.08	-18.24	-17.26	-10.73	-20.36	1.30	-24.55
	1.3412	-4.63	-5.77	-3.49	-4.44	-4.56	-7.66	-15.31	-7.33
	1.8831	-2.97	-3.96	-2.00	-2.51	-3.65	-4.73	-9.11	-5.76

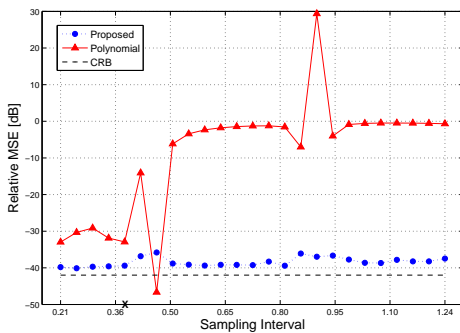




(a)



(b)



(c)

Figure 3.7: Estimation error comparison for an AR(2) process. Figures 3.7(a), 3.7(b), 3.7(c) show the estimation error for parameters  $a_0$ ,  $a_1$  and  $\sigma$  respectively.

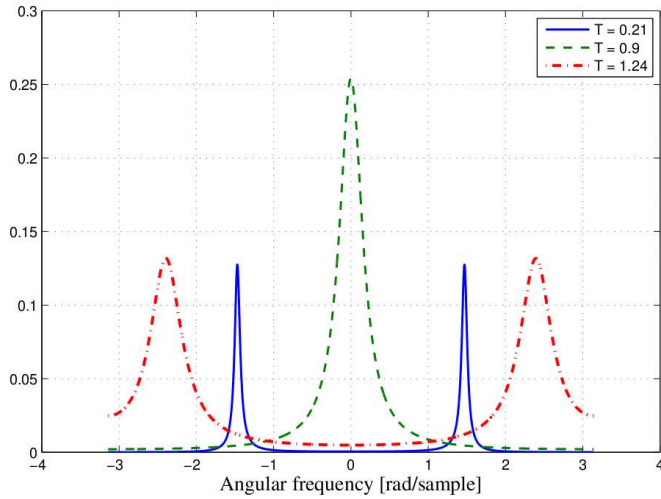


Figure 3.8: Power spectra of sampled processes. Shown here are power spectra of discrete-domain processes that originate from the same continuous domain process. Every power spectrum corresponds to a different sampling interval value. The poles of the continuous domain AR process are  $s_{1,2} = -0.2 \pm 7i$

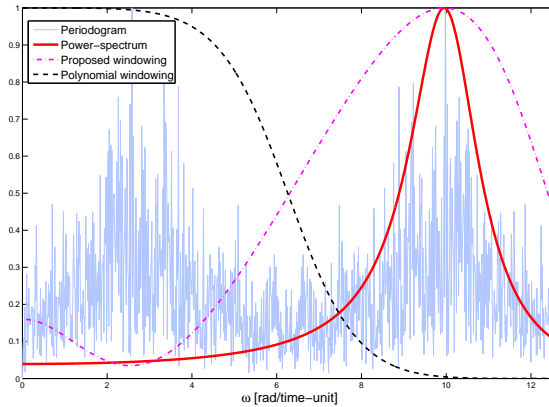


Figure 3.9: A frequency domain description of the proposed estimation approach.

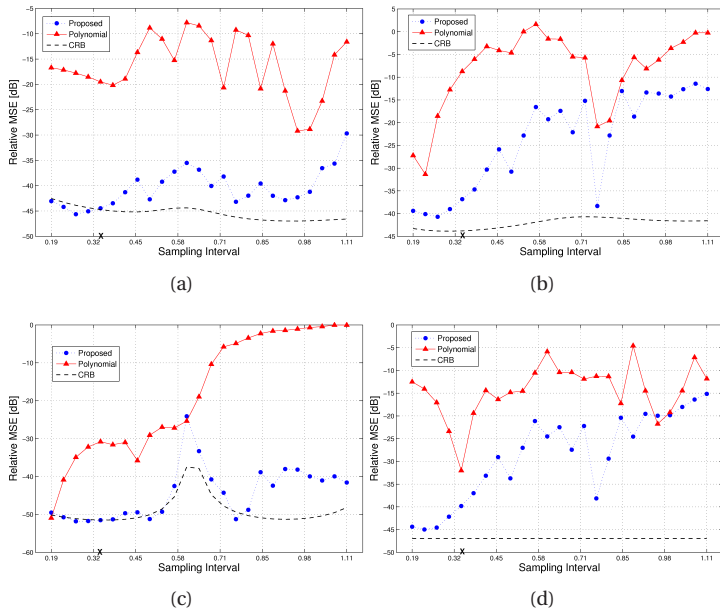


Figure 3.10: Similar to figure 3.7 for an ARMA(2,1) process. The poles of the process are  $s_{1,2} = -1 \pm 5i$ , the zero is  $r_1 = -3$  and the variance is  $\sigma^2 = 1$ . The number of samples is  $N = 100,000$ . Figures 3.10(a), 3.10(b), 3.10(c) and 3.10(d) show the estimation error for parameters  $a_0$ ,  $b_0$ ,  $a_1$  and  $\sigma$  respectively.

### **3.7 CARMA for ultrasound tissue characterization**

In ultrasound diagnostic, the acquired image is a rich source of information pertaining to the imaged tissue, even though some of it is not directly accessible at human visual inspection. To reveal this hidden information it is necessary to resort to computer-aided tissue characterization. This technique can capture several features carried by medical sonography and correlate them to the state of the imaged tissue. In particular, the statistical modeling of discrete ultrasound signals has been widely employed to derive interesting properties from estimated model parameters that are useful in the context of tissue characterization. Among them, shift-variant ARMA processes were found suitable to model ultrasound signals and images and to make the acquisition more robust with respect to the estimation of the uncorrupted tissue response [6, 93]. More recently, 2D ARMA modeling was proposed to improve computer-aided detection of breast tumors [94, 95].

Although CARMA stochastic processes have been successfully exploited for image modeling [78, 61], the continuous domain approach has never been investigated in the processing of ultrasound images. In this study, we consider the radio-frequency (RF) ultrasound signal as a sampled CARMA process and recover its CARMA parameters from the traditional ARMA coefficients through an indirect approach.

The ARMA coefficients are easily computed by minimizing the prediction error [55]. Then, the CARMA parameters can be estimated from their discrete-domain counterpart by mapping the ARMA model back into the continuous domain [67]. In doing so, both the exact discretization of CARMA model and the exponential properties of the autocorrelation function in the discrete and continuous domains must be taken into account. As was shown in [86], the exact link between ARMA and its corresponding CARMA model can be established, provided one is willing to interpolate the discrete signal by exponential B-splines specific to the considered model [56, 57]. In situations where a sufficiently high sampling frequency is employed, it is beneficial to exploit the dependence of the exponential B-spline on the ARMA parameters of the sampled model. This, along with a numerical optimization scheme, allows for the accurate estimation of the continuous domain model parameters.

Here, we analyze the CARMA parameters of ultrasound signals in terms of their ability to capture information about the concen-

tration of scatterers in imaged materials. For this purpose, we have created ultrasound images of phantoms with various concentrations of scatterers. The validation of CARMA and ARMA models in terms of discrimination performance shows that our continuous-based approach is more efficient than the traditional one. We believe that our unconventional ultrasound modeling is robust with respect to the effects of the system and has a good potential for tissue characterization.

We start by assuming that the spatial dependence of the effects of the acquisition system is weak. This allows us to focus on a segment of the ultrasound image and to consider its characteristics as locally shift-invariant [41]. As a consequence, the 1D discrete-time ultrasound signal  $x[n]$  can be modeled as a spatially invariant sampled CARMA process  $x(t)$ , where the continuous-time Gaussian white noise  $w(t)$  with variance  $\sigma^2$  represents the innovation signal.

For a CARMA model of order  $(n, m)$ , the spectrum of  $x(t)$  is expressed as

$$\Phi(j\omega; \theta) = \sigma^2 \left| \frac{\prod_{k=1}^q j\omega - r_k}{\prod_{k=1}^p j\omega - s_k} \right|^2 \quad (3.61)$$

$$= \sigma^2 \left| \frac{\sum_{k=0}^q b_k (j\omega)^k}{\sum_{k=0}^p a_k (j\omega)^k} \right|^2, \quad (3.62)$$

where  $\theta = (\mathbf{s}, \mathbf{r}) = (s_1, \dots, s_p, r_1, \dots, r_q)$  is the collection of parameters that contains the poles  $s_k$  and the zeroes  $r_k$  of the system, while  $a_k$  and  $b_k$  are the related polynomial coefficients.

In corollary equation (3.25) of section 3.2, it was shown that sampling the model (3.62) leads to a discrete ARMA model of order  $(p, p-1)$ . For the sake of simplicity, in the sequel we'll consider a unit-sampling interval. The discrete spectrum is then given by

$$\Phi_d(z; \theta) = \sigma_d^2 H_d(z; \theta) H_d(z^{-1}; \theta), \quad (3.63)$$

where

$$H_d(z; \theta) = \frac{\prod_{k=1}^{p-1} (1 - \nu_k z^{-1})}{\prod_{k=1}^p (1 - \rho_k z^{-1})} = \frac{\sum_{k=0}^{p-1} d_k z^{-k}}{\sum_{k=0}^p c_k z^{-k}}. \quad (3.64)$$

The sampled ultrasound signal, which represents the available data, is thus modeled as a discrete ARMA  $(p, p-1)$ . In this context the

autocorrelation function of (3.62) is derived from its discrete samples thanks to interpolation by shifts of the exponential B-spline function  $\beta(t; \theta)$  expressed in equation (3.21) and reported here for clarity:

$$\beta(t; \theta) = \mathcal{F}^{-1} \left\{ \prod_{k=1}^q (j\omega - r_k)(-j\omega - r_k) \cdot \prod_{k=1}^p \frac{(1 - e^{j\omega + s_k})(1 - e^{-j\omega + s_k})}{(j\omega - s_k)(-j\omega - s_k)} \right\} (t) \quad (3.65)$$

For this reason, the zeroes of the ARMA filter in (3.64) are the roots of the  $z$ -transform of the exponential B-spline samples  $\mathcal{B}_d(z; \theta) = \sum_{k \in \mathbb{Z}} \beta[k; \theta] z^{-k}$  that are located on the unit circle:

$$\mathcal{B}_d(z; \theta) = \sigma_d^2 \sum_{k=0}^{p-1} d_k(\theta) z^{-k} \sum_{k=0}^{p-1} d_k(\theta) z^k. \quad (3.66)$$

These expressions present a useful link between the discrete- and continuous domain ARMA models, and describe the mapping required to extract CARMA parameters from ARMA coefficients.

When the sampling frequency is high enough, the estimation of the CARMA parameters through indirect methods is efficient [67], and there should be no need to consider approaches focusing on several bands as presented in section 3.5. Fortunately, high frequency sampling operations are prevalent in ultrasound diagnostic. Thus, the derivation of a CARMA model by the inverse mapping of ARMA parameters offers good estimation performances.

We assume here to know the ARMA model orders, which can be estimated as the orders providing a good fitting of the model to the signal power spectrum. In this selection it is necessary to respect the constraint that the relative order of the model must be 1 for the signal to be a sampled CARMA process. The CARMA model order  $n$  is constrained to the ARMA orders, while order  $m < n$  can be fixed to provide an equal number of parameters for CARMA and ARMA models.

The ARMA parameters  $(c_k, d_k)$  and  $(\rho_k, \nu_k)$  can be computed by the standard minimization of the prediction error [55], which provides the estimated values  $(\hat{c}_k, \hat{d}_k)$  and  $(\hat{\rho}_k, \hat{\nu}_k)$ . While the link between the continuous domain poles  $s_k$  and the discrete-domain poles  $\rho_k$  is a trivial exponential mapping (in high frequency sampling conditions), the expression of the discrete-domain zeroes  $\nu_k$  is a complicated function of both continuous domain poles and zeroes, as can be seen from (3.66).

Since  $\mathcal{B}_d(z; \theta)$  depends solely on the CARMA parameters, it is possible to estimate the zeros  $r_k$  by the numerical minimization

**Algorithm 3.1** Estimation of CARMA parameters

- Discrete-domain estimation of parameters

Estimate  $(\hat{c}_k, \hat{d}_k)$  and  $(\hat{\rho}_k, \hat{v}_k)$  by minimizing the prediction error from  $x[n]$

Set  $\hat{\sigma}_d^2 = \mathbf{VAR}\{x[n]\}$

- continuous domain estimation of parameters

Set  $\hat{s}_k = \log(\hat{\rho}_k) \Rightarrow \hat{\mathbf{s}}$

Set  $\hat{\mathbf{r}} = \operatorname{argmin}_{\mathbf{r}} \|\hat{\mathbf{d}} - \mathbf{d}(\hat{\mathbf{s}}, \mathbf{r})\|^2$

Set  $\hat{\sigma}^2 = \hat{\sigma}_d^2 \cdot \left( \beta[1; \theta] \cdot \prod_{k=1}^{p-1} \frac{-1}{v_k} \cdot \prod_{k=1}^p e^{s_k} \right)^{-1}$  (3.34)

of the distance between the estimated ARMA coefficients  $\hat{d}_k$  and the function  $d_k(\theta)$ . The latter can be expressed as  $d_k(\mathbf{s}, \mathbf{r})$ , which highlights the separation between the collection of poles  $\mathbf{s}$  and the collection of zeroes  $\mathbf{r}$ . The parameters  $d_k(\mathbf{s}, \mathbf{r})$  can be built from the coefficients of the polynomial that corresponds to the roots of  $\mathcal{B}_d(z; \theta)$  in the unit circle. The collection of coefficients is indicated by  $\mathbf{d}(\mathbf{s}, \mathbf{r})$ . Likewise, the collection of estimated coefficients of the ARMA model numerator is  $\hat{\mathbf{d}}$ .

Setting  $\hat{\sigma}_d^2$  as the variance of the RF signal samples  $x[n]$ , we can estimate  $\sigma^2$  from the link between the variances of the discrete and continuous innovation signal (3.34).

