

Trabajo Invitado

Numerical Issues in Speckled Imagery Analysis with Maximum Likelihood Parameter Estimation

Alejandro C. Frery¹ Francisco Cribari-Neto²
Marcelo O. de Souza²

¹Centro de Informática
Universidade Federal de Pernambuco,
CP 7851, 50732-970 Recife, PE – Brazil
frery@cin.ufpe.br

²Departamento de Estatística
CCEN, Universidade Federal de Pernambuco,
Cidade Universitária, 50740-540 Recife, PE – Brazil

Invited Paper

Abstract

This paper deals with numerical problems arising when performing maximum likelihood parameter estimation. The distribution under study arises when processing and analyzing images that obtained using coherent illumination, as is the case of sonar, laser, ultrasound-B and synthetic aperture radar data. The noise that appears in this kind of imagery is called *speckle*, and it can neither be assumed Gaussian nor additive. The properties of speckle noise are well described by the Multiplicative Model, a statistical framework from which stem several important distributions. Amongst these distributions, one is regarded as the Universal Model for speckled data, namely the \mathcal{G}^0 law. This paper deals with amplitude data, so the \mathcal{G}_A^0 will be used. The literature reports that techniques for obtaining estimates (maximum likelihood, based on moments and on order statistics) of the parameters of the \mathcal{G}_A^0 distribution require samples of hundreds, even thousands, of observations in order to converge to sensible values. This is verified for maximum likelihood estimators, and a proposal based on alternated optimization is proposed to alleviate this situation. This proposal is assessed with simulated real and with simulated data. A Monte Carlo experience is devised to estimate the quality of maximum likelihood estimators with small samples, and real data is successfully analyzed with this alternated procedure.

Keywords: image analysis, inference, likelihood, computation, optimization.

1 Introduction

Remote sensing by coherent illumination can be used to obtain information about unaccessible and/or unobservable scenes. The surface of Venus, remote and unobservable due to constant cloud cover, was mapped using radar sensor. Similar sensors are used to monitor inaccessible earth regions, such as the Amazon, the poles, etc. Ultrasound-B imagery is employed to diagnose without invading the body. Sonar images are used to map the bottom of the sea, lakes and deep or dark rivers, and laser illumination can be used to trace profiles of microscopic entities.

This kind of images are formed by active sensors (since they carry their own source of illumination) that send and retrieve signals whose phases are recorded. The imagery is formed detecting the echo from the target, and in this process a noise is introduced due to interference phenomena.

The noise that appears in these imagery is called *speckle*, and departs from classical hypothesis: it is not Gaussian in most cases, and it is not added to the true signal. Classical techniques derived from the assumption additive noise with Gaussian distribution may lead to suboptimal procedures, or to the complete failure of the processing and analysis of the data.

Several models have been proposed in the literature to cope with this departure from classical hypothesis, being the K and \mathcal{G}_A^0 distributions the more successful ones (see [13] for an introduction to the subject of SAR image processing and analysis.) These are parametric models, so inference is a crucial step in every procedure. In many applications inference based on sample moments is used but, whenever possible, maximum likelihood estimators are preferred due to their nice asymptotic properties.

Since the family of \mathcal{G}_A^0 laws is regarded as an universal model for speckled imagery, this work concentrates on maximum likelihood inference of the parameters of this distribution. The literature reports severe numerical problems when estimating these parameters. The solution commonly proposed consists of using big samples, in spite of being small samples desirable for minute feature analysis and for techniques that do not introduce unacceptable blurring.

This paper presents the performance analysis of several classical techniques for maximum likelihood parameter estimation in the \mathcal{G}_A^0 model, showing that none of them is reliable for practical applications. A proposal based on alternated optimization of the reduced log-likelihood is then made and assessed with real and simulated data. The computational platform is Ox, well known for its numerical soundness.

The rest of the paper is organized as follows: Section 2 presents the main properties of the \mathcal{G}_A^0 model, our main object of interest. Section 3 recalls the main algorithms involved in maximum likelihood inference for the \mathcal{G}_A^0 model, with special emphasis on their availability in the Ox platform.

Once verified that these algorithms fail to produce acceptable estimators, section 4 describes and assesses the proposal that overcomes this problem, and applications are discussed in section 5. Conclusions and future research directions are commented in section 6.

2 The Universal Model

As proposed and assessed in [11], the family of \mathcal{G}^0 distributions can be successfully used to describe data contaminated by speckle noise. This family of distributions stems from assuming the following hypothesis about the signal formation in every image coordinate:

1. The observed data (return) can be described by the random variable $Z = XY$, where the independent random variables X and Y describe the ground truth and the speckle noise, respectively.
2. The random variable $X : \Omega \rightarrow \mathbb{R}_+$ follows a square root of reciprocal of gamma law, characterized by the density

$$f_X(x) = \frac{2^{\alpha+1}}{\gamma^\alpha \Gamma(-\alpha)} x^{2\alpha-1} \exp\left(-\frac{\gamma}{2x^2}\right) \mathbb{I}_{\mathbb{R}_+}(x),$$

where $\alpha < 0$ and $\gamma > 0$ and \mathbb{I}_A denotes the indicator function of the set A .

