PARAMETRIC IMAGE ESTIMATION IN ANALYZER-BASED PHASE
CONTRAST IMAGING

BY
ORIOL CAUDEVILLA

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering in the Graduate College of the Illinois Institute of Technology

Approved: [Signature]
Adviser

Chicago, Illinois
July 2014
ACKNOWLEDGEMENT

I do not want to miss the opportunity of thanking to my Thesis Adviser, Dr. Brankov, for his priceless guidance and support in my first steps in medical imaging research. Without him I would had never had the chance of working in Medical Imaging. Also, I would like to say thank you to my coworkers, who have collaborated on creating this fantastic working environment where I felt like home: any good job I did was partly because of them.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td>viii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Phase contrast imaging fundamentals</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Analyzer-Based Imaging system</td>
<td>6</td>
</tr>
<tr>
<td>1.3. Introduction to the reconstruction problem</td>
<td>10</td>
</tr>
<tr>
<td>2. RECONSTRUCTION METHODS</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Physical and statistical model of MIR</td>
<td>12</td>
</tr>
<tr>
<td>2.2 Classical Estimator (CE)</td>
<td>15</td>
</tr>
<tr>
<td>2.3 Maximum Likelihood (ML)</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Least squares estimator (LS)</td>
<td>17</td>
</tr>
<tr>
<td>2.5 Regularization: Maximum a Posteriori</td>
<td>18</td>
</tr>
<tr>
<td>3. IMPLEMENTATION: L-CONJUGATE ALGORITHM</td>
<td>20</td>
</tr>
<tr>
<td>3.1 Introduction to conjugate gradient (Quadratic case)</td>
<td>20</td>
</tr>
<tr>
<td>3.2 Classic nonquadratic extension (Polak-Ribière)</td>
<td>21</td>
</tr>
<tr>
<td>3.3 L-Conjugate algorithm</td>
<td>22</td>
</tr>
<tr>
<td>3.4 Extension for MAP</td>
<td>30</td>
</tr>
<tr>
<td>4. RESULTS, CONCLUSIONS AND FUTURE WORK</td>
<td>32</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>48</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Absorption estimation performance of ML LS and CE</td>
<td>37</td>
</tr>
<tr>
<td>2. Referaction estimation performance of ML LS and CE</td>
<td>37</td>
</tr>
<tr>
<td>3. USAXS estimation performance of ML LS and CE</td>
<td>37</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>2</td>
</tr>
<tr>
<td>2.</td>
<td>3</td>
</tr>
<tr>
<td>3.</td>
<td>4</td>
</tr>
<tr>
<td>4.</td>
<td>4</td>
</tr>
<tr>
<td>5.</td>
<td>5</td>
</tr>
<tr>
<td>6.</td>
<td>7</td>
</tr>
<tr>
<td>7.</td>
<td>10</td>
</tr>
<tr>
<td>8.</td>
<td>15</td>
</tr>
<tr>
<td>9.</td>
<td>29</td>
</tr>
<tr>
<td>10.</td>
<td>33</td>
</tr>
<tr>
<td>11.</td>
<td>38</td>
</tr>
<tr>
<td>12.</td>
<td>39</td>
</tr>
<tr>
<td>13.</td>
<td>39</td>
</tr>
<tr>
<td>14.</td>
<td>40</td>
</tr>
<tr>
<td>15.</td>
<td>41</td>
</tr>
<tr>
<td>16.</td>
<td>42</td>
</tr>
<tr>
<td>17.</td>
<td>43</td>
</tr>
</tbody>
</table>
18. PR vs LC convergence $\beta = 450$ ........................................................................... 43
19. PR vs LC convergence for $\beta = 800$ ........................................................................ 44
20. PR vs LC convergence for $\beta = 2000$ ........................................................................ 44
21. LC vs PR refraction reconstructions vs number of iterations ($\beta = 450$) ............ 45
22. PR vs LC comparison ($\beta = 450$) ........................................................................... 45
LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Absorption MIR parameter</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Refraction MIR parameter</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Ultra Small Angle Scatter MIR parameter</td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>Position vector</td>
</tr>
<tr>
<td>$x$</td>
<td>Vector of MIR parameters</td>
</tr>
<tr>
<td>$R$</td>
<td>Rocking curve of the system</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Angle of acquisition</td>
</tr>
<tr>
<td>$f$</td>
<td>Object function</td>
</tr>
<tr>
<td>$y_I$</td>
<td>Acquired angular samples</td>
</tr>
<tr>
<td>$L$</td>
<td>Log-Likelihood function</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Regularization parameter</td>
</tr>
<tr>
<td>$I(x)$</td>
<td>Fisher information matrix of the parameters</td>
</tr>
<tr>
<td>$D_g$</td>
<td>Differential matrix of $g$</td>
</tr>
<tr>
<td>$H_g$</td>
<td>Hessian matrix of $g$</td>
</tr>
<tr>
<td>$h$</td>
<td>Gradient of the Log-Likelihood function</td>
</tr>
<tr>
<td>$H_L$</td>
<td>Hessian matrix of the Log-Likelihood</td>
</tr>
</tbody>
</table>
ABSTRACT

Since the discovery of X-rays, by W.E. Roentgen in 1895, absorption-contrast imaging has been widely used in medical diagnostic, for example: mammography, chest X-ray, angiography etc.

Over the last 25 years many researchers have shown that phase-contrast X-ray imaging is capable of obtaining a better contrast then classical absorption-contrast imaging. This is especially true in soft tissue imaging like mammography or at higher X-rays energies.

During the past five years a significant effort has made in development of a tabletop analyzer-based phase-contrast system. In this system a perfect crystal is used to generate a quasi-monochromatic beam, which after interacting with an object is analyzed by a second perfect crystal called analyzer. The need of quasi-monochromatic beam and the low brilliance of conventional X-ray sources (CXS) has been the major limiting factor for tabletop systems.

The work presents in this thesis aims to develop novel reconstruction methods for tabletop analyzer-based phase-contrast imaging (ABI) systems. The presented reconstruction methods goal is to minimize the impact of CXS low brilliance by utilizing phase-contrast image formation model and maximum or maximum-a-posteriori Poisson likelihood approach. Finally, a fast-convergent conjugate gradient optimization algorithm has been derived specifically for Poisson likelihood function maximization.
CHAPTER 1
INTRODUCTION

This chapter introduces the general concepts of phase-contrast imaging and the particular Analyzer-based method. The goal is introducing the physical phenomena that phase-contrast techniques image and the systems used for in this process.

1.1. Phase contrast imaging fundamentals

Traditional X-ray imaging systems contrast is based on a difference in the X-ray photon absorption by different material (or tissue) types. This is a very common technique used for diverse applications such as airport security, industrial X-ray imaging and diagnostic medical imaging.

Photons in an X-ray beam can be absorbed by matter transferring its energy to the electrons which are expelled leaving its atoms ionized, Figure 2. The photons absorption probability depends on the density of the matter. Dense objects like bones absorb much more photons than light matter like air. This property makes possible, for example, to different soft and hard tissue in medical X-ray imaging, see Figure 1 for example.
Absorption is an intensity phenomena, it reduces the energy of the incident beam. However, measuring the intensity of the attenuated beam cannot properly image objects with low absorption contrast such as different types of soft tissue. Different structures in soft tissue have similar absorption properties making them indistinguishable in classical radiography. Nevertheless, subtle differences in the soft tissue density induce phase distortions of the X-rays. In addition, since phase-contrast does not depend on absorption, it can be observed at higher photon energies. This is important because can potentially reduce the harmful radiation dose.
Phase-contrast imaging techniques measure the change of the photon direction. Different areas of the imaged samples have a different index of refraction in consequence photons are refracted in different directions (Figure 3).

Finally ultra-small angle X-ray scatter (USAXS) measures the incoming beam dispersion due to refractive interaction with structures that are smaller than the detected pixels (a.k.a sub-pixel structure) (Figure 4).

In summary Figure 2, Figure 3 and Figure 4 schematically describe the three parameters that phase-contrast systems can measure and the corresponding resulting parametric images that we usually denote by the symbols $[\alpha, \mu, \sigma^2]^T$.  

Figure 2. Absorption process and its image.
Figure 3. Refraction process and its image.

Figure 4. USAXS process and its image.