The indirect method for CARMA model estimation is summarized in Algorithm 3.1. We used this algorithm to compute ARMA and CARMA parameters from a sequence of phantom images with various concentrations of scatterers. The two models were compared in terms of classification performance, as described in the next section.

### 3.8 Comparative study on phantom images

The ultrasound images were realized using as scatterers ultrafine polyamide particles (Orgasol<sup>®</sup>, Arkema, France) of diameter  $10 \pm 2 \mu\text{m}$  and density  $1030 \text{ kg/m}^3$ . The tissue-mimicking phantoms were prepared by mixing a specific concentration of Orgasol<sup>®</sup> particles with distilled water. A magnetic agitator preserved the solution homogeneous throughout the acquisition. In our experiments, we used five different mixtures having Orgasol<sup>®</sup> concentrations of 0.5%, 0.75%, 3%, 6%, and 12%. Low and high concentrations mimic random and dense mediums respectively.

The RF signal was acquired by a high-resolution ultrasound scanner (Vevo 770, Visualsonics, Toronto, Canada) at a central fre-

quency of 30 MHz and at a sampling frequency of 500 MHz with 8-bit resolution. For each phantom three video acquisitions were made. A selection of 90 scanlines from the central frame of each video sequence was used (about 300 samples per concentration). Along each scanline, the RF signal was used to compute the ARMA and CARMA parameters, as explained in section 3.7, except that we found it is more convenient to numerically optimize real polynomial coefficients, as opposed to complex poles and zeros.

In order to ensure a good fit of the available data, the orders of the CARMA model were chosen as  $m = 2$  and  $n = 4$ . Without loss of generality, we constrain the first polynomial coefficients to take a unit value, so that  $a_0 = b_0 = c_0 = d_0 = 1$ . The process is thus characterized by the four pole parameters  $\mathbf{a} = (a_1, a_2, a_3, a_4)$ , two zero coefficients, and the variance of the innovation signal, which we represent by the collection  $\mathbf{b} = (b_1, b_2, \sigma^2)$ . The orders of the ARMA model are constrained because of the sampling process performed on the continuous-time model: the number  $n = 4$  of poles is the same, while there are  $n - 1 = 3$  zeroes. The ARMA model parameters are represented by the collections  $\mathbf{c} = (c_1, c_2, c_3, c_4)$  and  $\mathbf{d} = (d_1, d_2, d_3)$ . Finally, the full set of parameters is a seven-attribute vector for both models.

The Mahalanobis distance along a single feature provides information about the individual discrimination power of the considered feature for distinguishing between two concentrations of scatterers [8]. We show in table 3.5 the Mahalanobis distance projected onto the direction corresponding to the coefficients on the rows belonging to the two analyzed models. We chose the most significant coefficients, separately for each model. The columns of the table represent the binary classification task of two consecutive concentrations of scatterers, where  $J_{[x,y]}$  is the Mahalanobis distance between the sample clusters with concentration  $x$  and  $y$ . From this first analysis, it is possible to observe that *i*) the higher the concentration of the scatterers, the more difficult the discrimination by a single coefficient is; *ii*) the CARMA coefficients clearly provide higher distances, which eases the classification task.

To extend these results to the classification of joint parameters, we performed several multiclass discrimination experiments where all available parameters were involved, for both models. The learning scheme used was a support-vector machine with Gaussian kernel, whose parameter had been tuned by cross-validation, separately for each classification model. We report in table 3.6 the rate of misclassification (averaged on 20 experiments) for the four



Table 3.5: Mahalanobis distance for various concentrations.

<b>Model</b>		$J_{[.5\%,.75\%]}$	$J_{[.75\%,3\%]}$	$J_{[3\%,6\%]}$	$J_{[6\%,12\%]}$
CARMA	$a_4$	<b>14.01</b>	10.55	<b>2.13</b>	<b>4.44</b>
	$b_2$	11.30	<b>39.18</b>	$5 \cdot 10^{-4}$	0.17
ARMA	$c_4$	0.56	5.22	0.33	0.37
	$d_3$	2.73	12.02	0.19	0.08

multiclass discrimination problems listed in the first column. The classifier was first trained on CARMA and ARMA pole coefficients **a** and **c**. Then, it was trained on the zero coefficients and variances **b** and **d** of both models. Finally, the learning was performed on all the available coefficients for each model.

The classification results show that the proposed CARMA coefficients are more sensitive to differences in the concentration of scatterers, above all when the discrimination is multi-class. continuous domain parameters favour a better recognition score and represent tissues of different concentrations in more compact and better discernible clusters, as illustrated in figures 3.11 and 3.12 for the discrimination problem with 5 categories of scatterers concentration. In these figures linear discriminant analysis [8] is used to reduce data dimensionality and to find the number  $N_{classes} - 1$  dimensional space, where mapped samples present maximum inter-class distance and minimum intra-class variance. The seven-dimensional samples are, thus, projected onto a four dimensional space and normalized; two planes of the projection space are visualized in figures 3.11 and 3.12. The bivariate distribution of samples in the projection planes is described by a 2D Gaussian fitting for each category, whose ellipse at a probability density function of 80% is shown in the figures. We observe that the clusters are better defined in the space of CARMA coefficients than in the space of the traditional ARMA parameters. Furthermore, as regards higher concentrations [3%, 6%, 12%] samples, although the discrimination task is more difficult, a good classification of these categories is provided by CARMA parameters analysis considering all the projection directions, while ARMA coefficients do not bring more separation, as shown in figure 3.12.

The difference of performances between CARMA-based and ARMA-based classifiers of the last two columns in Table 3.6 is also verified by resampled paired t-tests. These statistic tests, perfor-

Table 3.6: Rate of misclassification: [mean value  $\pm$  standard deviation] of 20 experiments (<sup>1</sup>CARMA model, <sup>2</sup>ARMA model).

Classes Concentration	Features for classification	
	poles	
	a <sup>1</sup>	c <sup>2</sup>
[0.5%, 0.75%]	0.007 $\pm$ 0.005	0.024 $\pm$ 0.006
[0.5%, 0.75%, 3%]	0.010 $\pm$ 0.003	0.024 $\pm$ 0.004
[0.5%, 0.75%, 3%, 6%]	0.116 $\pm$ 0.006	0.127 $\pm$ 0.006
[0.5%, 0.75%, 3%, 6%, 12%]	0.126 $\pm$ 0.007	0.135 $\pm$ 0.006
	zeroes	
	b <sup>1</sup>	d <sup>2</sup>
[0.5%, 0.75%]	0.013 $\pm$ 0.005	0.008 $\pm$ 0.003
[0.5%, 0.75%, 3%]	0.008 $\pm$ 0.003	0.032 $\pm$ 0.006
[0.5%, 0.75%, 3%, 6%]	0.156 $\pm$ 0.007	0.270 $\pm$ 0.008
[0.5%, 0.75%, 3%, 6%, 12%]	0.273 $\pm$ 0.009	0.368 $\pm$ 0.010
	all parameters	
	(a, b) <sup>1</sup>	(c, d) <sup>2</sup>
[0.5%, 0.75%]	<b>0.007</b> $\pm$ 0.003	0.010 $\pm$ 0.003
[0.5%, 0.75%, 3%]	<b>0.003</b> $\pm$ 0.002	0.012 $\pm$ 0.003
[0.5%, 0.75%, 3%, 6%]	<b>0.069</b> $\pm$ 0.006	0.113 $\pm$ 0.006
[0.5%, 0.75%, 3%, 6%, 12%]	<b>0.085</b> $\pm$ 0.006	0.122 $\pm$ 0.006

med on the misclassification errors obtained in 20 experiments, indicate in all the four classification problems an increment in performance of CARMA-based classifier respect to ARMA-based classifier at a 5% significance level. The improvement of performance is significant and encourages the application of continuous-time models on *in vivo* images for the characterization of ultrasound tissues.

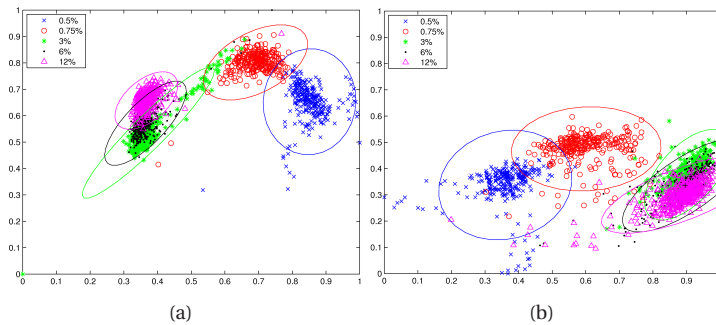


Figure 3.11: Scatter plots of normalized CARMA parameters in figure (a) and ARMA coefficients in figure (b) for several scatterer concentrations. The plane visualized here is given by the first two directions selected by linear discriminant analysis.

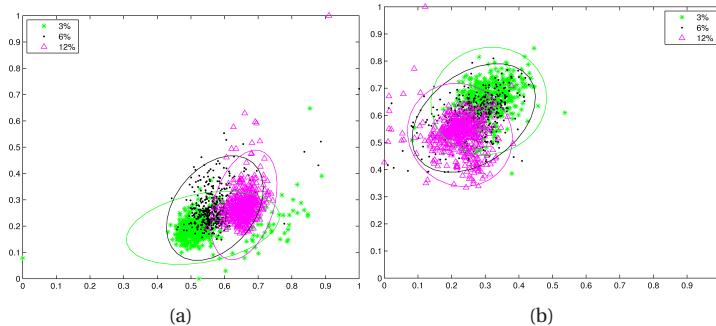


Figure 3.12: Scatter plots of normalized CARMA parameters in figure (a) and ARMA coefficients in figure (b) for concentrations [3%, 6%, 12%]. The plane visualized here is given by the last two directions selected by linear discriminant analysis.

### 3.9 Conclusion

In this chapter we introduced a new formulation of CARMA models based on exponential splines. We further proposed a maximum likelihood estimator for continuous domain ARMA parameters from sampled data. It utilizes the exponential B-spline framework while introducing an exact zero-pole coupling for the sampled process. The relation between the autocorrelation function and the autocorrelation sequence of the sampled process was investigated in both time- and frequency domains. Our approach

relies on the fact that the Cramér-Rao bound can be made arbitrarily small by increasing the number of samples while fixing the sampling interval at an arbitrary value. The likelihood function of the sampled process was investigated too. In particular, it was shown that it possesses local minima that originate from aliasing. The global minimum, however, corresponds to the maximum likelihood value.

The only assumptions that were made throughout this work are that the number of available samples is relatively large and that the model orders are known, allowing one to replace the whitening matrix by a digital filter. This approximation was then shown to be valid when considering expected values of the likelihood function. Experimental results indicate that the proposed exponential-based approach closely follows the Cramér-Rao bound for various aliasing configurations, while imposing no restrictions on sampling rate values.

Additionally we presented a continuous-time ARMA model for RF ultrasound signals and proposed an indirect approach to compute CARMA parameters based on the properties of exponential B-splines. This work includes a comparative study of the proposed CARMA coefficients and of the traditional ARMA parameters that are widely used for the analysis of ultrasound images. We performed experiments on phantom images specifically designed with various concentrations of scatterers and we focused on the comparison of CARMA and ARMA features in terms of discrimination performances.

Experimental results reveal that the new CARMA features are able to capture the information about the concentration of scatterers more efficiently. Thus, they can be extremely useful for the characterization of *in vivo* ultrasound tissues.

# Four

---

## IMPROVING STANDARD PROSTATE BIOPSY PROTOCOL

---

*It is a very sad thing that nowadays there is so little useless information.*

---

Oscar Wilde (1854-1900)

THE practical application of this thesis is the improvement of standard prostate biopsy protocols guided by ultrasound, used in clinical routine as the more reliable tool for prostate cancer diagnosis.

During a biopsy examination the urologist removes from 8 to 12 tissue samples, usually with a needle guided by some imaging tool. Then a pathologist checks for cancer cells. As the bioptic procedure is generally performed in the doctor's office with local anesthesia, the most preferred guiding tool is transrectal ultrasound.

The current clinical procedure for prostate cancer biopsy has several limitations, above all due to the variable appearance of cancer on ultrasound. Because of this variability, TRUS images visual inspection is characterized by low specificity. For this reason TRUS guided biopsy is performed by uniformly sampling the gland without any information about the disease probability of the imaged tissues.

Nevertheless computer-aided analysis of TRUS images can be exploited to highlight suspicious regions, providing useful information to the physician and guiding the standard prostate biopsy both as a positioning and detection tool.

As stated in section 1.3, standard image-guided biopsy is not lesion-directed, but it is based on a systematic sampling of those regions of the prostate with higher cancer incidence, as the peripheral zone. As the disease may present several foci, this standard procedure is quite inefficient with a detection rate of 25%.

The uniform sampling of some regions of the prostate aims at maximizing the efficiency of the procedure, but carries some randomness in the selection of cores which can be compensated for by lesion-directed methods.

The computer-aided analysis of TRUS images proposed in this work is aimed at improving the current biopsy protocol by marking the interesting regions and discarding the healthy tissues. As a consequence, a reduction in the number of unnecessary sampled cores will be possible by lesion-directed biopsy without losing diagnostic power.

An essential requirement for lesion-directed biopsy to be feasible is that the algorithm performing lesion detection is fast enough to perform real-time processing. This is necessary to have US biopsy sessions as short as the ones performed today and to preserve the main advantage of the US respect to others imaging modalities: the real time feedback.

Furthermore in order to maximally exploit the current technology already present in healthcare institutions and prevent extra costs due to physician training or longer biopsy sessions, radiologists must be provided with a fast and user-friendly tool. Non real-time detection tools are interesting for research purposes but not appropriate for clinical employment.

In this chapter a tool for real-time Computer-Aided Biopsy (rt-CAB) is presented and the procedure used to train it on a medical ground truth is described. rtCAB enhances TRUS video stream with a false color overlay image, and suggests the physician where to sample thus reducing the total number of cores.

The proposed detection tool was implemented using parallel programming on a specific graphic card, such that the recognition is processed at the same time as the visualization of the standard US video sequence, generally acquired at 24-50 fps.

On the contrary, the training of the detection algorithm, which is not included in the classification task performed during biopsy, is implemented using Matlab<sup>TM</sup>.

## **4.1 Database collection**

The main limits preventing the realization of an automatic detection tool in ultrasound imaging are the non standard orientation of the images and the absence of a public database for algorithms testing and validation.

The current TRUS guided biopsy protocol overcomes the problem of free US image orientation, in fact it exploits a standard visualization of the gland which appears in a fixed region of the image with a fixed position with respect to the biopsy needle. This provides robust clinical references for images interpretation.

In order to train the detection tool for suspicious regions recognition, a large TRUS video sequences database was collected in collaboration with the department of Urology of S. Orsola Hospital in Bologna. A ground truth database was realized, by recording radio-frequency US video sequences during biopsy sessions since November 2009.

The probe used during the sessions is an Esaote<sup>TM</sup> EC123, employed with US central frequency of 7.5 MHz, as in the study described in chapter 2. The dataset employed in this work contains about 400 video sequences, one for each sampled core, belonging to 42 patients. This ground truth is continuously updated and increased, alongside the proceeding of the research project, and so far about 1000 US video sequences belonging to 120 patients have been recorded.

The histopathological analysis of each core performed by an expert physician provides information about the pathological state of the imaged tissue, reporting, in case of disease, the percentage of tumour in the core and the corresponding Gleason score.

Additionally the PSA value, the DRE result, the prostate volume and the patient clinical history are likewise recorded to complete the ground truth database. Table 4.1 shows the list of the different typologies of tissue collected in the database and the number of cores belonging to each category.

An example of the form used during biopsy sessions, which will be filled with clinical information, DRE and TRUS analysis results, is shown in figure 4.1.

Table 4.1: Ground Truth Dataset

code	# cores	Description	Label
M	18	Cores with tumour percentage $\geq 70\%$	Malignant
UM	46	Cores with tumour percentage $< 70\%$	Unknown
BB	143	Benign prostate tissue <sup>1</sup>	Benignant
BB*	36	Other benignant tissue <sup>1</sup>	Benignant
BM	81	Benignant prostate tissue <sup>2</sup>	Benignant
BM*	39	Other benignant tissue <sup>2</sup>	Benignant
U	40	Precancerous lesions: Prostatic intraepithelial neoplasia PIN(26) Atypical small acinar proliferation ASAP(14)	Unknown
TOT	403	Total number of cores 42 patients (22 pathological)	

<sup>1</sup> from healthy patient

<sup>2</sup> from unhealthy patient

Although several video sequences are available for each examined patient, only a small portion of the ultrasound image is used in this study. Since the only certain pathological information is in the cylinder sampled by the needle, it is necessary to select the very region covered by the needle profile in the frames preceding the insertion.

The detection of the biopsy needle position is easily performed in the frames where the insertion is visible, due to the high reflectivity of the needle and its fixed location. This information is used in the frames preceding the insertion to select the core region, which is then segmented in 20 ROIs of size  $101 \times 11$ . Only the ROIs included in the core region are used in this study, since there is no sure knowledge about the tissues not submitted to pathological analysis.

In particular automatic needle region segmentation is possible in the scan-converted space by analysing the arithmetical difference between consecutive US frames, after being turned to grey-level images. Applying a suitable threshold to the difference image we can detect the frame where needle insertion begins and also highlight the needle zone.

The application of a threshold to the image given by the difference between the insertion frame and the previous frame does not provide a perfect shape of the needle, because of the noisy appearance of US images. What appears in this image is a cluster of small unconnected regions grouped in the needle zone.





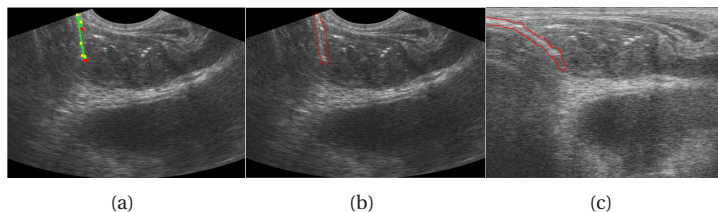


Figure 4.2: Segmentation process for needle region detection. In figure 4.2(a) it is shown the connection of small regions by Hough transform; the needle edges are visualised in the scan-converted space in figure 4.2(b) and in the RF space in figure 4.2(c).

These regions are then connected exploiting the Hough transform [96] and the fact that the needle is a line in the scan-converted space. Since the needle line is always almost vertical the angle range for connections search in the Hough transform can be constrained to  $[-20, 20]$  degrees.

Once the needle profile is detected, it is regularised by morphological operations of dilation and erosion. Finally the needle region contour is outlined by edge detection. The found needle profile is converted in the RF space where it appears curve and finally mapped onto frames preceding the insertion to perform ROI selection. The segmentation process is illustrated in figure 4.2.

In spite of the histopathological analysis, there is still some uncertainty in the ground truth, due to the unknown distribution of the tumour in the core. For this reason, only malignant cores with a tumor covering more than 70% of the sample (category M in table 4.1) are used to train the classification system, as it will be explained in the next section. Furthermore cores diagnosed with precancerous lesions as PIN and ASAP are ambiguous and cannot be included in the learning process as healthy or unhealthy regions. These observations highlights the necessity of a complex learning scheme also able to deal with unlabelled data.

## 4.2 Processing and learning scheme

In order to attain real-time processing capabilities, rtCAB makes use of a twofold signal processing datapath, a schematic representation of which is depicted in figure 4.3. Along the first branch, the incoming RF ultrasound signal undergoes envelope detection and

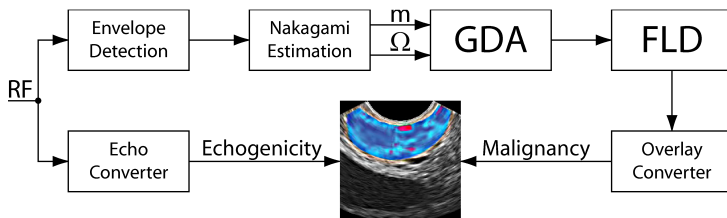


Figure 4.3: Signal processing flow graph of rtCAB: from left to right the incoming RF signal undergoes envelope detection, statistical estimation, non-linear projection onto the feature space and linear classification. Results are scan-converted and superimposed to the original ultrasound image.

scan conversion; this path has precedence on the second branch as during prostate biopsy the physician is mostly concerned about having a real-time video feedback of the organ he is sampling.

According to this requirement, a fast digital envelope detection algorithm [97] and an high speed digital scan converter [98] based on bilinear interpolation were chosen. Algorithms were optimized and combined to seamlessly perform detection, logarithmic compression and conversion in one step. As the chosen envelope detection algorithm operates memoryless only on a few input samples, parallel processing techniques were also employed to further speedup this step. By exploiting the computational resources made available by CUDA<sup>TM</sup> [99] enabled graphic card, a processing time around 1 ms was obtained on a GPU with 240 cores during the conversion of a  $2792 \times 250$  RF image to a  $300 \times 300$  pixel US image.

The second branch is devoted to the classification task. First a list of ROIs is automatically generated by tessellating the RF image with fixed size rectangular regions. Then, an improved version of the envelope detection algorithm used in the first branch is applied to the incoming data. Comparing the output of this stage to the output of a gold standard, Hilber Transform based, envelope detector, an accuracy of  $-49.2$  dB nRMSE was measured on standard acquired US data.

Following envelope detection, Nakagami distribution [26] parameter  $m$  and scaling factor  $\Omega$ , introduced in section 2.1, are estimated over each ROI. This distribution can give a good description of RF envelope statistics and is able to well capture different scattering structures. As discussed in [26], the most discriminant

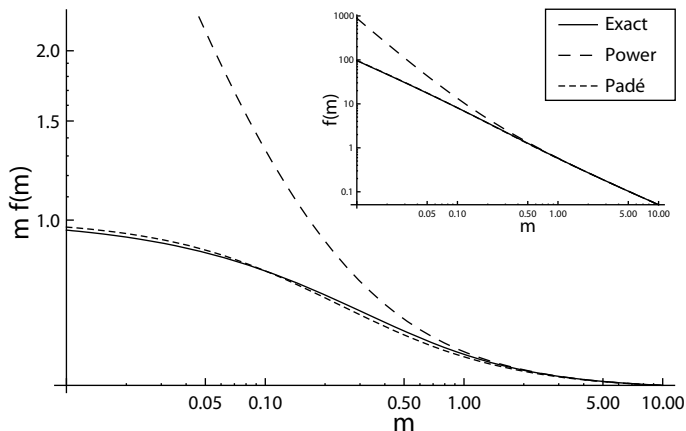


Figure 4.4: Comparison of the power series approximation in the ML Nakagami estimator and the corresponding Padé Approximant: the proposed solution is much more accurate, especially for low values of the shape parameter.

feature is the Nakagami parameter: in fact, different values of  $m$  corresponds to different scattering types. Also the scaling factor is important: by conveying information about the local scatter energy, its logarithm is strongly related to the log-compressed image the physicians usually employ to detect prostate cancer.

While estimating  $\Omega$  is not computational intensive, the ML estimation of  $m$  requires solving the nonlinear equation  $f(m) = \psi(m) - \ln(m) = \Delta$  where  $\psi(m)$  is the digamma function and  $\Delta$  is a quantity depending on data. As no closed form solution is known for this equation, authors in [100] suggest to asymptotically expand  $f(m)$  by means of a power series with strictly negative exponents. However a much better approximation can be obtained by means of Padé Approximants [101]. In fact, substituting  $f(m)$  with its approximant of order (1,2), a rational function in the form  $\hat{f}(m) = \frac{p_0 + p_1 m}{q_0 + q_1 m + q_2 m^2}$  can be used to obtain a closed form ML estimator. Parameters  $p_i$  and  $q_i$  were obtained by applying numerical optimization techniques. As shown in figure 4.4, the proposed solution is much more accurate, especially for low values of  $m$ .

The next stage is responsible for the classification of each ROI as malignant or not malignant basing on both Nakagami distribution parameters. The main blocks of the classification scheme are

a non linear feature extraction stage based on Generalized Discriminant Analysis (GDA) with Gaussian kernel[49] and a linear classification through the Fisher Linear Discriminant (FLD)[35] with an optimized shift of the decisional threshold.

Classification of each ROI, being each other independent, is performed exploiting CUDA<sup>TM</sup> parallel processing acceleration. The classification of a single  $101 \times 11$  ROI requires a mean time of  $47.6 \mu\text{s}$  while  $71.4 \mu\text{s}$  are necessary for a larger  $200 \times 20$  one. Results are then superimposed to the scan converted image to form a malignancy map which helps the urologist during the sampling procedure.

The classification model selection is based on a two-step learning scheme. The first step concerns the selection of the training set and the variance  $\sigma_{rbf}$  of the GDA kernel, tuned to obtain the maximum sensitivity. The second step optimizes the linear classifier threshold in order to provide the classification system with maximum positive predictive value.

The performance statistics considered in the first step are ROI-based, such that the sensitivity to be optimized is the number of well classified malignant ROIs over the total number of ROIs classified as unhealthy in the validation set. The data set for this step includes only a fraction of the database shown in table 4.1. In particular, it contains the benign ROIs coming from healthy patients (143 BB, 2860 ROIs) and the malignant ROIs belonging to the cores with tumor volume percentage larger than 70% (18 M, 360 ROIs). The latter choice is necessary to make negligible the uncertainty on the distribution of the disease in the core.

In the first step of the model selection, 500 random data sets of 360 unhealthy and 360 healthy ROIs are generated and separated in training set (600 ROIs) and validation set (120 ROIs). For each data set, a cross-validation over 30 values of  $\sigma_{rbf}$ , equally distributed between 0.1 and 6, is performed. The output of the model selection step is constituted by the training set  $[\hat{\mathbf{X}}_{trn}, \hat{\mathbf{c}}_{trn}]$ , including the selected matrix data  $\hat{\mathbf{X}}_{trn}$  and the corresponding classes  $\hat{\mathbf{c}}_{trn}$ , and the GDA kernel variance  $\hat{\sigma}_{rbf}$  corresponding to the maximum sensitivity in the classification of the validation set. The process for the first part of the learning scheme is shown in algorithm 4.1.

The choice of sensitivity as ROI-based criterion to select the model is arbitrary: accuracy, specificity and PPV may be used as well. Other criteria were also tested and the sensitivity was chosen as the metrics providing the best core-based classification perfor-

**Algorithm 4.1** Training set and GDA parameter selection

---

```

 $[X_{trn}, c_{trn}]$ : training set data matrix and labels
 $[X_{val}, c_{val}]$ : validation set data matrix and labels
for  $iter = 1$  to 500 do
  select random  $[X_{trn}, c_{trn}]$  and  $[X_{val}, c_{val}]$ 
  for  $\sigma_{rbf} = 0.1$  to 6 do
    Train on  $[X_{trn}, c_{trn}]$  using  $\sigma_{rbf}$  as GDA kernel variance
    Test on  $[X_{val}, c_{val}]$ 
    Compute ROI-based Accuracy:  $ACC([X_{trn}, c_{trn}], \sigma_{rbf})$ 
  end for
end for
Output:  $(\hat{X}_{trn}, \hat{c}_{trn}), \hat{\sigma}_{rbf} = \text{argmax}_{ACC}([X_{trn}, c_{trn}], \sigma_{rbf})$ 

```

---

mance as it will be shown next.

In the second step of the learning process the threshold of the FLD is tuned according to core-based statistics in order to maximize the PPV of the classification model over the whole dataset. The training set for this step is composed of all the cores (benign, malignant and unknown) of all the 22 unhealthy patients for a total of 204 cores.

Using the classification model selected in the previous step, a cross-validation is performed on the values of the threshold varying between the minimum and the maximum value of the FLD discriminant function. The number of thresholds considered is 1000, ranging from a model with 100% ROI-based specificity and a model with 100% ROI-based sensitivity. For any value of the threshold, the classification model is applied to the 204 cores and some core-based statistics are computed for each patient. In this case a core is considered unhealthy if at least one of its ROIs is classified as malignant.