Note that all of the incoming photon beam must travel in the same forward direction in order to observe described effects. This beam property is measured by beam spatial coherence. In addition since the refraction depends on the photon frequency (as described by the Snell’s law), the incoming beam must also be quasi-monochromatic (its frequency spectrum must be very narrow), this beam property is measured by beam temporal coherence. So a parallel and monochromatic beam is spatial and temporal coherent and it is usually called *coherent light*. 
As an example, Figure 5 shows the results that can be achieved with phase contrast imaging of soft tissue. Absorption image (left, a), shows good contrast between soft and hard tissue, bones are perfectly visible. The central image (b) corresponds to the refraction image and it shows good contrast for soft tissue like the trachea (arrow). Finally, USAXS (right, c) shows areas with high textural content, like lungs (arrow). Figure 5 is one of the most recent achievements in phase-contrast imaging: the first in-vivo image acquired of a small animal [2]. Those results were obtained using the so-called Grating-based Imaging (GBI) which uses an interferometry technique to measure phase effects. GBI is one of the most important methods in phase-contrast jointly with Analyzer-based Imaging (ABI), which is the technique used in this dissertation.

![Figure 5. Phase-contrast images obtained with Interferometry method: (a) absorption, (b) refraction, (c) ultra-small angle scattering. Source: [2].](image)

As it will be explained in section 1.2, ABI require more monochromatic light than GBI. This is an advantage for GBI systems because monochromating the beam causes
noisy reconstructions due to the loose of photons. However, GBI must be used with low-
energy photons, which means that bigger samples (like human bodies) could absorb most 
of the radiation leaving a few photons to reconstruct refraction and USAXS images. As it 
will be shown, the number of photons that reach the detector is the limiting parameter in 
phase-contrast imaging. Generally, low photon counts cause noisy reconstructions due to 
the Poisson noise. A complete explanation of GBI can be found in Section II chapter 3 in 
[3], and a comparison with ABI was developed in [4].

1.2. Analyzer-Based Imaging system

Analyzer-based phase contrast imaging (ABI) is a technique that measures the 
absorption, refraction and USAXS effects by using a perfect crystal called analyzer. The 
analyzer is an angular filter, which is used to measure the number of photons traveling at 
specific propagation angle with respect the horizontal plane.

Before describing the system, a more formal definition of absorption, refraction 
and USAXS will be given. To do so, first we must introduce the so-called Angular 
Intensity Profile (AIP). An ideal coherent beam contains X-ray contains photons traveling 
in a single direction and all have same energy (frequency). However, a real beam 
contains photons traveling in multiple directions around the main propagation vector. The 
number (or density) of photons traveling at each direction is described by Angular 
Intensity Profile (AIP) curve. Figure 6 shows two examples of AIP: horizontal axis 
represents the propagation angle with respect to the horizontal plane of the imaging 
system and the vertical axis is the number of photons at each angle.
As it was shown in [5] and [6], AIP distortions can be approximately described by three effects: absorption, refraction and ultra-small angle scatter (USAXS). Those three effects correspond to the general AIP attenuation (lose of photons caused by their absorption), AIP shifting (beam deflection induced for the changes in the object density) and AIP dispersion (angular divergence of the beam caused by textural elements in the object). Both the refraction and USAXS effects are due to phase changes in the incident X-ray beam when it goes through an object under imaging. In Figure 6 the three effects are exemplified. The AIP after the object has a lower integral, which means that some photons are absorbed. Next, it has been refracted by 2µrad measured as an AIP shift. Also, the curve is wider, which is caused by the USAXS effect.

Figure 6. AIP distortions. Left: AIP without sample. Right: AIP after an object.

Phase effects on the AIP are usually in the order of µrad and depend on the frequency of the X-ray.
An analyzers-based phase contrast system is schematically shown in Figure 7. This system is also known as Bonse-Hart camera (named after its inventors) [7]. The next paragraphs describe the four steps for acquiring the raw data.

First, since the beam generated with the conventional X-ray tube have a wide frequency spectrum it needs to be monochromatized and collimated. This us achieved by using a monochromator constituting of two perfectly crystal aligned to the Bragg’s angle. Second, in order to measure the resulting transmitted AIP, the beam is reflected by the analyzer crystal also aligned to the Bragg’s angle. Finally image is captured by a photon detector.

Usually, the monochromator and analyzer use a perfect crystal of Germanium or Silicon. As it was described by William Lawrence Bragg in [8], perfect crystals show a very high reflectivity for a very thin range of angles around a resonance point (named Bragg angle after his discoverer) and high absorption elsewhere. This phenomenon is also known as Bragg’s law and it is due to the constructive interference of the scatter radiation caused by the regular atomic structure at a certain incidence angle.

By setting up the analyzer position at different angles the detector captures the radiation intensity at different propagation directions, which is equivalent to sampling the AIP of the transmitted beam. This process is known as ‘rocking’ the analyzer. The analyzer is rocked vertically only, as it can be appreciated in Figure 7. Because of that, ABI systems are phase-sensitive in the vertical axis: they measure vertical refraction and USAXS. Since this process obtains multiple images of the object and different propagation directions, some times the ABI methods are called Multiple-Image Radiography (MIR). The three parametric images describing object phase effects were
introduced in [5, 9]. MIR can be considered as an extension of Diffraction Enhanced Imaging (DEI) [10]. DEI is an ABI method that samples the AIP at two specific analyzer angular positions corresponding to the Full Width at Half Maximum (FWHM) of the incident AIP.

Usually, the AIP of the incident beam is known as rocking curve because it can be measured by rocking the analyzer without setting up any object to image. The rocking curve is an intrinsic AIP of the system and it depends on the x-ray source, monochromatic, analyzer, etc.

As mentioned before, a real ABI system for medical purposes must use a regular X-ray tube source. Regular sources used in medical imaging give a very good results for conventional radiography: they provide enough photons to obtain a reasonable Signal to Noise Ratio (SNR). However, when setting up a ABI system with a regular source, the amount of photons that finally hit the detector is much lower due to the need of coherent light. Conventional radiography systems can use a wide spectrum for measuring absorption properties but, in ABI systems, the monochromator gets rid of all photons with a different energy from the desired. Also, the need of spatial coherence requires the use of a collimating device which absorb radiation too. Therefore, applying phase-contrast to medical imaging is usually limited by the SNR of the detected samples. However, this is not a problem in synchrotron light sources because of its high brilliance (outcoming flux of photons). During the past years, significant progress has been done in obtaining a table top ABI system (and other phase-contrast technologies) by pushing the hardware performance and the reconstruction accuracy.
The dissertation presented here is a step forward in the image reconstruction algorithm that allows to obtain better and faster MIR parameters estimations.

![Figure 7. ABI System of acquisition.](image)

1.3. **Introduction to the reconstruction problem**

In section 1.2 the data acquisition system has been introduced. As explained there, an Analyzer-Based Imaging (ABI) system measures samples of the Angular Intensity Profile (AIP) using the analyzer crystal.

The reconstruction problem is the estimation of the three phase effects introduced in section 1.2, according to the propagation model developed in [6] that will be extensively explained in 2.1.

As it was also introduced in section 1.2, the lack of photons is a problem in tabletop ABI systems. In all photon-counting systems, like in several medical imaging techniques, the discrete nature of photons leads to Poisson noise in the acquired data. Poisson noise is critical in low light environments (with low photons) where the Signal to Noise Ratio (SNR) is worse. All those factors make a real medical applicable ABI system
potentially noisy. According to that fact, a reconstruction method tries to make the best out of the poor quality data acquired.

Reconstruction methods for ABI systems are the central topic of this work. In the following chapters, different options will be discussed and evaluated according to two basic criteria: maximum estimation accuracy and minimum processing time. Chapter 2 introduces the different statistics developed and their performance evaluation. Maximum Likelihood Estimator (MLE) will be proved to be the best approach for ABI parameter estimation. In consequence, efforts in pushing the MLE estimation speed have been done. From that work, chapter 3 shows the good results obtained using a new optimization method for Poisson Likelihood Maximization that we call L-conjugate algorithm.
CHAPTER 2

RECONSTRUCTION METHODS

In this chapter, we introduce the reconstruction methods that are called to substitute the original estimation algorithm proposed in [5, 9] and the Least Squares (LS) method developed by [11]. Those previously proposed estimators do not consider the Poisson nature of the noise in the samples. The natural step forward is using a Maximum Likelihood approach (ML) to improve the estimations in terms of mean Variance and mean squared Bias. ML indeed considers the Poisson nature of the noise.

2.1 Physical and statistical model of MIR

Section 1.2 introduces the general principles behind Analyzer-Based Phase Contrast Imaging. As it has been described, we attempt to detect how tissue interacts with X-ray radiation not only in terms of absorption but in terms of phase. ABI methods evaluate phase effects by measuring three kind of distortions on the known Angular Intensity Profile (AIP) of the incoming beam: absorption ($\alpha$), refraction ($\mu$) and ultra-small angle scattering ($\sigma^2$).