As a consequence a core-based true positives value (TP) is the number of malignant cores correctly classified; likewise the core-based false positives value (FP) is the number of misclassified benign cores. As stated in section 1.2, the PPV value is given by  $\frac{TP}{TP+FP}$  and represents the probability of prostate cancer positive-ness of a pathological gland according to the detection tool.

The final value of the threshold is selected as the one providing the largest median PPV over all pathological patients. Whether there are several threshold values corresponding to the maximum PPV, the highest value is selected in order to maximize the sensitivity SE constrained to the PPV maximization. The second step of the learning scheme is shown in algorithm 4.2.

**Algorithm 4.2** Threshold selection

---

```

 $th = 0$  initial threshold at zero
C classification model defined by  $(\hat{\mathbf{X}}_{trn}, \hat{\mathbf{c}}_{trn}, \hat{\sigma}_{rbf}, th)$ 
 $d$ : FLD discriminant function
for  $th = \min(d)$  to  $th = \max(d)$  do
  for  $patient = 1$  to 22 do
    Classify by C any core of this patient
    Compute core-based PPV( $patient, th$ )
    Compute core-based SE( $patient, th$ )
  end for
end for
 $PPV_m(th) = \text{med}(PPV(patient, th))$  median over patients
Set  $\hat{th} = \arg \max PPV_m(th)$ 
if  $\hat{th}$  not unique then
   $\hat{th} = \max(\hat{th})$  corresponding to the largest SE
else
   $\hat{th} = \tilde{th} \rightarrow$  SE constrained to PPV
end if
Output:  $\hat{th}$ 

```

---

### 4.3 Classification results

The described learning scheme selects a threshold for the classification model providing a median sensitivity of 100% and a median PPV of 39%, as shown in figure 4.5.

The standard double sextant biopsy is equivalent to considering all the sampled cores as *a priori* malignant, thus it is characterized by a PPV over the studied data set of 24%. Therefore the proposed classification scheme improves the positive predictive value of 65% without losing diagnostic power. In fact, called  $ppv_0$  the PPV of the TRUS guided biopsy, the probability of correctly diagnosing prostate cancer after extracting  $N_0$  cores, can be computed as  $1 - (1 - ppv_0)^{N_0}$ . Since the PPV of rtCAB in guiding biopsies,  $ppv_1$ , is higher than  $ppv_0$ , the number of cores may be reduced keeping the same diagnostic power:  $N_1 = N_0 \frac{\log(1 - ppv_0)}{\log(1 - ppv_1)}$ . Consequently, according to the results of this study, it is possible to achieve the same diagnostic value of a standard prostate biopsy (8-12 cores), with no more than 7 cores in rtCAB guided biopsy.

### 4.4 The future

The detection tool rtCAB can be further improved by adding more ultrasonic features to Nakagami parameters and by assuring their real-time computation. Interesting attributes, which can improve performance detection by keeping low time consumes, are the Unser textural features, described in appendix A.

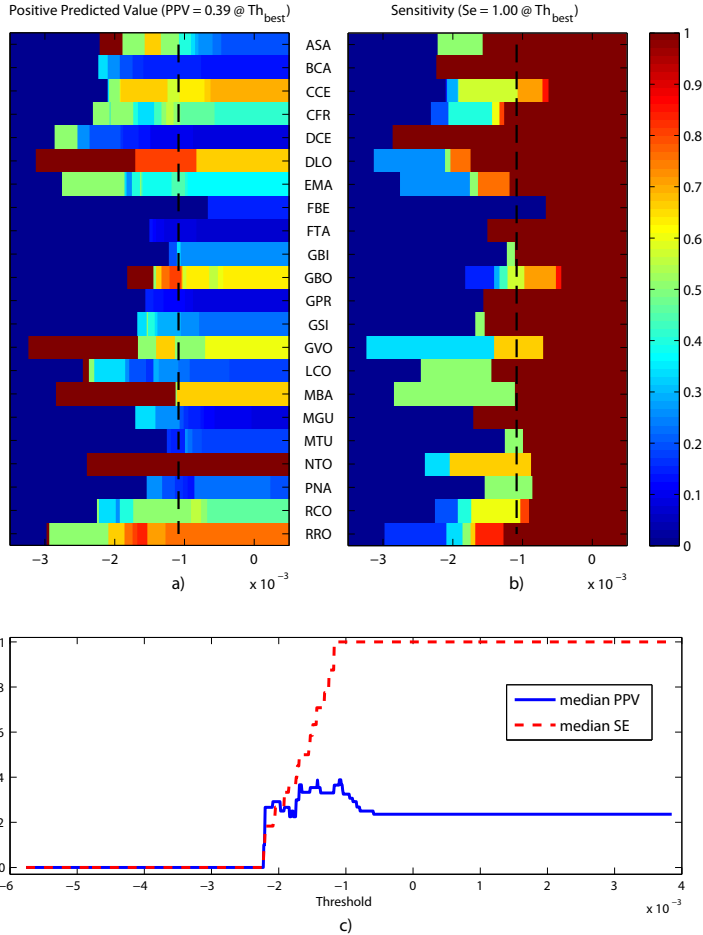


Figure 4.5: Learning process results. figure 4.5(a) shows the core-based PPV of the classification system. At right the color bar shows the color coding for the sensitivity and PPV values. figure 4.5(b) shows the core-based sensitivity of the classification system for any patients, listed on the y-axis with their representative code. The x-axis represents the 1000 threshold values, while its optimal value is marked by a vertical line. figure 4.5(c) shows the median results over all patients.

When performance results will be clinically satisfactory on a ground truth dataset containing about 100 pathological patients, the designed detection tool will be installed on a machine for ultrasound acquisition, giving complete freedom to radiologists of tun-



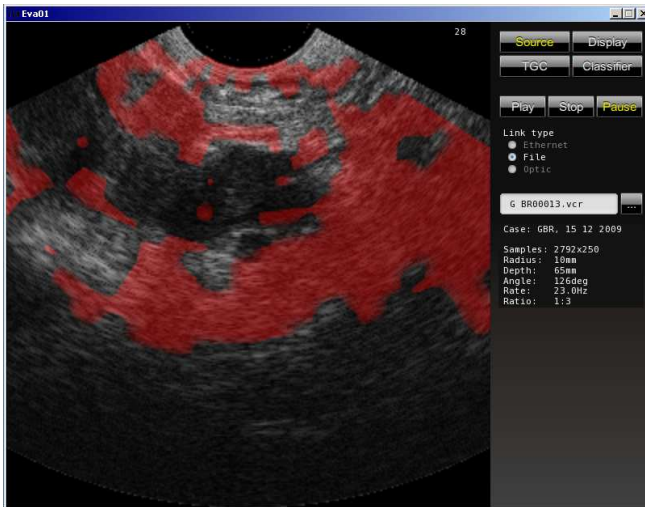


Figure 4.6: Interface for real-time lesion detection proposed to radiologists to be used during prostate biopsy.

ing detection characteristics and location through a user-friendly graphical interface. An example of such interface is shown in figure 4.6.

A double screen or a double window on the machine display will provide both the malignancy map visualised on the US image and the traditional B-mode signal as imaged in standard machines. Radiologists will thus receive on-line information about suspected regions and will choose, according to his experience, the cores of the gland to sample.

Once a prototype machine will be ready to use, a clinical trial will be performed on new patients, analysing both the standard double sextant sampled cores and the targeted cores obtained by rtCAB examination. The results of the clinical trial will show the actual efficiency of the new US-guided lesion-directed protocol for prostate biopsy.

## 4.5 Conclusion

TRUS guided biopsy is a fundamental step in the current clinical procedure for prostate cancer diagnosis. So far TRUS images have been used to correctly position the probe during biopsy, while the information extracted from TRUS images, which can be very use-

ful in detecting suspicious regions, has never been systematically exploited to guide biopsy. The standard double sextant biopsy protocol equally samples the gland, randomly selecting the tissues to be analyzed. This procedure does not take into account the risk of sampling a possibly benignant core, not significant for cancer detection.

In this work we propose a real-time Computer Aided Biopsy tool, which employs Nakagami parameters extracted from TRUS images to detect suspicious regions and guides prostate biopsy. The proposed detection tool uses a non linear classification process which takes advantage of Nakagami parameters computed through an original Padé Approximant estimator. The non linear classification algorithm was trained on a 400 cores ground truth database, collected thanks to the department of Urology at S. Orsola Hospital in Bologna. In addition, the resulting algorithm was implemented through CUDA parallel processing and it is characterized by a frame rate of 30 fps, suitable for a clinical application.

The proposed rtCAB is able to reduce the biopsy core number from 8-12 to 7, preserving the same diagnostic power of a classical double sextant protocol.

---

## CONCLUSIONS

---

THE research activity presented in this work was developed during a PhD program at University of Bologna in Italy from 2008 to 2010 and was partly carried out at École Polytechnique Fédérale de Lausanne in Switzerland.

This research focused on ultrasound image processing for tissue characterization with the purpose to support minimally invasive detection of pathologies. An important clinical target has been the spine of the different studies proposed in this thesis: the realisation of a lesion-directed ultrasound-guided biopsy to improve the detection of prostate cancer. For this reason the research activity described in this work was conducted side by side with the researchers of the department of Urology of S.Orsola hospital in Bologna.

As the investigated problem is the discrimination by a binary classification scheme of benignant tissue and a tissue diagnosed with adenocarcinoma, most of the experiments performed throughout the research activity were based on medical ground truth derived from biopsy findings. The classifier takes advantage of some characteristics of ultrasound signals, examined on medical ground truth, to type imaged tissues. Several aspects of this problem were investigated in this work: *i)* the ultrasound image analysis techniques to improve characterization performance, *ii)* the design of an appropriate learning scheme on medical ground truth to assure a good quality classification, *ii)* the modeling of ultrasound signal to design new features useful for tissue characterization, *iv)* the optimization of the developed techniques for real-time computation to allow their clinical employment.

The first part of this thesis concentrated on biomedical image analysis techniques involved in the tissue characterization procedure. Here it was proposed to include a pre-processing deconvolution step in the traditional procedure before the extraction of ul-

trasonic characteristics from the signal. This additional step was shown to improve the diagnostic skill of ultrasonic features, allowing a better discrimination between healthy and pathological tissues.

The second part of the work focused on ultrasound signal modeling. In particular it was proposed to use a sampled continuous domain autoregressive moving average model to fit the ultrasound data. This modeling provides the possibility to analyze the behaviour of the continuous-time signal which is than sampled in the acquisition process to provide the actual available data. In order to derive information about the continuous-domain signal, it was proposed a new mathematical formulation of CARMA process based on exponential B-splines. These functions, appropriately designed, allow to restore by interpolation the continuous time signal from the information about its samples. In addition, a novel maximum-likelihood CARMA identification algorithm based on exponential B-splines was proposed. This algorithm exploits an exact discretization of CARMA model and does not introduce any approximation or assumption on the sampling interval. It is, thus, able to provide good estimates of CARMA process parameters also in aliasing conditions and it can be useful in several image processing applications. Since in ultrasound applications the sampling frequency is quite high, there is no risk of aliasing and a simpler CARMA identification was also proposed for this case. Thanks to the presented techniques CARMA parameters can than be extracted from the ultrasound signal. A study performed on phantom images showed that these new continuous-time descriptors are more efficient than traditional discrete-domain ARMA parameters in characterizing the concentration of small particles in ultrasound images, called scatterers and responsible of ultrasound pulse diffraction.

Finally a clinical study was conducted on a large medical database to assess the feasibility of the improvement of the standard prostate cancer biopsy protocol by image processing techniques. The developed techniques were optimized for speed up and implemented on a GPU enabled with CUDA<sup>™</sup> architecture in order to allow the ultrasound image processing to run in parallel with the real-time visualization of B-mode echo. The result of this part of the research is the development of a real-time Computer-Aided Biopsy tool working at a frequency rate compatible with standard ultrasound machines and providing radiologists with a visual malignancy map, displayed on standard B-mode. This tool allows to

perform a lesion-detected biopsy with the same diagnostic power of the standard prostate biopsy protocol, but with a reduction in the number of cores to sample from 8-12 to 7.

Future research will be targeted to the improvement of the classification scheme on a larger medical database with the addition of textural features and to the installation of the rtCAB tool on a standard machine in order to start a clinical trial at S.Orsola hospital.

Another aspect to take care of is a faster implementation of the deconvolution algorithm and its assessment on a large medical database for a future inclusion of this pre-processing step in the rt-CAB processing.

In addition, further investigations should be performed in order to validate the usefulness of continuous-domain ARMA parameters on *in vivo* images to characterize biological tissues.

Finally, a delicate aspect of this research is the presence in the medical database of regions on which we cannot obtain full knowledge. Precancerous lesions, as well as regions with a very small tumour percentage, cannot be treated as labelled data. For this reason, and in order to capture as much information as possible from the medical ground truth, we are also investigating the properties of semi-supervised classification approaches to improve tissue characterization performances.



---

## FEATURES FOR ULTRASONIC TISSUE TYPING

---

The multi-feature approach used in this work, based on attributes of different nature, will be briefly described in the following paragraphs. In general, every single group of features produces several attributes, but, as described in section 2.1, a correlation analysis performed on each group provides a selection of attributes from a certain set. Once an output feature image is obtained, a statistical parameter is computed on feature values in each ROI to have a local information.

**Wavelet Transform:** this feature groups the wavelet packet coefficients of the RF signal, decomposed in four bands. According to the correlation analysis only the first and third bands coefficients are considered meaningful.

**Polynomial Fit of Wavelet Packet Transform:** the computation of this feature consists of the extraction of the four bands decomposition WT coefficients of the RF signal. The second step is a fitting of the computed WT coefficients for each point of RF signal with a third order polynomial [28]. The result is a projection of the WT coefficients in a lower dimensionality space. Only the second order coefficient is considered meaningful by correlation analysis.

**Wavelet Decomposition:** this feature consists on the decomposition of RF signal in its coherent and diffuse parts [44]. The RF signal is modelled as the sum of a diffuse component, due to the interaction of the pulse with resolvable scatterers, and a diffuse component due to the randomly located unresolvable scatterers. When a coherent component is present in time, the scale-averaged wavelet power is characterized by larger peaks, used to detect the time location of coherent scatterers through thresholding. The coherent component is then reconstructed according to the model, through the superposition of Gaussian modulated si-

nusoids. The diffuse component is derived from the difference of the RF signal and the coherent component. An inverse wavelet packet transform is applied to both coherent and diffuse signals and the coefficients in the 8<sup>th</sup> and 9<sup>th</sup> node of the packet wavelet tree are used as attributes. The generated feature set is then composed of four features but, according to correlation analysis, just the mean of the third parameter, *WDES Diff. Proj. n.8*, is meaningful for classification purposes. The name *WDES* indicates the used decomposition algorithm: wavelet-based despeckling [102].

**Central Frequency:** this feature is an estimate of the mean central frequency of the RF signal. According to the algorithm based on amplitude spectrum magnitude, the integrated backscatter coefficient is first evaluated as an average of the PSD of the RF signal, computed by means of FFT. Next, the mean central frequency is estimated as the first order momentum of RF signal PSD, i.e. it is computed as  $\frac{\sum_{BW} f * PSD(f)}{\sum_{BW} PSD(f)} = \frac{\sum_{BW} f * PSD(f)}{IBS}$ .

**Attenuation:** this group of features is based on the main idea that slope and intercept of the linear fitting of the RF signal central frequency are measures of signal attenuation and thus of backscatter [20]. The first step to compute this feature consists of RF signal PSD evaluation, which can be performed through several algorithms: zero crossing applied on a sliding window, FFT, fitting with an auto-regressive (AR) model of a certain order. In the last method the meaningful frequencies corresponding to non trivial maximums and minimums of the AR spectrum can be computed analytically since they only depend on AR coefficients. A last way to compute this attenuation features is based on modified PSD whose linear trending is removed and a smoothing procedure is applied to minimize the variance of spectral slope estimation [103]. After PSD estimation, the central frequency over the signal and its linear fitting are computed. The whole feature set contains 18 attributes but the correlation analysis select 13 of them, privileging computation based on AR model. In particular, in the ranked feature sets in Table 2.2, only the intercept of central frequency fitting in the case of 2<sup>nd</sup> order AR model estimated through Burg method (*Intercept\_AR(2) burg*) and the intercept of the second frequency extracted in the 3<sup>rd</sup> order AR model computed through a LMS procedure (*Intercept2\_AR(3) lmsd*) are selected.

**B-mode:** this single attribute represents the B-mode value of the analyzed RF signal, computed through its Hilbert Transform.

**Nakagami:** this group of features comes from a statistical model



of the RF signal based on the Nakagami distribution. Shape and scale parameters of Nakagami distributions estimated on the RF signal are shown to be correlated to the density and intensity of scatterers and can be computed through estimation of some RF signal statistical moments [26]. The attributes used in this work are the logarithm of the shape and scale parameters computed on the envelope of the RF signal (*Nakagami logm* and *Nakagami logOmega*). The same parameters can be extracted from the coherent and diffuse components of the RF signal. Correlation analysis selects shape and scale parameters when they are extracted from the RF signal, while only the scale parameter (*Naka w Diff. Proj. n. node\_number*, *Naka w Coher. Proj. n. node\_number*) is selected when these attributes are extracted from the RF signal diffuse and coherent component, thus this feature set contains four different attributes.

**Statistic:** both the mean and the variance of the pure RF signal in each ROI are considered by the correlation analysis as important attributes for diagnostic purposes. Thus these feature set contains two different attributes.

**Haralick:** this textural feature contains statistical attributes generated from co-occurrence matrix of the ROI in the direction of  $45^\circ$  [30]. Several different statistical parameters can be computed from the co-occurrence matrix in each direction, but the correlation analysis selects as meaningful features only four of them: sum of squares (*Haralick Sum of Squares*), correlation, entropy and average of the histogram of the sum of grey levels.

**Unser:** this feature contains statistical attributes generated from histogram of the sum and difference of grey levels in the ROI [31]. Nine statistical parameters can be computed from sum and difference histogram: mean, variance, contrast, homogeneity, cluster shade, cluster prominence, energy, correlation and entropy. These measures can be replicated on all the four angular directions ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ), giving a total of 36 attributes. Correlation analysis selects the following attributes: mean for the horizontal direction, correlation for the direction of  $45^\circ$ , mean and contrast for the vertical direction and correlation, cluster prominence, entropy and homogeneity for the direction of  $135^\circ$ , providing a 9 attributes feature set.

**Fractal:** this group of parameters is based on computation of progressive binarized versions of the B-mode image [32], [36] obtained through ten different binary thresholding operations. Lattice of different grid size highlights the number of squares contain-

ing active pixels, expressed through the lattice size and the fractal dimension. Finally, fractal features are computed through linear regression of the number of active squares with respect to the lattice size. Both slope (*Alpha*) and intercept (*Beta*) are used as features, providing 20 different fractal attributes, named *Fractal(1) Alpha/Beta num\_attribute*. A different kind of fractal feature is based on a sequence of imaginary extensions of original image toward upper and lower grey levels, considered as surfaces of a volume with a certain radius surrounding the original image [104], [36]. The surface area contains information about the difference of two subsequent extended images and it is expressed through fractal dimension and volume radius. Choosing ten different radius values and performing a piece-wise linear regression of the surface area, the derived slopes (*Alpha*) and intercepts (*Beta*) are used as textural features. This fractal feature provides 20 more different attributes, named *Fractal(2) Alpha/Beta num\_attribute*. Correlation analysis selected 7 attributes from the first group of fractal features and 9 from the second group, providing a set of 16 features.

---

## STATE OF THE ART IN TISSUE CHARACTERIZATION

---

The technical characteristics of previously published methods for prostate tissue characterization, discussed in chapter 2, are summarized in table B.1 in order to show a complete framework about the state of the art in ultrasound-based computer-aided detection techniques.

Table B.1: Published methods for ultrasound-based prostate tissue characterization

Work		Dataset							Techniques				Results			
Author	Year	Gold Standard	# images/ # patients	Malign/ Benign	# ROIs	Image size	ROI size	Train/Test	Segmentation	Features	Feature selection	Classifier	SE	SP	Acc	Az
Basset [16]	1993	biopsy	37/16	6/10 patients	37	512×512	1.45 cm <sup>2</sup>	-/-	rectangular around needle	Textural	no	one parameter decision tree	83	71	-	-
Huynen [17]	1994	biopsy	98/51	46/5 patients	-	512×512	3×3 mm <sup>2</sup>	66/32	rectangular around needle	Textural	no	multi-parameter decision tree	80	88.20	-	-
Houston [18]	1995	biopsy	25/25	11/14 patients	25	512×512	121×20	-/-	rectangular around needle	Textural	no	Decision tree	73	86	80	-
Schmitz [19]	1999	prostatectomy	200/33	33/0 patients	3400	-	128×16 0.1 cm <sup>2</sup>	32/1	-	acoustical and textural	SFS on Mahalanobis distance	-	82	88	-	-
Scheipers [20]	2003	prostatectomy	-/100	40517/129967 ROIs	170000	-	128×16 0.1 cm <sup>2</sup>	99/1	manual by radiologist	attenuation, scattering, textural	ROC area based wrapper	2 parallel neuro-fuzzy systems	-	-	75	86
Feleppa [21]	2004	biopsy and prostatectomy	-/200	110/909 ROIs	1019	-	64×64	180/20	-	spectral and clinical data	-	artificial neural network	-	-	80	85
Mohamed [22]	2005	physician visual inspection	33/20	20/0 patients	96	-	irregular	80/16	Gabor filter	Textural	mutual information	SVM	83.3	100	93.75	-
Llobet [53]	2007	biopsy	4593/289		4593	384 × 288	2500 pixels	202/87 patients	rectangular around needle	Textural	no	hidden Markov model	68	53	61.6	60.1
Mohamed [23]	2008	physician visual inspection	33/20	20/0 patients	108	-	irregular	90/18	Gabor filter	PSD, rotational invariance technique	Particle swarm optimization	SVM	83.3	100	94.4	-
Han [24]	2008	-	51/51	51/0 patients	200	350 × 300	25 × 25	5/46	thresholded histogram equalization	textural, location, ellipticity	no	SVM	92	95.9	-	-
HistoScanning <sup>TM</sup> [54]	2009	prostatectomy	29/29	29/0 patients	-	3D	0.08 cm <sup>2</sup>	15/14	-	statistical	-	integration of 3 algorithms	95	-	-	-

---

## INITIAL CONDITIONS IN CARMA ESTIMATION

---

The task of band-pass power spectrum allocation is repeatedly required by the algorithm of section 3.5. Each such allocation corresponds to a single local minima extraction and it is based on equating the derivative of the power spectrum  $\Phi(s; \theta)$  to zero while substituting  $s = j\omega_k$ , where  $\omega_k$  is the required frequency of maximum response. The coefficients of  $\Phi_L(s)$  are then obtained by numerical minimizing the derivative value at  $\omega_k$ . The initial value for the innovation variance is determined by the autocorrelation function of the model, which is the inverse Laplace transform of the power spectrum, and by the variance of the available data. A description of the allocation workflow is given in Figure C.1.

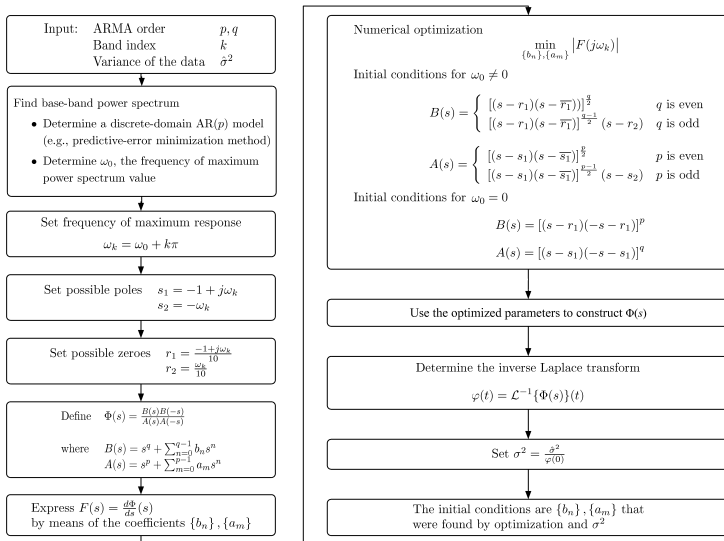


Figure C.1: Allocating initial conditions. Every local minima of the likelihood function (3.49) is obtained by a different set of initial conditions. This workflow is part of the estimation algorithm of Figure 3.5.

---

## NON SYMMETRIC EXPONENTIAL B-SPLINES

---

Considering a LTI operator  $L_{\mathbf{s},\mathbf{r}}$  representing a the whitening model of a CARMA process of orders  $(p, q)$  with poles  $\mathbf{s}$  and zeroes  $\mathbf{r}$ , an exponential spline ([105, 56, 57]) is defined as

$$s(t) = \sum_{k \in \mathbb{Z}} a[k] h(t - t_k; \mathbf{s}, \mathbf{r}) \quad (\text{D.1})$$

where  $h(t; \mathbf{s}, \mathbf{r})$  is the Green function of  $L_{\mathbf{s},\mathbf{r}}$ . The function  $s(t)$  can be reconstructed in exponential B-spline basis, more appropriate descriptor thanks to their compact support unlike the Green function:

$$s(t) = \sum_{k \in \mathbb{Z}} c[k] \beta(t - k; \mathbf{s}, \mathbf{r}) \quad (\text{D.2})$$

The exponential B-spline is a combination of shifted and scaled version of the Green function and can be defined by applying a localization operator  $\Delta_{\mathbf{s}}$  to the Green function itself:

$$\beta(t; \mathbf{s}, \mathbf{r}) = \Delta_{\mathbf{s}} \{h(t; \mathbf{s}, \mathbf{r})\} \quad (\text{D.3})$$

The localization operator is defined in the Fourier domain:

$$\Delta(\omega; \mathbf{s}) = \prod_{k=1}^p (1 - e^{-j\omega + s_k}) \quad (\text{D.4})$$

As the Green function is, by definition, the impulsive response of the inverse operator  $L_{\mathbf{s},\mathbf{r}}^{-1}$ , its Fourier transform is the CARMA system transfer function (3.9). Thus the Fourier transform of  $\beta(t; \mathbf{s}, \mathbf{r})$  is

$$\hat{\beta}(\omega; \mathbf{s}, \mathbf{r}) = \prod_{k=1}^p \frac{1 - e^{-j\omega + s_k}}{j\omega - s_k} \prod_{k=1}^q j\omega - r_k \quad (\text{D.5})$$

In this context the Green function  $h(t; \mathbf{s}, \mathbf{r})$  can be expressed as an exponential spline:

$$h(t; \mathbf{s}, \mathbf{r}) = \sum_{k \in \mathbb{Z}} p[k; \mathbf{s}] \beta(t - k; \mathbf{s}, \mathbf{r}) \quad (\text{D.6})$$

where the coefficients of the interpolation present the following z-transform:

$$P_d^-(z; \mathbf{s}) = \frac{1}{\prod_{k=1}^p (1 - e^{s_k} z^{-1})} \quad (\text{D.7})$$

The autocorrelation function  $\varphi(t; \mathbf{s}, \mathbf{r})$  of the CARMA process can be expressed in function of exponential B-splines:

$$\begin{aligned} \varphi(t; \mathbf{s}, \mathbf{r}) &= \sigma^2 h(t; \mathbf{s}, \mathbf{r}) * h(-t; \mathbf{s}, \mathbf{r}) \\ &= \sigma^2 \sum_{k \in \mathbb{Z}} p[k; \mathbf{s}] \beta(t - k; \mathbf{s}, \mathbf{r}) * \sum_{k' \in \mathbb{Z}} p[k'; \mathbf{s}] \beta(-t - k'; \mathbf{s}, \mathbf{r}) \\ &= \sigma^2 \sum_{k \in \mathbb{Z}} p[k; \mathbf{s}] \sum_{k' \in \mathbb{Z}} p[k'; \mathbf{s}] \beta(t - k; \mathbf{s}, \mathbf{r}) * \beta(-t - k'; \mathbf{s}, \mathbf{r}) \\ &= \sigma^2 \prod_{i=1}^p e^{s_i} \sum_{k \in \mathbb{Z}} p[k; \mathbf{s}] \sum_{k' \in \mathbb{Z}} p[k'; \mathbf{s}] \beta(t - k; \mathbf{s}, \mathbf{r}) * \beta(t + k' + p; -\mathbf{s}, -\mathbf{r}) \end{aligned}$$