The physical model of Multiple-Image Radiography was first proposed by [5], and then validated in [6]. Those publications prove that the three parameters already introduced define an object function $f$ that accurately describes the possible distortions in the AIP of an incoming beam, also called rocking curve $R$. The model describes the interaction at pixel level as the convolution of the object function and the rocking curve.
where $g$ is the AIP of the transmitted beam. For a more visual description of the image model formation, see figure 8.

$$g(\theta) = R(\theta) * f(\theta)$$  \hspace{1cm} (2.1)

In order to derive the object function, in [6] an approximation is obtained using radiative transfer theory [12]. First, the sample is considered invariant across the plane orthogonal to the beam path $z$. The medium is characterized by its absorption coefficient, $\mu_m(\vec{r}) = \mu_m(z)$, and its refractive index, $n(\vec{r}) = n_0 + n_y(z)$ where $\vec{r} = (x, y, z)$. Then, a scatter elements distribution in the medium is considered, $\rho_n(\vec{r}) = \rho_n(z)$. The scatters are characterized their cross section $\sigma_s$, their absorption cross section $\sigma_a$ and the difference between the refractive index of the scatters and the medium, $\delta$. Those parameters define the so-called albedo (2.2) and the beam broadening due to a single scattering center (or scatter) (2.3).

$$W_0 = \frac{\sigma_s}{\sigma_s + \sigma_a}$$ \hspace{1cm} (2.2)

$$\alpha_p = \left(4\delta^2 \ln\left(\frac{2}{\delta}+1\right)\right)^{-1}$$ \hspace{1cm} (2.3)

The assumptions described above define a eight parameters model for the object function (2.4) where $\mu(z) = \rho_n(z)\sigma_a + \mu_m(z)$, $Z$ is the thickness of the sample at a given pixel position, $\Delta \theta = \int_0^Z n_y(z)/n_0 \, dz$ and $\tau(Z) = \int_0^Z (\rho_n(z)\sigma_{ext} + \mu(z)) \, dz$.

$$f(\theta) = \exp\left(\int_0^\Delta \mu(\vec{r}) \right) \sum_{k=0}^{\infty} \left[ \frac{\left(\tau(Z)W_0\right)^k}{k!} \exp(-\tau(Z)W_0) \sqrt{\frac{\alpha_p}{k\pi}} \exp\left(-\frac{\alpha_p}{k}(\theta - \Delta \theta)^2\right) \right]$$ \hspace{1cm} (2.4)
As it was proved in [13], the object function of any pixel can be approximated as an attenuated Gaussian function (2.5) defining the new three parameters $[\alpha, \mu, \sigma^2]^T$ as shown in (2.6), (2.7) and (2.8). Those three parameters correspond to the absorption, refraction and USAXS effects defined in section 1.1.

$$f(\theta; \alpha, \mu, \sigma^2) = \frac{e^{-\alpha}}{\sigma^2} \exp\left(\frac{-1}{2\sigma^2}(\theta - \mu)\right)$$

(2.5)

$$\alpha = \int_0^Z \overline{n}(z) dz = \int_0^Z \left(\rho_n(z)\sigma_a + \mu_m(z)\right) dz$$

(2.6)

$$\mu = \int_0^Z n_y(z) / n_0 dz$$

(2.7)

$$\sigma^2 = \frac{1}{2} \frac{\tau(Z) W_0}{\alpha_p} = \frac{1}{2} \int_0^Z \rho_n(z) \sigma_a / \alpha_p dz$$

(2.8)

As introduced in 1.2, ABI systems take samples of the rocking curve using the analyzer crystal. However, those samples are affected by Poisson noise because of the discrete nature of photons. According to equation (2.1) our angular samples $y_i$ are Poisson distributed with mean $g(\theta_l; \alpha, \mu, \sigma^2)$ where $\theta_l$ correspond to the $l$th angular sample acquired. Each angular sample is treated as an independent Poisson random variable.

$$y_i \sim \text{Poisson}\left(g(\theta_l; \alpha, \mu, \sigma^2)\right)$$

(2.9)
According to the physical model, we consider each pixel’s AIP independent of each other. So, MIR parameters can be estimated sequentially pixel by pixel. This is the approach for the Classical Estimator (CE), Least Squares (LS) and Maximum Likelihood (ML) introduced in sections 2.2, 2.4 and 2.3 respectively. As shown in 2.5 a regularized method is introduced which forces to reconstruct the images as a whole.

### 2.2 Classical Estimator (CE)

The Classical Estimator was the first reconstruction method proposed for MIR [5]. CE attempts to measure \( \left[ \alpha, \mu, \sigma^2 \right]^T \) as the logarithm of the integral (2.11), the first (2.12) and the second (2.13) moment of the AIP where \( I_0 = \int R(\theta) d\theta \), the input radiation.

\[
\hat{I}_0 = \sum_{l=0}^{L-1} y_l
\]

\[
\hat{\alpha} = -\log \left( \frac{\hat{I}_0}{I_0} \right)
\]
CE is designed based on the physical understanding of the absorption, refraction and USAXS phenomena. However, from the statistical and object model point of view, it is not considering the noise nature of the samples or the rocking curve’s shape. As it will be seen in section 4.2, this leads to highly biased and noisy reconstructed images.

2.3 Maximum Likelihood (ML)

Maximum Likelihood (ML) is a standard method for probabilistic estimation widely used in many applications. It is a very well known approach and there are several results involving the ML estimator proving its good properties.

As it will be discussed in section 4.1, our estimators will be evaluated in terms of variance and bias (or, equivalently, bias squared). Therefore, we seek for an efficient (unbiased and with minimum variance) estimator, if possible. The minimum variance that an unbiased estimator can achieve is called Cramér-Rao Lower Bound [14]. A useful result for ML estimators is that if there exists an efficient estimator of a parameter, it has to be the ML estimator [15].

ML estimator maximizes the sampling probability distribution of the acquired data \( p(y | \alpha, \mu, \sigma^2) \), called Likelihood, as a function of the parameters \( [\alpha, \mu, \sigma^2]^T \), where \( y \) is the vector of samples. Since the MIR statistical model considers independent
Poisson samples, the joint Likelihood function is the product of the individual Poisson Likelihoods (2.14).

\[
p(y | \alpha, \mu, \sigma^2) = \prod_{l=0}^{L-1} \frac{1}{\gamma_l!} g(\theta_l; \alpha, \mu, \sigma^2)^{y_l} \exp\left(-g(\theta_l; \alpha, \mu, \sigma^2)\right)
\]  

For computational efficiency reasons, ML methods usually maximize the log-Likelihood function (2.15), which is equivalent because the logarithm is a monotonic increasing function.

\[
L(\alpha, \mu, \sigma^2) = \log\left(p(y | \alpha, \mu, \sigma^2)\right) = \\
= \sum_{l=0}^{L-1} \left(-g(\theta_l; \alpha, \mu, \sigma^2) + y_l \log\left(g(\theta_l; \alpha, \mu, \sigma^2)\right) - \log(y_l!)\right)
\]  

As a result, the ML estimator, or MLE, is described by (2.16), and it is computed by maximizing (2.15).

\[
\left[\hat{\alpha}, \hat{\mu}, \hat{\sigma}^2\right] = \text{argmax}\left(L(\alpha, \mu, \sigma^2)\right)
\]  

2.4 Least squares estimator (LS)

Least Squares estimator attempts to minimize the squared distance between the image formation model and the empirical samples. \(y_i, i = 0 \ldots (L-1)\) are the samples of the Angular Intensity Profile (AIP) where \(L\) is the number of samples taken at positions \(\theta_i\). According to the system model explained in 2.1, the AIP samples follow a Poisson distribution of parameter \(g(\theta_i)\), the convolution of the Rocking curve \(R\) of the system and the object function \(f\), (2.1), (2.5) and (2.9).
Therefore, LS method attempts to minimize the objective function

\[ J(\alpha, \mu, \sigma^2) = \sum_{i=0}^{L-1} \left( y_i - g(\theta_i; \alpha, \mu, \sigma^2) \right)^2 \]

with respect the parameter vector \([\alpha, \mu, \sigma^2]^T\),

(2.17).

\[
\begin{bmatrix}
\hat{\alpha}, \hat{\mu}, \hat{\sigma}^2
\end{bmatrix} = \text{argmin}\left(J(\alpha, \mu, \sigma^2)\right).
\]

(2.17)

LS method is implicitly assuming Gaussian noise in the samples. This is because minimizing the distance between the model and the observations is equivalent to the Maximum Likelihood method for the Gaussian case (same variance for all angular samples). However, Poisson random variables have higher variance for higher expected value.

### 2.5 Regularization: Maximum a Posteriori

Regularization is a common procedure in image reconstruction to obtain less noisy results. An image is a mapping of a physical property that can be considered continuous. This means that close pixels should have similar values because they correspond to close points of the imaged object. However, when reconstructing the MIR, like other medical imaging techniques, the pixel values are estimated independently without considering this prior knowledge. This leads to images that, despite they are statistically correct estimations, are noisy.

In order to obtain smoother images, the log-Likelihood function is modified with a regularization term that constrains the squared distance between each \(i\)th pixel with the ones in its 4-neighbor set \(\mathcal{N}_i\). Equation (2.18) shows the modified log-Likelihood where

\[ J_{\text{reg}}(\alpha, \mu, \sigma^2) = \sum_{i=0}^{L-1} \left( y_i - g(\theta_i; \alpha, \mu, \sigma^2) \right)^2 + \lambda \sum_{i=0}^{L-1} \sum_{j \in \mathcal{N}_i} \left( \alpha_{ij} - \alpha \right)^2 \]

and \(\lambda\) is a regularization parameter.
$\beta$ is the trade off parameter between the ML term and the smoothing term and $y'_i$ is the raw data measurements for the ith pixel $\left[\alpha_i, \mu_i, \sigma_i^2\right]^T$.

$$L(\alpha_i, \mu_i, \sigma_i^2) = \sum_{i=0}^{L-1} \left( -g(\theta_i, \alpha_i, \mu_i, \sigma_i^2) + y'_i \log\left( g(\theta_i, \alpha_i, \mu_i, \sigma_i^2) \right) - \log\left( y'_i \right) \right) +$$

$$-\frac{1}{2} \beta \cdot \sum_{k \in \mathbb{R}} \left[ (\alpha_i - \alpha_k)^2 + (\mu_i - \mu_k)^2 + (\sigma_i - \sigma_k^2)^2 \right]$$ (2.18)

Since equation (2.18) involves the reconstructed values for the neighbor pixels, the image must be reconstructed as a whole and not pixel-by-pixel. Considering the vector of all pixel values $\mathbf{x}$ and $\mathbf{x}_i = \left[\alpha_i, \mu_i, \sigma_i^2\right]^T$, the whole objective function to maximize in its compact form is (2.19). Since the smoothing term in (2.18) is quadratic, $\mathbf{R}$ is a constant matrix that corresponds to the smoothing term’s Hessian matrix.

$$J(\mathbf{x}) = \sum_i L(\mathbf{x}_i) - \beta \mathbf{x}^T \mathbf{R} \mathbf{x}$$

$$\hat{x} = \arg\max (J(\mathbf{x}))$$ (2.19)

The regularization proposed in (2.18) is equivalent to a Maximum a Posteriori (MAP) with a Gaussian prior $p(\mathbf{x})$ and it is equivalent to maximize the posterior probability of the parameters $p(\mathbf{x} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}) \cdot p(\mathbf{x})$ [16].
Section 2.3 describes the ML approach to parameter estimation as the maximization of the likelihood function. In chapter 3, a practical numerical implementation is proposed for this task. There are several numerical methods for maximizing functions; most of them based on the very well known steepest ascend. In addition, improvements like Conjugate Gradient (CG) [17], also very common in optimization, achieve faster convergence rates. Here we present a faster conjugate gradient approach to maximize the Likelihood of the MIR parameters or any Poisson ML problem.

3.1 Introduction to conjugate gradient (Quadratic case)

Given a function quadratic function \( f(x) = c + b^T x + \frac{1}{2} x^T Q x \) and a starting point \( x_0 \), it is possible to sequentially find a set of \( N = \text{dim}(x) \) pairs \( (\alpha_k, d_k) \) such that the update formula \( x_{k+1} = x_k + \alpha_k d_k \) leads to the \( f \) optimal in \( N \) steps. This method is known as conjugate gradient descend because the directions \( \{d_0, d_1, \ldots, d_{N-1}\} \) used are conjugate for matrix \( Q \).

Definition: A set of vectors \( \{d_0, \ldots, d_{N-1}\} \) are said to be conjugate for \( f \) if \( d_i^T Q d_j = 0 \) for any \( i \) and \( j \).
Theorem: The update formula (3.1) for finding \((\alpha_k, d_k)\), where \(h_k\) is the gradient vector at the point \(x_k\), and considering an arbitrary starting point \(x_0\) and \(d_0 = h_0\) leads to the \(f\) optimal in \(N = \dim(x)\) steps through conjugate directions \(d_k\).

\[
\begin{align*}
x_{k+1} &= x_k + \alpha_k d_k \\
\alpha_k &= -\frac{h_k^T d_k}{d_k^T Q d_k} \\
d_{k+1} &= -h_{k+1} + \beta_k d_k \\
\beta_k &= \frac{h_{k+1}^T Q d_k}{d_k^T Q d_k}
\end{align*}
\] (3.1)

Proof: A full demonstration can be found in [18], section 9.3. Our method proposed for non-quadratic optimization is proved to converge in approximately \(N\) steps by adapting this proof.

As it can be appreciated in (3.1), each conjugated direction \(d_{k+1}\) can be computed by conjugating with respect to the last ascend direction. This is possible due to the fact that the Hessian matrix \(Q\) is constant for all \(x\).

3.2 Classic non-quadratic extension (Polak-Ribière)

Generally, explicit conjugation for non-quadratic problems is not possible. But, compared with gradient descend; faster convergence rates can be achieved approximating the quadratic CG to the non-quadratic case.

Usually, a second order Taylor approximation at each step is performed using the local Hessian matrix \(H_f\) as the quadratic form in (3.1) as seen in (3.2). Also, the step size
\( \alpha_k \) is not computed with a closed form. Instead, a line search algorithm is used to approximate the directional minimum \( \alpha_k = \arg \min_\alpha (f(x_k + \alpha d_k)) \).

\[
x_{k+1} = x_k + \alpha_k d_k \\
d_{k+1} = -h_{k+1} + \beta_k d_k \\
\beta_k = \frac{h_{k+1}^T H_j d_k}{d_k^T H_j d_k}
\] (3.2)

Polak-Ribière (PR), in addition, uses an alternative expression for computing \( \beta_k = (g_{k+1} - g_k)^T g_k / g_k^T g_{k+1} \) which is an equivalent formula for the quadratic case. This is useful to avoid the computation of \( H_j \).

### 3.3 L-Conjugate algorithm

In section 3.2 the classical Polak-Ribière approach for optimizing a non-quadratic function is explained. However, it is possible to perform an actual CG algorithm for the Poisson Log-Likelihood function, which converges faster than the PR method. This section defines L-conjugate directions, describes how to sequentially find a L-conjugate base of the parameter space and a method to achieve the optimal in approximately \( N = \dim(x) \) steps.

L-conjugate algorithm is valid for any Log-Likelihood of a joint independent Poisson random variables with a smooth transformation of parameters \( g(x) \), (3.3) and (3.4). This is an extension and formalization of the method developed in [19] which was designed for a PET reconstruction system by Maximum Likelihood. For the PET case, the transformation of parameters is linear while, for our ABI parameter reconstruction, it...
is not, (2.1). Also, we produce an approximate analytical expression for the ascend step, which can be computed much faster than the proposed line search.

\[ \mathbf{g} : \mathbb{R}^N \rightarrow \mathbb{R}^L \]

\[ \mathbf{x} \rightarrow \mathbf{g} (\mathbf{x}) \]

\[ L (\mathbf{x}) = \sum_{i=0}^{L-1} \left( y_i \cdot \log \left( g_i (\mathbf{x}) \right) - g_i (\mathbf{x}) - \log (y_i! \right) \]  

(3.4)

When applied to MIR parameter estimation, the function \( \mathbf{g} \) components correspond to the \( L \) angular samples acquired and the variables correspond to the parameters. However, we can also consider all the angular samples for all pixels in an image as the function components and all the MIR parameters for all pixels as the variables. This means that we can still use the L-conjugate algorithm for reconstructing a whole image at the same time and not only pixel-by-pixel. This property is important because when we introduce spatial smoothing to the ML optimization, neighbor pixels values must be considered in the optimization.

First, a new definition of conjugation is needed. We pretend to find ascend directions locally conjugate to all past directions.