The last passage is possible thanks to the symmetry property of exponential B-splines [56]. In addition the convolution of two exponential B-spline is still an exponential B-spline with as coefficients the full set of parameters of the first and the second splines. In this case the resultant exponential B-splines has parameters  $-\mathbf{s}, -\mathbf{r} : \mathbf{s}, \mathbf{r}$ , where the symbol  $:$  indicates the union of coefficients. This property leads to the following development:

$$\begin{aligned} \varphi(t; \mathbf{s}, \mathbf{r}) &= \sigma^2 \prod_{i=1}^p e^{s_i} \sum_{k \in \mathbb{Z}} p[k; \mathbf{s}] \sum_{k'' \in \mathbb{Z}} p[k - k'' - p; \mathbf{s}] \beta(t - k''; -\mathbf{s}, -\mathbf{r} : \mathbf{s}, \mathbf{r}) \\ &= \sigma^2 \sum_{k'' \in \mathbb{Z}} \tilde{p}[k'' + p; -\mathbf{s} : \mathbf{s}] \beta(t - k''; -\mathbf{s}, -\mathbf{r} : \mathbf{s}, \mathbf{r}) \end{aligned} \quad (\text{D.8})$$

The definition of the autocorrelation function of a CARMA process and its spectrum depends on the symmetric exponential B-splines  $\beta(t - k''; -\mathbf{s}, -\mathbf{r} : \mathbf{s}, \mathbf{r})$ , which was for this reason introduced in section 3.2. In particular, since considering all parameters  $(-\mathbf{s}, -\mathbf{r} : \mathbf{s}, \mathbf{r})$  is redundant, because they are opposite each other, in section 3.2 we take a single parameter vector  $\theta$  representing poles and zeroes of the system. Unlike (3.29), here  $\Phi_d(z, \theta) = \sigma^2 P_d(z; \theta) \cdot \mathcal{B}_d(z; \theta)$  uses some slightly different definitions of  $P_d(z; \theta)$



and  $\beta(t; \theta)$ . The definition of  $P_d(z; \theta)$  as  $z$ -transform of interpolation coefficients  $\tilde{p}[k]$  in (D.8) is as follows

$$P_d(z; \theta) = \frac{z^p \prod_{k=1}^p e^{s_k}}{\prod_{k=1}^p (1 - e^{s_k} z)(1 - e^{s_k} z^{-1})} \quad (\text{D.9})$$

As  $\beta(t; \theta) = \mathcal{F}^{-1} \{ \hat{\beta}(\omega; \mathbf{s}, \mathbf{r}) \cdot \hat{\beta}(\omega; -\mathbf{s}, -\mathbf{r}) \}$ , let us consider directly the definition of  $\hat{\beta}(\omega; \theta)$  in the frequency domain:

$$\begin{aligned} \hat{\beta}(\omega; \theta) &= \prod_{k=1}^q (j\omega - r_k)(j\omega + r_k) \cdot \prod_{k=1}^p \frac{(1 - e^{-j\omega + s_k})(1 - e^{-j\omega - s_k})}{(j\omega - s_k)(j\omega + s_k)} \\ &= (-1)^q \prod_{k=1}^p e^{-j\omega - s_k} \cdot \prod_{k=1}^q (j\omega - r_k)(-j\omega - r_k) \\ &\quad \cdot \prod_{k=1}^p \frac{(1 - e^{j\omega + s_k})(1 - e^{-j\omega + s_k})}{(j\omega - s_k)(-j\omega - s_k)} \end{aligned} \quad (\text{D.10})$$

These definitions allow to define the alternative expression for the variance of the innovation signal of the sampled CARMA process expressed in corollary 1.



---

# NOTATIONS

---

## Chapter One. Background

Symbol	Meaning
$\mathbf{x}$ :	generic ROI feature vector of length $N_f$
$x_i$ :	$i$ -th feature of a ROI
$N_f$ :	number of features
$x[n, m]$ :	2D image data
$w[n, m]$ :	2D tissue response
$\mathbf{H}(\cdot)$ :	acquisition system convolution operator
$h[n, m]$ :	acquisition system point spread function
$\mu[n, m]$ :	additive noise corrupting tissue response
$\mathbf{x}_k$ :	$k$ -th ROI feature vector of length $N_f$
$c_k$ :	$k$ -th ROI class label
$f(\mathbf{x})$ :	generic predictive model
$\mathbf{w}$ :	linear classifier coefficients vector
$b$ :	linear classifier trained threshold
$\alpha_i$ :	coefficients for support vector linear predictive model
$\phi(\cdot)$ :	non linear mapping for kernel-based classification
$\mathbf{K}(\cdot, \cdot)$ :	kernel function
$\mathbf{z}$ :	remapped ROI feature vector in kernel space
$sv$ :	set of reference feature vectors
$a_i$ :	$i$ -th coefficient for kernel-based classification rule
$\sigma_{rbf}^2$ :	variance of radial basis function kernel
$J$ :	Mahalanobis distance operator
$\mu_i$ :	intra-class mean value of samples belonging to class $i$
$\Sigma_i$ :	covariance matrix of samples belonging to class $i$
$TP$ :	number of well classified unhealthy (positive) samples
$FP$ :	number of bad classified healthy (negative) samples
$TN$ :	number of well classified healthy (negative) samples
$FN$ :	number of bad classified unhealthy (positive) samples
$ACC$ :	number of well classified samples over the total
$SE$ :	number of well classified unhealthy samples over all the unhealthy ones
$SP$ :	number of well classified healthy samples over all the healthy samples
$PPV$ :	n. of w. c. unhealthy samples over the samples classified as unhealthy
$N_0$ :	number of cores in standard prostate biopsy protocol
$ppv_0$ :	PPV of standard prostate biopsy protocol

---

## Chapter Two. Ultrasound image analysis and interpretation

Symbol	Meaning
$S$ :	ROI-based feature data matrix, feature set
$\mathbf{f}_i$ :	$i$ -th feature vector of length $N_{ROI}$
$N$ :	number of features
$N_{ROI}$ :	number of regions of interest
$N_{pixel}$ :	number of pixels
$\mathbf{c}$ :	data class vector of length $N_{ROI}$
$D(.,.)$ :	feature set relevance with respect to class
$R(.)$ :	feature set redundancy
$I(.,.)$ :	mutual information of feature set with respect to the class
$C$ :	K-means output segmented image, clustering of pixels
$K$ :	number of clusters for segmentation
$d(.,.)$ :	generic distance between pixels
$\mathbf{X}_i$ :	feature vector of the $i$ -th pixel of length $N_f$
$X_{i,j}$ :	$j$ -th feature value for the $i$ -th pixel
$\mathbf{X}$ :	pixel-based feature data matrix of size $N_{pixel} \times N_f$
$W(.)$ :	energy function of K-means algorithm
$y_k$ :	$k$ -th cluster center
$\Sigma$ :	final estimated segmented image
$P(\Sigma)$ :	<i>a priori</i> probability of segmented image
$P(\Sigma \mathbf{X})$ :	<i>a posteriori</i> probability of segmented image, given the observed data
$P(\mathbf{X} \Sigma)$ :	probability of the observed data conditioned to the segmented image
$C_i = k$ :	assignment of class $k$ to pixel $i$ by K-means
$\Sigma_i = k$ :	final assignment of class $k$ to pixel $i$
$\mu_{i,j}^k$ :	mean value of Gaussian distribution modelling $j$ -th feature value
$\sigma_{i,j}^k$ :	standard deviation of Gaussian distribution modelling $j$ -th feature value
$U(.)$ :	functional to minimize by segmentation algorithm
$V(.)$ :	Markov Random Field potential function
$\delta_i$ :	set of pixels in the neighbourhood of pixel $i$
$x[n]$ :	RF discrete signal
$w[n]$ :	true tissue response
$h[n]$ :	acquisition system point spread function
$\epsilon_k$ :	predictive deconvolution coefficients
$\hat{x}[n]$ :	estimated prediction of $x[n]$
$e[n]$ :	discrete innovation signal, estimated tissue response
$E(.,.)$ :	predictive deconvolution error function
$\epsilon$ :	predictive deconvolution coefficients vector
$\lambda$ :	predictive deconvolution forgetting factor
$F$ :	feature space
$\phi(.)$ :	non linear mapping
$\mathbf{x}$ :	generic feature vector of one ROI
$\mathbf{z}$ :	mapped ROI feature vector
$\mathbf{w}$ :	linear classifier coefficients vector
$b$ :	linear classifier trained threshold
$D_F(.)$ :	Fisher criterion
$S_B$ :	between-class scatter matrix for mapped data
$S_W$ :	within-class scatter matrix for mapped data
$\mathbf{K}(.,.)$ :	kernel function
<b>Appendix A. Features for ultrasonic tissue typing</b>	
$f$ :	frequency
$PSD(f)$ :	power spectral density of RF signal

### Chapter *Three*. A CARMA model for ultrasound signals

Symbol	Meaning
$f(x; y_1, y_2)$ :	continuous-time function $f$ is expressed in $x$ and depends on parameters $y_1$ and $y_2$
$f[n; y_1, y_2]$ :	discrete-time sequence $f$ is expressed in $n$ and depends on parameters $y_1$ and $y_2$
$\omega$ :	radial frequency [rad/time-unit] in Fourier transform or normalized frequency [rad/sample] in DTFT
$x[n]$ :	RF signal, discrete ARMA process
$\mathbf{x}$ :	discrete ARMA process samples vector
$c_k$ :	AR coefficients of $x[n]$
$d_k$ :	MA coefficients of $x[n]$
$x(t)$ :	continuous-time ARMA process, continuous-time RF signal
$w(t)$ :	continuous-time ARMA model innovation signal, CD tissue response
$D$ :	continuous-time derivative operator
$A(D)$ :	LTI operator (AR part)
$B(D)$ :	LTI operator (MA part)
$\sigma^2$ :	variance of continuous-time innovation signal
$p$ :	order of $A(D)$
$q$ :	order of $B(D)$
$\theta$ :	CARMA process parameters vector
$r_k$ :	zeros of CARMA model
$s_k$ :	poles of CARMA model
$\mathbf{r}$ :	vector of CARMA model zeroes
$\mathbf{s}$ :	vector of CARMA model poles
$\theta_0$ :	CARMA process real parameters vector
$a_k$ :	coefficients of $A(D)$
$b_k$ :	coefficients of $B(D)$
$\varphi(t)$ :	CARMA process autocorrelation function
$\varphi[n]$ :	sampled CARMA process autocorrelation sequence
$\Phi(j\omega)$ :	power spectrum of $x(t)$ ; Fourier transform of autocorrelation function
$\Phi(s)$ :	Laplace transform of $\varphi(t)$
$\Phi_d(e^{j\omega})$ :	power spectrum of $x[n]$ ; DTFT of sampled autocorrelation sequence
$\Phi_d(z)$ :	$z$ -transform of $\varphi[n]$
$h(t)$ :	CARMA system impulse response
$\hat{h}(\omega)$ :	CARMA system transfer function
$\alpha_k$ :	coefficients of partial fraction decomposition of $\Phi(s)$
$\beta(t)$ :	exponential B-spline
$\beta[n]$ :	samples of exponential B-spline
$\Delta(s)$ :	Laplace transform of localization filter
$p[n]$ :	$\phi(t)$ interpolation coefficients in exponential B-spline basis
$P_d(z)$ :	$z$ -transform of the sequence $p[n]$
$\mathcal{B}_d(z)$ :	$z$ -transform of the sequence $\beta[n]$
$\hat{\beta}_d(\omega)$ :	DTFT of the sequence $\beta[n]$
$\hat{\beta}(\omega)$ :	Fourier transform of $\beta(t)$
$H_d(z)$ :	ARMA model given by sampling of CARMA model
$\sigma_d^2$ :	discrete-time innovation signal variance in sampled CARMA model
$\rho_k$ :	poles of ARMA model
$\nu_k$ :	zeroes of ARMA model
$\mathcal{B}_d^-(z)$ :	causal part of $\mathcal{B}_d(z)$
$\mathcal{B}_d^+(z)$ :	anti-causal part of $\mathcal{B}_d(z)$
$P_d^-(z)$ :	causal part of $P_d(z)$
$P_d^+(z)$ :	anti-causal part of $P_d(z)$
$\eta(t)$ :	fundamental interpolation kernel of sequence $\varphi[n]$ to rebuild $\varphi(t)$
$I_{k,l}(\theta)$ :	element of Fisher information matrix
$T$ :	sampling interval
$N$ :	number of samples of discrete process
$pdf(\mathbf{x})$ :	probability density function of $\mathbf{x}$
$\Sigma$ :	autocorrelation matrix of $\mathbf{x}$