**Definition:** Given an ordered set \( \{ \mathbf{d}_0, \ldots, \mathbf{d}_{N-1} \} \) of vectors in the parameter space \( \mathbb{R}^N \), the scalar values \( \{ \alpha_0, \ldots, \alpha_{N-1} \} \) and a starting point \( \mathbf{x}_0 \), we say that \( \{ \mathbf{d}_0, \ldots, \mathbf{d}_{N-1} \} \) are L-conjugate for the set of matrices \( \{ \mathbf{H}_0, \ldots, \mathbf{H}_{N-1} \} \) if (3.5). Where

\[ \mathbf{H}_k = \mathbf{D}^T \mathbf{g} (\mathbf{x}_k) \cdot \mathbf{A}^{-1} (\mathbf{x}_k) \cdot \mathbf{D} \mathbf{g} (\mathbf{x}_k) \] is the information matrix (4.10) evaluated in \( \mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_{k-1} \mathbf{d}_{k-1} \), \( \mathbf{D} \mathbf{g} \) is the differential matrix of \( \mathbf{g} \) and the diagonal matrix \[ [\mathbf{A}]_{ll} = g_{li} (\mathbf{x}) \].
\[ d_i^\top H_{j} d_k = 0 \quad \forall i < k \quad (3.5) \]

The expression of the Fisher information matrix is derived in section 4.1. However, the fact that the matrices used for conjugation are the Fisher information is not relevant for the mathematical discussion of this chapter.

Second, some properties of L-Conjugate vectors need to be provided in order to justify the fast convergence of the final algorithm.

**Lemma:** The Fisher information matrix of the MIR parameters is positive definite for any parameter value \( x \).

**Proof:** We need to prove that given any vector \( v \), \( v^\top I(x)v > 0 \) for all \( x \). We know that \( A^{-1} \) is a positive definite matrix because it is diagonal and all the elements are positive. Then, given \( u = D_g v \), the product \( u^\top A^{-1}u > 0 \). So the Fisher information matrix is positive definite too (3.6).

\[ v^\top I(x)v = v^\top D_g^\top D_g v = (D_g v)^\top A^{-1} D_g v = u^\top A^{-1}u > 0 \quad (3.6) \]

**Lemma:** L-Conjugate vectors are linear independent.

**Proof:** Given an ordered set \( \{d_0 \ldots d_{N-1}\} \) of conjugate vectors in the parameter space \( \mathbb{R}^N \) for the set of matrices \( \{H_0 \ldots H_{N-1}\} \) we want to prove (3.7).

\[ \sum_{j=0}^{N-1} \lambda_j d_j = 0 \iff \lambda_0 = \cdots = \lambda_{N-1} = 0 \quad (3.7) \]

The left implication is obvious. To prove the right implication we start by left multiplying the linear combination of the L-conjugate vectors for \( d_0^\top H_0 \). Since the vectors are L-conjugate, there is only one surviving term because \( \lambda_j d_0^\top H_0 d_j = 0 \) for all
\( j \) greater than zero. Since \( H_k \) are all positive definite at all \( x_k \), \( d_0^T H_0 d_0 \neq 0 \) and we can conclude (3.8).

\[
\lambda_0 d_0^T H_0 d_0 = 0 \Rightarrow \lambda_0 = 0 \Rightarrow \sum_{j=1}^{N-1} \lambda_j d_j = 0
\]  

(3.8)

Sequentially we can repeat the same procedure left multiplying for all \( d_j^T H_k \) to conclude that all \( \lambda_k \) are 0.

**Observation:** If \( \dim(x) = N \), then \( N \) L-Conjugate vectors are a base of the parameter space.

Third, we will show that an approximately optimal step sizes \( \alpha_j \) for a given L-Conjugate base can be computed at each step to reach the ML-optimal point \( x^* \).

**Theorem:** Given a set of L-conjugate vectors base \( \{d_0, \ldots, d_{N-1}\} \) and a starting point \( x_0 \), the ML-optimal point \( x^* \) can be approximately expressed as

\[
x^* - x_0 = \alpha_0 d_0 + L + \alpha_{N-1} d_{N-1} \quad \text{where} \quad \alpha_k = d_k^T h_k / d_k^T H_k d_k.
\]

**Proof:** Since \( \{d_0, \ldots, d_{N-1}\} \) are a base, there is a linear combination of the vectors to form the difference vector (3.9).

\[
x^* - x_0 = \alpha_0 d_0 + \cdots + \alpha_{N-1} d_{N-1}
\]  

(3.9)

Since \( x_k = x_0 + \alpha_0 d_0 + L + \alpha_{k-1} d_{k-1} \), we can express equation (3.9) as follows:

\[
x^* - x_k = \alpha_j d_j + L + \alpha_{N-1} d_{N-1} \quad \text{. Then, using the L-Conjugate property (3.5), we obtain equation (3.10).}
\]

\[
d_k^T H_k \left( x^* - x_k \right) = d_k^T H_k \left( \alpha_k d_k + \cdots + \alpha_{N-1} d_{N-1} \right) = \alpha_k d_k^T H_k d_k
\]  

(3.10)
Finally, using a Taylor expansion of the function \( g \) on the point \( x_k \) and approximating \( y = g(x^*) \), the expression (3.11) for \( \alpha_k \) is obtained.

\[
\alpha_k = \frac{d_k^T H_k (x^* - x_k)}{d_k^T H_k d_k} \approx \frac{d_k^T D^g(x_k) \cdot A^{-1}(x_k)(y - g(x_k))}{d_k^T H_k d_k} = \frac{d_k^T h_k}{d_k^T H_k d_k} \tag{3.11}
\]

**Observation:** It is possible to compute the approximation for \( \alpha_k \) using only the ascend direction, the gradient and the Fisher information matrix at \( x_k \). Actually, the approximation in (3.11) can be improved using a line search method when \( x_k \) is far from \( x^* \), where the Taylor approximation is worse. Also, pre-filtering the angular samples \( y \), the approximation \( y_j : g_j(x) \) is more accurate. The expression (3.11) corresponds to the maximum of the second order Taylor approximation at \( x_k \).

Finally, we present an algorithm for computing the conjugate directions at each step that in addition to the approximate formula of the step size \( \alpha_k \) allows us to approximately achieve the optimal in \( N \) iterations.

The procedure to conjugate directions is based on the Gram-Schmidt (GS) algorithm for orthonormalizing a base of vectors. In this case, we assume the gradients at each point are a base of \( \mathbb{R}^N \) and we want to form a conjugate base from them. To do so, the approach is, like in GS, subtracting to each element from the gradients base its projection to the subspace of the past conjugate directions. Where projection must be understood according to the “norm” defined by the matrices \( H_k \). This computation is summarized in (3.12), where \( -\beta_j^{(k)} \) are the projections of \( h_k \) to the past direction \( d_j \).
\[ \mathbf{d}_k = \mathbf{h}_k + \sum_{j=0}^{k-1} \beta_j^{(k)} \mathbf{d}_j \] (3.12)

To compute \( \beta_j^{(k)} \), we force each \( \mathbf{d}_k \) to be L-conjugate to the past directions resulting in equation (3.13).

\[
\begin{align*}
\mathbf{d}_j^T \mathbf{H}_j \mathbf{d}_k &= \mathbf{d}_j^T \mathbf{H}_j \left( \mathbf{h}_k + \sum_{j=0}^{k-1} \beta_j^{(k)} \mathbf{d}_j \right) = 0 \\
\Rightarrow \sum_{j=0}^{k-1} \beta_j^{(k)} \mathbf{d}_j^T \mathbf{H}_j \mathbf{d}_j &= -\mathbf{d}_j^T \mathbf{H}_j \mathbf{h}_k & i = \{0, \ldots, k-1\} 
\end{align*}
\] (3.13)

As a result, a system of \( k \) equations must be solved at each iteration to conjugate the ascend direction \( \mathbf{M}^{(k)} \cdot \mathbf{b}^{(k)} = \mathbf{c}^{(k)} \). This might seem an arduous computational task, however, the system has a triangular matrix that can be built sequentially at each iteration, which saves a lot of computational time (3.14).

\[
\mathbf{M}^{(k)} = 
\begin{pmatrix}
\mathbf{d}_0^T \mathbf{H}_0 \mathbf{d}_0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
\mathbf{d}_{k-1}^T \mathbf{H}_{k-1} \mathbf{d}_0 & \cdots & \cdots & \mathbf{d}_{k-1}^T \mathbf{H}_{k-1} \mathbf{d}_{k-1}
\end{pmatrix} \quad (3.14).
\]

\[
\mathbf{c}^{(k)} = 
\begin{pmatrix}
-\mathbf{d}_0^T \mathbf{H}_0 \mathbf{h}_k \\
\vdots \\
-\mathbf{d}_{k-1}^T \mathbf{H}_{k-1} \mathbf{h}_k
\end{pmatrix} \quad (3.15)
\]