$ \Sigma $ :	determinant of $\Sigma$
$l(\theta; \mathbf{x})$ :	likelihood function of the model to data $\mathbf{x}$
$\tilde{l}(\theta; \mathbf{x})$ :	approximation of $l(\theta; \mathbf{x})$
$G_d(z)$ :	inverse of digital filter $H_d(z)$
$g_\theta[n]$ :	coefficients of digital filter $G_d(z)$
$\kappa(\theta)$ :	corrective term of log-likelihood approximation
$c[n]$ :	Fourier coefficients of logarithm of ARMA spectrum
$L[m, n]$ :	lower triangular matrix describing filter $g_\theta$
$T_M$ :	infinite Toeplitz matrix describing ARMA spectrum
$t_M[n]$ :	sequence associated to the matrix $T_M$
$\tilde{l}_W(\theta; \mathbf{x})$ :	Whittle approximation of log-likelihood function
$\tilde{\mathbf{x}}_d(\omega)$ :	DFT of $\mathbf{x}$
$e(a_k)$ :	estimation error over parameter $a_k$
$\hat{a}_{k,n}$ :	estimation of parameter $a_k$ in the $n$ -th experiment
$\omega_{peak}$ :	frequency at which $\Phi(\omega)$ is maximum
$\omega_{max}$ :	frequency at which $\Phi(\omega)$ is at $-10$ dB respect to $\Phi(\omega_{peak})$
$T_{max}$ :	sampling time corresponding to frequency $\omega_{max}$
<b>a</b> :	vector of coefficients $a_k$ of continuous-domain AR part
<b>b</b> :	vector of coefficients $b_k$ of continuous-domain MA part
<b>c</b> :	vector of coefficients $c_k$ of discrete-domain AR part
<b>d</b> :	vector of coefficients $d_k$ of discrete-domain MA part
$d_k(\mathbf{s}, \mathbf{r})$ :	coefficients of the casual polynomial of $\mathcal{B}_d(z)$
$\mathbf{d}(\mathbf{s}, \mathbf{r})$ :	vector of coefficients $d_k(\mathbf{s}, \mathbf{r})$
$J_{[x,y]}$ :	Mahalanobis distance between samples corresponding to scatterer concentration $x$ and $y$
$N_{classes}$ :	number of scatterers concentrations in the classification problem

---



---

#### Appendix B. Initial conditions in CARMA estimation

---

$k$ :	band index for multi band likelihood optimization
$\omega_0$ :	peak frequency for the base band solution of likelihood maximization
$\omega_k$ :	estimated peak frequency in the $k$ -th band
$A(s)$ :	casual part of denominator of $\Phi(s)$
$B(s)$ :	casual part of numerator of $\Phi(s)$
$F(s)$ :	first derivative of $\Phi(s)$
$\hat{\sigma}^2$ :	variance of data $x[n]$

---



---

#### Appendix C. Non symmetric exponential B-splines

---

$L_{\mathbf{s}, \mathbf{r}}$ :	LTI operator of whitening model of a CARMA process, inverse CARMA system
$s(t)$ :	exponential spline
$h(t)$ :	Green function of $L_{\mathbf{s}, \mathbf{r}}$ , impulse response of CARMA model ( $L_{\mathbf{s}, \mathbf{r}}^{-1}$ )
$a[k]$ :	coefficients of a generic spline interpolation by Green function of $L_{\mathbf{s}, \mathbf{r}}$
$c[k]$ :	coefficients of a generic spline interpolation by exponential B-splines
$\Delta_{\mathbf{s}}$ :	localization operator of green function in the time/space domain
$p[k]$ :	coefficients for green function interpolation by exponential B-splines
$\tilde{p}[k]$ :	coefficients for autocorrelation function interpolation by exponential B-splines

---



---

## Chapter *Four*. Improving current prostate biopsy protocol

Symbol	Meaning
$m$ :	Nakagami shape parameter
$\Omega$ :	Nakagami scale parameter
$\psi(m)$ :	Digamma function
$f(m)$ :	non linear function to determine $m$
$\hat{f}(m)$ :	approximation of $f(m)$
$p_k$ :	$\hat{f}(m)$ numerator coefficients
$q_k$ :	$\hat{f}(m)$ denominator coefficients
$C$ :	generic classification model
$\mathbf{X}_{trn}$ :	generic ROI feature matrix of size $N_{ROI}$ in training set $\times N_f$
$\mathbf{X}_{tst}$ :	generic ROI feature matrix of size $N_{ROI}$ in testing set $\times N_f$
$\mathbf{c}_{trn}$ :	generic training set class label vector
$\mathbf{c}_{tst}$ :	generic testing set class label vector
$\hat{\mathbf{X}}_{trn}$ :	selected $\mathbf{X}_{trn}$
$\hat{\mathbf{c}}_{trn}$ :	labels corresponding to the selected $\mathbf{X}_{trn}$
$\sigma_{rbf}$ :	variable variance of RBF kernel for GDA processing
$\hat{\sigma}_{rbf}$ :	selected RBF kernel variance
$d$ :	linear classification discriminant function
$th$ :	variable threshold to apply to discriminant function
$\hat{th}$ :	set of thresholds $th$ corresponding to maximum PPV
$\hat{\hat{th}}$ :	maximum threshold in the set $\hat{th}$
$ACC$ :	ROI-based accuracy
$TP$ :	core-based true positives
$FP$ :	core-based false positives
$PPV$ :	core-based positive predictive value
$ppv_0$ :	standard protocol PPV
$ppv_1$ :	proposed rtCAB PPV
$N_0$ :	standard protocol number of cores to sample
$N_1$ :	proposed rtCAB number of cores to sample

## Abbreviations

Acronym	Meaning
ACC:	accuracy
AR:	auto-regressive
ARMA:	auto-regressive moving-average
ASAP:	atypical small acinar proliferation
BPH:	benign hyperplasia
BW:	band width
CAR:	continuous-domain auto-regressive
CAD:	computer-aided diagnosis/detection
CARMA:	continuous-domain auto-regressive moving-average
CD:	continuous-domain
CRB:	Cramér-Rao bound
CT:	computerized tomography
CZ:	central zone
DFT:	discrete Fourier transform
DRE:	digital rectal examination
DTFT:	discrete-time Fourier transform
FFT:	fast Fourier transform
FLD:	fisher linear discriminant
FN:	false negatives
FP:	false positives
FS:	feature selection
GDA:	generalized discriminant analysis
GPU:	graphic processing unit
GUI:	graphical user interface
IQ:	in-phase quadrature signal
LDA:	linear discriminant analysis
LTI:	linear time invariant
MA:	moving-average
MAP:	maximum a posteriori
MIHFS:	mutual information hybrid feature selection
ML:	maximum likelihood
MLE:	maximum likelihood estimator
mRMR:	min-redundant max-relevance
MRF:	Markov random field
MRI:	Magnetic resonance imaging
nRMSE:	normalized root mean squared error
PCA:	principal component analysis
PIN:	prostatic intraepithelial neoplasia
PPV:	positive predictive value
PSA:	prostate-specific antigen
PSD:	power spectral density
PSF:	point spread function
PZ:	peripheral zone
RBF:	radial basis function
RF:	radio-frequency
RLS:	recursive least squares
ROI:	region of interest
rtCAB:	real-time computer-aided biopsy
SE:	sensitivity
SFS:	sequential forward selection
SP:	specificity
SVM:	support vector machine
TGC:	time-gain compensation
TN:	true negatives
TP:	true positives
TRUS:	trans-rectal ultrasound
TZ:	transition zone
US:	ultrasound



---

# PUBLICATIONS

---

- 2011 (*under revision*)  
IEEE Transactions on Signal Processing  
H. Kirshner, S. Maggio, M. Unser.  
*A Sampling Theory Approach for Continuous ARMA Identification*
- 2011 (*under revision*)  
IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control  
M. Alessandrini, S. Maggio, J. Porée, L. De Marchi, N. Speciale, E. Franceschini, O. Bernard, O. Basset.  
*An expectation maximization framework for improved tissue response characterization*
- 2011 May  
Proceedings SampTA2011  
H. Kirshner, S. Maggio, M. Unser.  
*Maximum-Likelihood Identification of Sampled Gaussian Processes*
- 2011 March  
Proceedings ISBI2011  
S. Maggio, M. Alessandrini, N. Speciale, O. Bernard, D. Vray, O. Basset, M. Unser.  
*Continuous-domain ARMA modeling for ultrasound tissue characterization*
- 2011 February  
SPIE Medical Imaging  
M. Alessandrini, S. Maggio, J. Porée, L. De Marchi, N. Speciale, E. Franceschini, O. Bernard, O. Basset.  
*An expectation maximization framework for an improved tissue characterization using ultrasounds*
- 2010 November  
17 Congresso Nazionale SIEUN  
N. Testoni, N. Speciale, A. Bertaccini, D. Marchiori, M. Fiorentino, F. Manfredi, R. Schiavina, R. Cividini, F. Galluzzo, S. Maggio, E. Biagi, L. Masotti, G. Masetti, G. Martorana.  
*A retrospective study to reduce prostate biopsy cores by a real time interactive tool*
- 2010 October  
Proceedings IEEE IUS2010  
N. Testoni, S. Maggio, F. Galluzzo, L. De Marchi, N. Speciale.  
*rtCAB: a tool for reducing unnecessary prostate biopsy cores*

- 2010 August  
Proceedings EUSIPCO 2010  
S. Maggio, H. Kirshner, M. Unser  
*Continuous-time AR model identification: does sampling rate really matter?*
- 2010 February  
IEEE Transactions on Medical Imaging  
S. Maggio, A. Palladini, L. De Marchi, M. Alessandrini, N. Speciale, G. Masetti  
*Predictive deconvolution and hybrid feature Selection for Computer-Aided Detection of prostate cancer*
- 2009 March  
Proceedings International Symposium on Acoustical Imaging  
M. Scebran, A. Palladini, S. Maggio, L. De Marchi, N. Speciale  
*Automatic regions of interests segmentation for computer aided classification of prostate TRUS images*
- 2008 November  
Proceedings IEEE IUS2008  
S. Maggio, L. De Marchi, M. Alessandrini, N. Speciale  
*Computer aided detection of prostate cancer based on GDA and predictive deconvolution*
- 2005 November  
WSEAS Transactions on Systems  
S. Maggio, N. Testoni, L. De Marchi, N. Speciale, G. Masetti  
*Ultrasound Images Enhancement by means of Deconvolution Algorithms in the Wavelet Domain*
- 2005 September  
WSEAS ISCGAV2005  
S. Maggio, N. Testoni, L. De Marchi, N. Speciale, G. Masetti  
*Wavelet-based Deconvolution Algorithms Applied to Ultrasound Images*

---

# BIBLIOGRAPHY

---

- [1] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 198–211, 2007.
- [2] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*. Elsevier Inc., 2004.
- [3] J. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Transaction on Medical Imaging*, vol. 25(8), p. 987:1010, 2006.
- [4] D. Pham, C. Xu, and J. Prince, "Current methods in medical image segmentation," *Annual Reviews in Biomedical Engineering*, vol. 2(1), pp. 315–337, 2000.
- [5] P. Boccacci and M. Bertero, *Introduction to Inverse Problems in Imaging*. Taylor & Francis, 1998.
- [6] O. Michailovich and D. Adam, *Blind Image Deconvolution: theory and applications*, ch. 5: Deconvolution of Medical Images from Microscopic to whole body images, pp. 169–230. CRC Press, 2007.
- [7] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.
- [8] A. R. Webb, *Statistical Pattern Recognition*. Wiley, 2002.
- [9] M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 821–837, 1964.
- [10] C. Abate-Shen and M. Shen, "Molecular genetics of prostate cancer," *Genes & Development*, vol. 14(19), p. 2410–2434, 2000.
- [11] J. McAninch and E. Tanagho, *Smith's General Urology*. McGraw-Hill/Appleton & Lange, 17th ed., 2008.
- [12] P. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Modern Pathology*, vol. 17(3), pp. 292–306, 2004.
- [13] J. Raja, N. Ramachandran, G. Munneke, and U. Patel, "Current status of transrectal ultrasound-guided prostate biopsy in the diagnosis of prostate cancer," *Clinical Radiology*, vol. 61, pp. 142–153, 2006.
- [14] J. De La Rosette, R. Giesen, A. Huynen, R. Aarmink, MP Van Iersel, F. Debruyne, and H. Wijkstra, "Automated analysis and interpretation of transrectal ultrasonography images in patients with prostatitis," *European urology*, vol. 27(1), pp. 47–53, 1995.
- [15] M. Essink-Bot, H. de Koning, H. Nijs, W. Kirkels, P. V. der Maas, and F. Schroder, "Short-term effects of population-based screening for prostate cancer on health-related quality of life," *Journal of The National Cancer Institute*, vol. 90, pp. 925–931, 1998.

- [16] O. Basset, Z. Sun, J. Mestas, and G. Gimenez, "Texture analysis of ultrasonic images of the prostate by means of co-occurrence matrix," *Ultrasonic Imaging*, vol. 15, pp. 218–237, 1993.
- [17] A. Huynen, R. Giesen, J. de la Rosette, R. Aarmink, F. Debruyne, and H. Wijkstra, "Analysis of ultrasonographic prostate images for the detection of prostatic carcinoma: the automated urologic diagnostic expert system," *Ultrasound in Medicine and Biology*, vol. 20, no. 1, pp. 1–10, 1994.
- [18] A. Houston, S. Premkumar, D. Pitts, and R. Babaian, "Prostate ultrasound image analysis: Localization of cancer lesions to assist biopsy," *Proceedings of the Eighth IEEE Symposium on Computer-Based Medical Systems*, pp. 94–101, 1995.
- [19] G. Schmitz, H. Ermert, and T. Senge, "Tissue-characterization of the prostate using radio frequency ultrasonic signals," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 46, pp. 126–138, 1999.
- [20] U. Scheipers, H. Ermert, H. Garcia-Schurmann, T. Senge, and S. Philippou, "Ultrasonic multifeature tissue characterization for prostate diagnostics," *Ultrasound in Medicine and Biology*, vol. 29, no. 8, pp. 1137–1149, 2003.
- [21] E. Feleppa, J. Ketterling, C. Porter, and J. Gillespie, "Ultrasonic tissue-type imaging (tti) for planning treatment of prostate cancer," *Proceedings of SPIE*, vol. 5373, no. 223, 2004.
- [22] S. Mohamed and M. Salama, "Computer-aided diagnosis for prostate cancer using support vector machine," *Proceedings of SPIE*, vol. 5744, no. 898, 2005.
- [23] S. Mohamed and M. Salama, "Prostate cancer spectral multifeature analysis using true images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 548–556, 2008.
- [24] S. Han, H. Lee, and J. Choi, "Computer-aided prostate cancer detection using texture features and clinical features in ultrasound image," *Journal of Digital Imaging*, vol. 21, no. Suppl 1, pp. S121–33, 2008.
- [25] M. Moradi, P. Mousavi, and P. Abolmaesumi, "Computer-aided diagnosis of prostate cancer with emphasis on ultrasound-based approaches: a review," *Ultrasound in Medicine and Biology*, vol. 33, no. 7, pp. 1010–1028, 2007.
- [26] P. Shankar, "Ultrasonic tissue characterization using a generalized Nakagami model," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 48, no. 6, pp. 1716–1720, 2001.
- [27] E. Feleppa, W. Fair, T. Liu, A. Kalisz, W. Gnadt, and F. Lizzi, "Two-dimensional and three-dimensional tissue-type imaging of the prostate based on ultrasonic spectrum analysis and neural network classification," *Proceedings of SPIE*, vol. 3982, no. 152, 2000.
- [28] L. Masotti, E. Biagi, S. Granchi, L. Breschi, E. Magrini, and F. D. Lorenzo, "Clinical test of rules (rules: radiofrequency ultrasonic local estimators)," *IEEE Ultrasonics Symposium*, vol. 3, no. 23–27 Aug., pp. 2173–2176, 2004.
- [29] G. Georgiou, F. Cohen, C. Piccoli, F. Forsberg, and B. Goldberg, "Tissue characterization using the continuous wavelet transform. part ii: Application on breast rf data," *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 48, no. 2, pp. 363–373, 2001.
- [30] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
- [31] M. Unser, "Sum and difference histograms for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 118–125, 1986.