This procedure is possible until \( k = N-1 \) since it does not make sense to obtain a bigger set of L-conjugate vectors in a \( N \)-dimensional space because they are linearly independent. However, after \( N \) iterations the algorithm might not have converged due to the approximation in (3.11). That is why after \( N \) iterations the algorithm must re-start or conjugate for the past \( N-1 \) directions only: (3.16) and (3.17).
\[
M^{(k)} = \begin{pmatrix}
    d_{k-N}^T H_{k-N} d_{k-N} & 0 & \cdots & 0 \\
    \vdots & \ddots & \ddots & \vdots \\
    \vdots & \ddots & \ddots & 0 \\
    d_{k-1}^T H_{k-1} d_{k-N} & \cdots & d_{k-1}^T H_{k-1} d_{k-1}
\end{pmatrix}
\]  

(3.16)

\[
c^{(k)} = \begin{pmatrix}
    -d_{k-N}^T H_{k-N} h_k \\
    \vdots \\
    -d_{k-1}^T H_{k-1} h_k
\end{pmatrix}
\]  

(3.17)

**Algorithm:** Initialize \( d_0 = h_0 \), \( \alpha_0 = \frac{d_0^T h_0}{d_0^T H_0 d_0} \), \( M^{(0)} = d_0^T H_0 d_0 \) and iterate the steps i, ii and iii for \( k = 1 \ldots N-1 \).

i. Update parameters: \( x_k = x_{k-1} + \alpha_{k-1} d_{k-1} \).

ii. Add the needed row and column to \( M^{(k-1)} \) to build \( M^{(k)} \) as in (3.14) and build the \( c^{(k)} \) following (3.15).

iii. Solve the system \( M^{(k)} \cdot \beta^{(k)} = c^{(k)} \), compute next direction and step (3.18).

\[
d_k = h_k + \sum_{j=0}^{k-1} \beta_j^{(k)} d_j
\]  

(3.18)

\[\alpha_k = \frac{d_k^T h_k}{d_k^T H_k d_k}\]

After \( N \) iterations, iterate the next steps until desired convergence criteria is accomplished:

iv. Update parameters: \( x_k = x_{k-1} + \alpha_{k-1} d_{k-1} \).

v. Delete first row and column and add the needed row and column to \( M^{(k-1)} \) to build \( M^{(k)} \) as in (3.16). Build \( c^{(k)} \) following (3.17).

vi. Solve the system \( M^{(k)} \cdot \beta^{(k)} = c^{(k)} \), compute next direction and step (3.19).
\[ d_k = h_k + \sum_{j=k-N}^{k-1} \beta_j \alpha_j d_j \]
\[ \alpha_k = \frac{d_k^T h_k}{d_k^T H_k d_k} \quad (3.19) \]

**Example:** Next, we show a small 2-D example in figure 9 where a Poisson Log-Likelihood (3.4) is maximized with a transformation of parameters like in (3.3) with \( N = 2 \) and \( L = 6 \). The methods compared are the classical Polak-Riviere Conjugate Gradient, the LC algorithm and the LC algorithm with the optimal (analytically obtained) step using the known optimal value \( x' \) (3.11). In one hand, the examples show how the conjugate direction leads to the optimal value using the analytical expression (Red) and that the step approximation (Line Search) leads to a very good approximation in two steps. In the other hand, classical Polak-Riviere methods do not conjugate the directions properly for this non-quadratic case.

![Figure 9. Two examples of Poisson Log-Likelihood optimization.](image)
3.4 Extension for MAP

In section 3.3 it has been developed an actual gradient conjugation for the joint Poisson Log-Likelihood called L-conjugate algorithm. However, when a regularization term is included the function to maximize is different. Despite that fact, explicit conjugation is still possible because the regularization term is a quadratic function.

Following the notation used in section 3.3, when computing the conjugated directions for the modified Log-Likelihood (2.19), the matrices $H_k$ are not the information matrices at each point. The matrices with respect the conjugation is done are modified with the quadratic form $R$ in of the regularization term as shown in (3.20).

$$H_k = D_d^T(x_k) \cdot A^{-1}(x_k) \cdot D_d(x_k) + \beta \cdot R$$

(3.20)

For the regularized case, the gradient of the objective function is slightly different from the standard log-Likelihood function (4.8) as seen in equation (3.21).

$$h_k = D_d^T(x_k) \cdot A^{-1}(x_k) \cdot (y - g(x_k)) - \beta \cdot R \cdot x_k$$

(3.21)

To see that with this matrices the conjugation is still possible we need to see that they are positive definite (so it is possible to build the conjugated basis) and that the step size can be computed from local properties such as gradient and the $H_k$ matrices.

First, since the matrix $R$ is a quadratic form, it is positive definite and therefore $H_k$ too. Second, we need to recompute the $\alpha_k$ values (3.11) for the regularized case (3.22) using that the gradient at the optimal point $h^*$ is zero,

$$h^* = D_d^T(x^*) A^{-1}(x^*) (y - g(x^*)) - \beta R x^* = 0.$$
\[
\alpha_k = \frac{d_k^T H_k (x' - x_k)}{d_k^T H_k d_k} = \frac{d_k^T D_g^T (x_k) \cdot A^{-1} (x_k) D_g (x_k) \cdot (x' - x_k)}{d_k^T H_k d_k} + \frac{\beta d_k^T R (x' - x_k)}{d_k^T H_k d_k} \\
= \frac{d_k^T D_g^T (x_k) \cdot A^{-1} (x_k) D_g (x_k) \cdot (x' - x_k)}{d_k^T H_k d_k} + \frac{d_k^T D_g^T (x^*) A^{-1} (x^*) (y - g(x^*)) - \beta d_k^T R x_k}{d_k^T H_k d_k} \\
= \frac{d_k^T D_g^T (x_k) \cdot A^{-1} (x_k) (y - g(x_k)) - \beta d_k^T R x_k}{d_k^T H_k d_k} = \frac{d_k^T h_k}{d_k^T H_k d_k}
\]
CHAPTER 4
RESULTS, CONCLUSIONS AND FUTURE WORK

4.1 Comparison of ML, LS and CE estimators with simulated data

In estimation problems, such MIR parameters reconstruction, performance is evaluated in terms of Mean Square Error (MSE) (4.1), Bias (4.2) and Variance (4.3). Note that in this section we are considering single pixel values $x = [\alpha, \mu, \sigma^2]^T$.

\[
\text{MSE}(\hat{x}) = E\left\{ (x - \hat{x})^2 \right\} \quad (4.1)
\]

\[
\text{Bias}(\hat{x}) = E\left\{ x - \hat{x} \right\} \quad (4.2)
\]

\[
\text{Var}(\hat{x}) = E\left\{ (\hat{x} - E\{\hat{x}\})^2 \right\} \quad (4.3)
\]

Those metrics are related by $\text{MSE}(\hat{x}) = \text{Bias}^2(\hat{x}) + \text{Var}(\hat{x})$, which is easy to prove developing the squares in the expressions. Ideally, we seek for Minimum Variance Unbiased estimator (MVU). Since $\text{Bias}(\hat{x}) = 0$ for MVU, they are also minimum MSE estimators. Usually, those performance metrics depend on the value of the parameters. In our case, this property means that the estimation accuracy is pixel-value dependent. That force us to evaluate the quality of the estimators in terms of mean squared bias (mBias), mean variance (mVar) and mean mean squared error (mMSE).

The only possible approach is designing an experiment with simulated data to empirically estimate the performance.

The simulated data is based in a human breast study data acquired in Brookhaven National Laboratory. Using a Synchrotron Light Source, like the one at Brookhaven, is
very convenient due to its high brilliance and coherence. The amount of photons is high enough to consider our data noiseless. For the Brookhaven experiment, photon energy was set to 60 keV because this corresponds to a standard energy value of a conventional x-Ray tube source. This set up recreates the environment that a tabletop machine would have except for the noise, that can be considered absent. Also, 25 analyzer positions were imaged in a range between $-2 \mu m$ and $2 \mu m$. The raw data was used to reconstruct the three parametric images $\mathbf{x} = \left[ \alpha, \mu, \sigma^2 \right]^T$ with the classical MIR algorithm and those noise-free images are used to simulate new noisy samples.

Using (2.1) and (2.5), we can obtain the object function $f$ and the image formation model $g$ for any rocking curve $R$ that our system could have and for each of the pixels of the noiseless images. To obtain the new raw data we only need to simulate a Poisson random variable $y$ with our model $g$ as a parameter evaluated at any angular sampling pattern desired.

Figure 10. ROI: breast calcification. Left: absorption image. Center: refraction image. Right: USAXS image.

To estimate the mBias, mVar and mMSE of our methods, we select a small image region of interest (ROI) with a diverse range of values for $\mathbf{x} = \left[ \alpha, \mu, \sigma^2 \right]^T$. The ROI
corresponds to a 56x56 pixels area containing a calcification, see Figure 10. Calcifications are important artifacts on breast cancer diagnosis and phase-contrast could potentially increase its detectability. Across the ROI, the bias, variance and MSE are estimated for each pixel running multiple noise simulation (and their corresponding estimations) as (4.4), (4.5) and (4.6) where \( \bar{x} \approx \frac{1}{N_{real}} \sum_{i=1}^{N_{real}} \hat{x}_i \) and \( \hat{x}_i \) are the estimations for each noise realizations.

\[
\text{bias}(x) = \frac{1}{N_{real}} \sum_{i=1}^{N_{real}} (\hat{x}_i - x) \tag{4.4}
\]

\[
\text{var}(x) = \frac{1}{N_{real}} \sum_{i=1}^{N_{real}} (\hat{x}_i - \bar{x})^2 \tag{4.5}
\]

\[
\text{MSE}(x) = \frac{1}{N_{real}} \sum_{i=1}^{N_{real}} (\hat{x}_i - x)^2 \tag{4.6}
\]

After that, averaging across the ROI the obtained \( \text{bias}^2(x) \), \( \text{var}(x) \) and \( \text{MSE}(x) \), we obtain an estimation for (4.1), (4.2) and (4.3).

In addition to compare the three estimators, we will compute the Cramér-Rao Lower Bound (CRLB). Cramér-Rao inequality is a well known result in estimation theory in which a theoretical lower bound is given for unbiased estimators variance. According to the result presented for Cramér and Rao in [14], the CRLB can be computed as (4.7), where \( x = [x_1, x_2, x_3]^T = [\alpha, \mu, \sigma^2]^T \) and \( I(x) \) is the Fisher Information matrix.
\[
\text{Var}(\hat{x}) \geq \text{diag}\left(\Gamma^{-1}(x)\right)
\]

\[
[I(x)]_{ij} = -E\left\{ \partial^2 \log(p(y|x)) \over \partial x_i \partial x_j \right\}
\]

\[
I(x) = -E\left\{ (\nabla L(x)) (\nabla^T) \right\}
\]

For MIR parameters, we can express that depending on the model function \( g \) presented in (2.1) and its differential matrix \( D_g \). To compute the Fisher information matrix, using expression (4.7), we need to develop an expression for the gradient (4.8) and for the Hessian matrix \( (H_L) \) (4.9) of the Log-Likelihood (2.15). The expected value is taken using \( E\{y_i\} = g_t(x) \). Finally, we obtain an expression with \( D_g \) and the diagonal matrix \( [A]_{ij} = g_{ti}(x) \) in (4.10).

\[
h = \nabla L(x) = \sum_{i=0}^{L-1} \nabla g_i(x) (y_i - g_i(x)) = D_g^T \Lambda^{-1} (y - g(x))
\]

\[
H_L = (\nabla L(x)) \nabla^T = D_g^T \Lambda^{-1} y \nabla^T - D_g^T \Lambda^{-1} g \nabla^T
\]

\[
= \left( \sum_{i=0}^{L-1} \frac{\partial g_i(x)}{\partial x} \frac{y_i}{g_i(x)} \right) \cdot \nabla^T - \left( \sum_{i=0}^{L-1} \frac{\partial g_i(x)}{\partial x_j} \right)_j \cdot \nabla^T
\]

\[
= \sum_{i=0}^{L-1} \frac{y_i}{g_i(x)} \left( H_{g_i} - \frac{\nabla g_i(x) \cdot (\nabla g_i(x))^T}{g_i(x)} \right) - \sum_{i=0}^{L-1} H_{g_i}
\]

\[
I(x) = -E\{H_L\} = -E\left\{ \sum_{i=0}^{L-1} \frac{y_i}{g_i(x)} \left( H_{g_i} - \frac{\nabla g_i(x) \cdot (\nabla g_i(x))^T}{g_i(x)} \right) - \sum_{i=0}^{L-1} H_{g_i} \right\}
\]

\[
= \sum_{i=0}^{L-1} \frac{\nabla g_i(x) \cdot (\nabla g_i(x))^T}{g_i(x)} = D_g^T(x) \cdot \Lambda(x) \cdot D_g(x)
\]
Using expression (4.10), we can compute the CRLB for a given value of $x$. It is important to realize that the CRLB is strictly dependent on the value of the parameters we pretend to estimate. Therefore, the theoretical bound we just derived is pixel dependent. Due to that, we will use the mean bound (mCRLB) to as a reference of the minimum mean Variance for unbiased estimators. We define mCRLB as the mean CRLB across the pixels.

### 4.2 Results and discussion

In order to compare the three MIR parameter estimators exposed in the past sections, its performance is presented in terms of mMSE, mBias Squared and mVariance in Table 1, Table 2 and Table 3. Also, the mean Cramer-Rao Lower Bound (mCRLB) is plotted for the three parameters. The mCRLB is the mean of the CRLB of each of the pixels in the ROI introduced in section 4.1.

This experiment has been done with a uniform sampling pattern between $-2$ and $2 \mu$rad using 7 angular samples. A total exposure of 1800 photons per pixel has been simulated. This is an standard dose for a conventional X-ray tube. Also, the rocking curve has been simulated as a Pearson type VII function (4.11). It has been shown in [20] and [21] that ABI systems intrinsic curve can be modeled by this function. In the case of using a Si(333) with a 60 keV source peak, the parameters values to fit the curve are $a = 0.5355$ and $m = 4.3005$. Since $c$ corresponds to the amplitude, it depends on the selected dose. In our 7 samples case and 1800 photons per pixel, $c = \frac{1800}{L \cdot 257.1429}$.

$$R(\theta) = c \left(1 + \frac{\theta^2}{ma^2}\right)^{-m}$$  \hspace{1cm} (4.11)
Table 1. Absorption estimation performance of ML, MAP, LS and CE

<table>
<thead>
<tr>
<th>CRLB</th>
<th>ML</th>
<th>MAP</th>
<th>LS</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.47 \cdot 10^{-3}</td>
<td>P-R</td>
<td>L-C</td>
<td>P-R</td>
<td>L-C</td>
</tr>
</tbody>
</table>

| mMSE | 3.50 \cdot 10^{-3} | 3.50 \cdot 10^{-3} | 6.84 \cdot 10^{-5} | 6.84 \cdot 10^{-5} | 3.68 \cdot 10^{-3} | 8.10 \cdot 10^{-3} |
| mBias\^2 | 1.50 \cdot 10^{-5} | 1.50 \cdot 10^{-5} | 6.01 \cdot 10^{-5} | 6.01 \cdot 10^{-5} | 1.34 \cdot 10^{-5} | 4.63 \cdot 10^{-3} |
| mVar  | 3.49 \cdot 10^{-3} | 3.49 \cdot 10^{-3} | 8.26 \cdot 10^{-6} | 8.26 \cdot 10^{-6} | 3.66 \cdot 10^{-3} | 3.48 \cdot 10^{-3} |

Table 2. Refraction estimation performance of ML, MAP, LS and CE

<table>
<thead>
<tr>
<th>CRLB</th>
<th>ML</th>
<th>MAP</th>
<th>LS</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.93 \cdot 10^{-3}</td>
<td>P-R</td>
<td>L-C</td>
<td>P-R</td>
<td>L-C</td>
</tr>
</tbody>
</table>

| mMSE | 1.94 \cdot 10^{-3} | 1.94 \cdot 10^{-3} | 2.62 \cdot 10^{-4} | 2.62 \cdot 10^{-4} | 2.30 \cdot 10^{-3} | 1.89 \cdot 10^{-3} |
| mBias\^2 | 6.67 \cdot 10^{-6} | 6.67 \cdot 10^{-6} | 2.02 \cdot 10^{-4} | 2.02 \cdot 10^{-4} | 7.75 \cdot 10^{-6} | 6.62 \cdot 10^{-6} |
| mVar  | 1.94 \cdot 10^{-3} | 1.94 \cdot 10^{-3} | 6.03 \cdot 10^{-5} | 6.03 \cdot 10^{-5} | 2.29 \cdot 10^{-3} | 1.89 \cdot 10^{-3} |