- [32] W. Lee, Y. Chen, and K. Hsieh, "Ultrasonic liver tissues classification by fractal feature vector based on m-band wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 382–392, 2003.
- [33] I. Witten and E. Frank, *Tools Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, 2005.
- [34] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [35] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2008.
- [36] T. Wagner, *Handbook of Computer Vision and Applications, Volume 2: Signal Processing and Pattern Recognition*, ch. 12, Texture Analysis, pp. 275–308. Academic Press, 1999.
- [37] T. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 40(4), pp. 901–914, 1992.
- [38] J. Ng, R. Prager, N. Kingsbury, G. Treece, and A. Gee, "Wavelet restoration of medical pulse-echo ultrasound images in an em framework," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 54, no. 3, pp. 550–568, 2007.
- [39] O. Michailovich and D. Adam, "Phase unwrapping for 2-d blind deconvolution of ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 1–7, 2004.
- [40] H. Shin, R. Prager, J. Ng, H. Gomersall, N. Kingsbury, G. Treece, and A. Gee, "Sensitivity to point-spread function parameters in medical ultrasound image deconvolution," *Ultrasonics*, vol. 49, no. 3, pp. 344–357, 2009.
- [41] O. V. Michailovich and D. Adam, "A novel approach to the 2-d blind deconvolution problem in medical ultrasound," *IEEE Transactions on Medical Imaging*, vol. 24, no. 1, pp. 86–104, 2005.
- [42] K. Peacock and S. Treitel, "Predictive deconvolution: Theory and practice," *Geophysics*, vol. 34, no. 2, pp. 155–169, 1969.
- [43] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 4th ed., 2002.
- [44] G. Georgiou and F. Cohen, "Tissue characterization using the continuous wavelet transform, part i: Decomposition method," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 48, no. 2, pp. 355–363, 2001.
- [45] L. De Marchi, A. Palladini, N. Testoni, and N. Speciale, "Blurred ultrasonic images as isiaffected signals: Joint tissue response estimation and channel tracking in the proposed paradigm," *IEEE Ultrasonics Symposium*, pp. 1270–1273, 2007.
- [46] R. Neelamani, H. Choi, and R. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, 2004.
- [47] S. Caporale, A. Palladini, L. D. Marchi, N. Speciale, and G. Masetti, "Wavelet-based algorithms for speckle removal from b-mode images," *Proceedings LASTED International Conference on Biomedical Engineering*, vol. 417, 2004.
- [48] T. Joachims. *SVMlight: Support Vector Machine. SVM-Light Support Vector Machine* <http://svmlight.joachims.org/>, University of Dortmund, 1999.
- [49] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

- [50] M. Scabia, E. Biagi, and L. Masotti, "Hardware and software platform for real-time processing and visualization of echographic radiofrequency signals," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 49, no. 10, pp. 1444–1452, 2002.
- [51] F. Lizzi, E. Feleppa, S. K. Alam, and C. Deng, "Ultrasonic spectrum analysis for tissue evaluation," *Pattern Recognition Letters*, vol. 24(4-5), pp. 637–658, 2003.
- [52] L. Masotti, E. Biagi, S. Granchi, L. Breschi, E. Magrini, and F. D. Lorenzo, "Tissue differentiation based on radiofrequency echographic signal local spectral content," *IEEE Symposium on Ultrasonics*, vol. 1, pp. 1030–1033, 2003.
- [53] R. Llobet, J. Perez-Cortes, A. Toselli, and A. Juan, "Computer-aided detection of prostate cancer," *International Journal of Medical Informatics*, vol. 76, pp. 547–556, 2007.
- [54] J. Braeckman, P. Autier, C. Garbar, M. P. Marichal, C. Soviany, R. Nir, D. Nir, D. Michielsen, H. Bleiberg, L. Egevad, and M. Emberton, "Computer-aided ultrasonography (histoscaning): a novel technology for locating and characterizing prostate cancer," *BJU International*, vol. 101, pp. 293–298, 2007.
- [55] L. Ljung, *System Identification - Theory For the User, 2nd ed.* PTR Prentice Hall, Upper Saddle River, N.J., 1999.
- [56] M. Unser and T. Blu, "Cardinal exponential splines: Part i - theory and filtering algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 5373–5383, 2005.
- [57] M. Unser and T. Blu, "Cardinal exponential splines: Part ii - think analog, act digital," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 5384–5394, 2005.
- [58] I. Lim and B. Lee, "Lossless pole-zero modeling for speech signals," *IEEE Transactions on Speech and Audio Processing*, vol. 1(3), pp. 269–276, 1993.
- [59] H. Kirshner and M. Porat, "On the role of exponential spline in image interpolation," *IEEE Trans. Signal Processing*, vol. 18(10), pp. 2198–2208, 2009.
- [60] J. M. Francos, A. Z. Meiri, and B. Porat, "A unified texture model based on a 2-d wold like decomposition," *IEEE Trans. Signal Processing*, vol. 41, pp. 2665–2678, 1993.
- [61] S. Nemirovski and M. Porat, "On texture and image interpolation using markov models," *Signal Processing: Image Communication*, vol. 24, pp. 139–157, 2009.
- [62] K. J. Aström, *Introduction to Stochastic Control Theory*. Academic Press, 1970.
- [63] B. Wahlberg, "Limit results for sampled systems," *International Journal of Control*, vol. 78(3), pp. 1267–1283, 1988.
- [64] B. Wahlberg, L. Ljung, and T. Söderström, "Sampling of continuous time stochastic processes," *Control - Theory and Advanced Technology*, vol. 9(1), pp. 99–112, 1993.
- [65] T. Söderström, *Discrete-Time Stochastic Systems, 2nd ed.* London: Springer-Verlag, 2002.
- [66] M. M. E. K. Larsson and T. Söderström, "An overview of important practical aspects of continuous-time arma system identification," *Circuits Systems Signal Processing*, vol. 25, pp. 17–46, 2006.
- [67] H. Garnier and L. Wang, eds., *Identification of Continuous-time Models from Sampled Data*. London: Springer-Verlag, 2008.
- [68] R. Johansson, "Identification of continuous-time models," *IEEE Trans. Signal Processing*, vol. 42(4), pp. 887–897, 1994.
- [69] D. Pham, "Estimation of continuous-time autoregressive model from finely sampled data," *IEEE trans. Signal Processing*, vol. 48(9), pp. 2576–2584, 2000.

- [70] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: Implementation issues and performance evaluation," *Int. J. Control*, vol. 76(13), pp. 1337–1357, 2003.
- [71] H. Garnier and P. Young, "Time-domain approaches to continuous-time model identification of dynamical systems from sampled data," in *Proc. Amer. Control Conf., Boston, MA*, 2004.
- [72] G. C. Goodwin, J. Yuz, and H. Garnier, "Robustness issues in continuous-time system identification from sampled data," in *Proc. 16th IFAC World Congress, Prague, Czech Republic*, 2005.
- [73] H. Fan, T. Söderström, M. Mossberg, B. Carlsson, and Y. Zou, "Estimation of continuous-time ar process parameters from discrete-time data," *IEEE Trans. Signal Processing*, vol. 47(5), pp. 1232–1244, 1999.
- [74] M. Mossberg, "Estimation of continuous-time stochastic signals from sample covariances," *IEEE trans. Signal Processing*, vol. 56(2), pp. 821–825, 2008.
- [75] R. Pintelon, J. Schoukens, and Y. Rolain, *Frequency-domain approach to continuous-time system identification: Some practical aspects*, ch. 8, pp. 214–248. London: Springer-Verlag, 2008.
- [76] J. Gillberg and L. Ljung, "Frequency-domain identification of continuous-time arma models from sampled data," *Automatica*, vol. 45, pp. 1371–1378, 2009.
- [77] T. Söderström, "Computing stochastic continuous-time models from arma models," *International Journal of Control*, vol. 53(6), pp. 1311–1326, 1991.
- [78] H. Kirshner, M. Porat, and M. Unser, "A stochastic minimum-norm approach to image and texture interpolation," in *European Signal Processing Conference, Denmark*, 2010.
- [79] E. K. Larsson, "Limiting sampling results for continuous-time arma systems," *International Journal of Control*, vol. 78(7), pp. 461–473, 2005.
- [80] B. Friedlander, "On the computation of the cramer-rao bound for arma parameter estimation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-32(4), pp. 721–727, 1984.
- [81] E. K. Larsson and E. G. Larsson, "The crb for parameter estimation in irregularly sampled continuous-time arma system," *IEEE Signal Processing Letters*, vol. 11(2), pp. 197–200, 2004.
- [82] G. Szegő, "On certain hermitian forms associated with the Fourier series of a positive function," *Communications du Séminaire Mathématique de l'Université de Lund*, vol. 9, pp. 228–238, 1952.
- [83] R. E. Hartwig and M. E. Fisher, "Asymptotic behavior of toeplitz matrices and determinants," *Archive for Rational Mechanics and Analysis*, vol. 32(3), pp. 190–225, 1969.
- [84] E. Basor, "Asymptotic formulas for toeplitz determinants," *Transactions of the American Mathematical Society*, vol. 239, p. 333–365, 1978.
- [85] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2(3), pp. 155–239, 2006.
- [86] S. Maggio, H. Kirshner, and M. Unser, "Continuous-time ar model identification: Does sampling rate really matter?," in *European Signal Processing Conference, Denmark*, 2010.
- [87] K. J. Aström and T. Söderström, "Uniqueness of the maximum likelihood estimates of the parameters of an arma model," *IEEE Trans. on Automatic Control*, vol. AC-19(6), pp. 769–773, 1974.

- [88] E. K. Larsson, M. Mossberg, and T. Söderström, "An overview of important practical aspects of continuous-time arma system identification," *Circuits Systems Signal Processing*, vol. 25(1), pp. 17–46, 2006.
- [89] H. Tsai and K. S. Chan, "Maximum likelihood estimation of linear continuous time long memory processes with discrete time data," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 67(5), pp. 703–716, 2005.
- [90] "Maximum likelihood estimation of a class of continuous-time long-memory processes," tech. rep., Department of Statistics and Actuarial Science, University of Iowa, Iowa City. Tech. Rep. 324, available at <http://www.stat.uiowa.edu/techrep/>, 2005.
- [91] R. H. Jones and A. V. Vecchia, "Fitting continuous arma models to unequally spaced spatial data," *Journal of the American Statistical Association*, vol. 88(423), pp. 947–954, 1993.
- [92] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. New Jersey: Prentice Hall, 1997.
- [93] S. Maggio, L. D. Marchi, M. Alessandrini, and N. Speciale, "Computer aided detection of prostate cancer based on gda and predictive deconvolution," *Proceedings IEEE IUS*, 2008.
- [94] A. Abdulsadda, N. Bouaynaya, and K. Iqbal, "Stability analysis and breast tumor classification from 2d arma models of ultrasound images," in *Proceedings of the Thirty-First Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBSÅ09), Minneapolis MN, USA, 2009*.
- [95] J. Zielinski, N. Bouaynaya, and D. Schonfeld, "Two-dimensional arma modeling for breast cancer detection and classification," in *Proceedings of the 2010 IEEE International Conference on Signal Processing and Communications (SPCOM2010), Bangalore, India, July 18-21, 2010*.
- [96] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 2007.
- [97] C. Fritsch, A. Ibanez, and M. Parrilla, "A digital envelope detection filter for real-time operation," vol. 48, pp. 1287–1293, Dec. 1999.
- [98] J. H. Chang, J. T. Yen, and K. K. Shung, "High-speed digital scan converter for high-frequency ultrasound sector scanners," *Ultrasonics*, vol. 48, pp. 444–452, September 2008.
- [99] M. Garland, S. Le Grand, J. Nickolls, J. Anderson, J. Hardwick, S. Morton, E. Phillips, Y. Zhang, and V. Volkov, "Parallel computing experiences with cuda," vol. 28, pp. 13–27, jul. 2008.
- [100] J. Cheng and N. Beaulieu, "Maximum-likelihood based estimation of the nakagami m parameter," vol. 5, pp. 101–103, mar. 2001.
- [101] G. A. Baker and P. Graves-Morris, "Padé approximants second edition," in *Encyclopedia of Mathematics and its Applications*, no. No. 59, p. 746, Cambridge University Press, 1996.
- [102] F. Argenti and L. Alparone, "Speckle removal from SAR images in the undecimated wavelet domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2363–2374, 2002.
- [103] D. Liu and M. Saito, "A new method for estimating the acoustic attenuation coefficient of tissue from reflected ultrasonic signals," *IEEE Transactions on Medical Imaging*, vol. 8, no. 1, pp. 107–110, 1989.
- [104] S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple resolution texture analysis and classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 68, no. 4, pp. 518–523, 1984.
- [105] I. Khalidov and M. Unser, "From differential equations to the constructio of new wavelet-like bases," *IEEE Transactions on Signal processing*, vol. 54, no. 4, 2006.