Table 3. USAXS estimation performance of ML, MAP, LS and CE

<table>
<thead>
<tr>
<th>CRLB</th>
<th>ML</th>
<th>MAP</th>
<th>LS</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.22 \cdot 10^{-3}</td>
<td>P-R</td>
<td>L-C</td>
<td>P-R</td>
<td>L-C</td>
</tr>
</tbody>
</table>

| mMSE | 2.23 \cdot 10^{-3} | 2.23 \cdot 10^{-3} | 1.35 \cdot 10^{-4} | 1.35 \cdot 10^{-4} | 3.72 \cdot 10^{-3} | 4.19 \cdot 10^{-3} |
| mBias\^2 | 7.64 \cdot 10^{-6} | 7.64 \cdot 10^{-6} | 7.52 \cdot 10^{-5} | 7.52 \cdot 10^{-5} | 2.10 \cdot 10^{-5} | 2.02 \cdot 10^{-3} |
| mVar  | 2.22 \cdot 10^{-3} | 2.22 \cdot 10^{-3} | 6.02 \cdot 10^{-5} | 6.02 \cdot 10^{-5} | 3.70 \cdot 10^{-3} | 2.17 \cdot 10^{-3} |

Those results prove the MLE to be the best estimator in terms of mMSE for absorption and USAXS. Also, despite it provides a higher mMSE for refraction, it has a lower bias and its variance is above the mCRLB. Considering Figure 12, using MAP, one can obtain a better result than with the CE or LS.

In [13], it was already proven that the CE was almost MVU estimator for more than 11 uniform samples. Here, we show that this result can be improved even for less samples and it can be easily extrapolated to any sampling pattern.
In addition, the results show, as expected, that P-R and L-C maximization algorithms achieve the same solution for ML and MAP estimators.

The MAP method performance shown in the tables are for $\beta = 750$. This value is selected to minimize the mMSE of the three dimensional estimation. The values tested for $\beta$ are 0, 100, 250, 500, 750, 1000, 2000 and 5000. MAP, as expected, causes a bias and reduces the variance showing a trade off between both accuracy metrics for the tree parameters. The choose of the optimal value for $\beta$ can be done in terms of mMSE or, in the future, with a task based optimization procedure. The trade-off behavior can be observed in Figure 11, Figure 12 and Figure 13 where the comparison with LS and CE estimators is shown in a different scale due to display purposes.

![Figure 11. ML absorption estimation performance for different values of $\beta$](image)
Figure 12. ML refraction estimation performance for different values of $\beta$

Figure 13. ML USAXS estimation performance for different values of $\beta$
The same trade-off showed in Figure 11, Figure 12 and Figure 13 is visually observed in Figure 14 with an example of reconstruction with several values for $\beta$.

![Reconstructed images for different values of $\beta$](image)

Figure 14. Reconstructed images for different values of $\beta$

To evaluate the performance of the L-Conjugate (LC) method described in section 3.3, we compare it with the convergence rate of the Polak-Ribiere (PR) algorithm. To do so, we repeat the experiment designed in section 4.2, and plot the convergence rate in terms of mMSE for the two iterative approaches.

First, we test the convergence speed without regularization term (MLE). The results can be examined in Figure 15 and Figure 16. One can easily observe that the L-conjugate converges in less number of iterations. In addition, L-conjugate method requires less computational time to converge, as it can be seen in Figure 16. Despite the
conjugation of the directions implies solving a linear equation system, since it is triangular, does not overload the computational time. Also, using the analytical approximation for the step size derived in (3.11) is less time consuming than a line-search approach, which is an iterative procedure. Figure 16 shows the average time consumed per iteration per pixel. In consequence, Figure 15 displays the mMSE for pixel at each iteration.

Figure 15. LC vs. PR convergence
To verify the extension to MAP, the ROC is reconstructed for several $\beta$ values and one noise realization. Figure 17, Figure 18, Figure 19 and Figure 20 show how the LC algorithm achieves a faster convergence than PR for different values of the regularization parameter $\beta$: Less iterations are needed for all cases. For a visual inspection of the faster convergence, Figure 21 shows reconstructed refraction images at different iterations for LC and PR and $\beta = 450$. Also, Figure 22 shows the same comparison between PR and LC but the images are aligned in terms of MSE (same column images have similar MSE).
Figure 17. PR vs. LC convergence for $\beta = 100$

Figure 18. PR vs. LC convergence $\beta = 450$
Figure 19. PR vs. LC convergence for $\beta = 800$

Figure 20. PR vs. LC convergence for $\beta = 2000$
<table>
<thead>
<tr>
<th>Iter.</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE L-C</td>
<td>2.093673·10⁻²</td>
<td>2.082284·10⁻²</td>
<td>2.072331·10⁻²</td>
<td>2.070920·10⁻²</td>
</tr>
<tr>
<td>L-C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time L-C</td>
<td>6.4065s</td>
<td>12.7298s</td>
<td>28.6215s</td>
<td>44.6063s</td>
</tr>
<tr>
<td>MSE P-R</td>
<td>2.102725·10⁻²</td>
<td>2.093364·10⁻²</td>
<td>2.081218·10⁻²</td>
<td>2.076018·10⁻²</td>
</tr>
<tr>
<td>P-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time P-R</td>
<td>67.2866s</td>
<td>134.2117s</td>
<td>303.7499s</td>
<td>636.1274s</td>
</tr>
</tbody>
</table>

Figure 21. LC vs. PR refraction reconstructions vs. number of iterations ($\beta = 450$)

<table>
<thead>
<tr>
<th>Iter.</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE L-C</td>
<td>2.093673·10⁻²</td>
<td>2.082284·10⁻²</td>
<td>2.072331·10⁻²</td>
<td>2.070920·10⁻²</td>
</tr>
<tr>
<td>L-C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time L-C</td>
<td>6.4065s</td>
<td>12.7298s</td>
<td>28.6215s</td>
<td>44.6063s</td>
</tr>
<tr>
<td>MSE P-R</td>
<td>2.102725·10⁻²</td>
<td>2.093364·10⁻²</td>
<td>2.081218·10⁻²</td>
<td>2.073534·10⁻²</td>
</tr>
<tr>
<td>P-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time P-R</td>
<td>67.2866s</td>
<td>134.2117s</td>
<td>303.7499s</td>
<td>636.1274s</td>
</tr>
</tbody>
</table>

Figure 22. PR vs. LC comparison ($\beta = 450$)
4.3 Conclusions and future work

In this thesis we presented developed novel reconstruction methods for tabletop, photon limited, analyzer-based phase-contrast imaging (ABI) systems. The presented reconstruction methods goal is to minimize the impact of CXS low brilliance by utilizing phase-contrast image formation model and maximum or maximum a posteriori Poisson likelihood approach. Finally, a fast-convergent conjugate gradient optimization algorithm has been derived specifically for Poisson likelihood function maximization.

The presented results shown that model-based maximum likelihood (ML) and maximum a posteriori likelihood (MAP) approaches for parametric image estimation can achieves a better mean-squared-error in comparison to classical MIR and simplistic model-based least squares estimation methods. Also, thanks to the conjugate gradient optimization algorithm derived specifically for Poisson likelihood function, developed in chapter 3, solution can be obtained faster.

In addition, the L-Conjugate algorithm proposed in chapter 3, is a general fast maximization method for any non-linear Poisson likelihood or a posteriori likelihood. The methodology derived here can be utilized other applications besides ABI reconstruction. Additionally, in this thesis we have proved that L-conjugate directions lead approximately to the ML estimate in approximately N steps (where N is the space dimension) and an analytical approximation to the step size is provided.

In global, this work is an important milestone in the reconstruction algorithm for ABI because it provides a statistically optimal frequentist estimator that can be computed efficiently.
We envision several extensions in future: First, we would like to extend this approach to tomography reconstruction. Second, we would like to explore other priors like total variation or more advance methods based on a Machine Learning approach [22]. Finally, ABI systems have an additional parameter that has a critical influence in the reconstruction: the angular sampling pattern. All the work presented here has been done for the standard uniform sampling with a fixed number of samples. However, the information acquired is very different depending on the number and position of the samples. Several work has been already done in this topic in [23] and [13], but a complete understanding of the sampling schemes influence on iterative reconstruction is still a pending issue.
BIBLIOGRAPHY