3. When linear detection is used, the random variable Y obeys a square root of gamma distribution, whose density is

$$f_Y(y) = \frac{L^L}{\Gamma(L)} y^{2L-1} \exp(-Ly^2) \mathbb{I}_{\mathbb{R}_+}(y),$$

where $L \geq 1$ is the (equivalent) number of looks, a parameter that can be controlled in the image generation process.

Under these assumptions, the density of Z is given by

$$f_Z(z) = \frac{2L^L \Gamma(L-\alpha)}{\gamma^\alpha \Gamma(L) \Gamma(-\alpha)} \frac{z^{2L-1}}{(\gamma + Lz^2)^{L-\alpha}} \mathbb{I}_{\mathbb{R}_+}(z), \quad (1)$$

where $-\alpha, \gamma > 0$ are the (unknown) parameters and $L \geq 1$ is the number of looks. The main properties of this distribution, denoted $\mathcal{G}_A^0(\alpha, \gamma, L)$, are presented in [7, 10, 11]. In particular,

moments of order r will be useful in this work. They are given by

$$E(Z^r) = \left(\frac{\gamma}{L}\right)^{r/2} \frac{\Gamma(-\alpha - r/2)\Gamma(L + r/2)}{\Gamma(-\alpha)\Gamma(L)} \quad (2)$$

if $\alpha < -r/2$, and are not finite otherwise. The mean and variance of a $\mathcal{G}_A^0(\alpha, \gamma, n)$ distributed random variable can be computed using equation (2), yielding

$$\mu_Z = \sqrt{\frac{\gamma}{n}} \frac{\Gamma(n + \frac{1}{2})\Gamma(-\alpha - \frac{1}{2})}{\Gamma(n)\Gamma(-\alpha)}, \quad (3)$$

$$\sigma_Z^2 = \frac{\gamma [n\Gamma^2(n)(-\alpha - 1)\Gamma^2(-\alpha - 1) - \Gamma^2(n + \frac{1}{2})\Gamma^2(-\alpha - \frac{1}{2})]}{n\Gamma^2(n)\Gamma^2(-\alpha)}, \quad (4)$$

provided $\alpha < -1/2$ and $\alpha < -1$, respectively. As previously said, in many applications estimators for (α, γ) are derived using the analogy method [9] and moment equations. When the first and second moments are used, besides the severe numerical instabilities that often appear, only samples from laws with $\alpha < -1$ can be analyzed.

The dependence of this distribution on the parameter $\alpha < 0$ can be seen in Figure 1. It is noticeable that the bigger α the less symmetric and the heavier-tailed the density is.

[Figure 1 about here.]

If Z_A follows a $\mathcal{G}_A^0(\alpha, \gamma, n)$ distribution, then its cumulative distribution function is given by

$$F_{Z_A}(z) = \frac{n^\alpha \Gamma(n - \alpha) z^{2n}}{\gamma^\alpha \Gamma(n) \Gamma(-\alpha)} H(n, n - \alpha; n + 1; -nz^2/\gamma), \quad (5)$$

with $-\alpha, \gamma, z > 0$ and $n \geq 1$, where

$$H(a, b; c; t) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)t^k}{\Gamma(c+k)k!}.$$

is the hypergeometric function (see [1]).

Equation (5) can also be written as

$$F_{Z_A}(z) = \Upsilon_{2n, -2\alpha} \left(\frac{-\alpha z^2}{\gamma} \right), \quad (6)$$

again with $-\alpha, \gamma, z > 0$, $n \geq 1$, where $\Upsilon_{2n, -2\alpha}$ is the cumulative distribution function of the Snedecor's \mathcal{F} law with $2n$ and -2α degrees of freedom. This form is useful for the following

reasons:

1. The cumulative distribution function of a $\mathcal{G}_A^0(\alpha, \gamma, n)$ random variable, needed to perform the Kolmogorov-Smirnov test and to work with order statistics, can be computed using the T_γ function, available in most statistical software platforms, and relation (6).
2. Since the function T_γ^{-1} is also available in most statistical platforms, outcomes from the $\mathcal{G}_A^0(\alpha, \gamma, n)$ law can be obtained using this inverse function [5, 15] and returning outcomes of the random variable $Z = (-\gamma T_{2n-2\alpha}^{-1}(U)/\alpha)^{1/2}$, with U uniformly distributed on $(0, 1)$. This was the method employed in the forthcoming Monte Carlo experience.

A crucial feature of the distribution characterized by eq. (1) is that its parameters are interpretable: γ is a scale parameter, while α is related to the roughness of the target. Small values of α (say $\alpha < -10$) describe smooth regions as, for instance, crops and burnt fields. When α is close to zero (say $\alpha > -5$) the observed target is extremely rough, as is the case of urban spots. Intermediate situations ($-10 < \alpha < -5$) are usually related to rough areas as, for instance, forests. The equivalent number of looks L is known beforehand or is estimated for the whole image using extended targets, i.e., very large samples. Note that estimating (α, γ) amounts to making inference about the unobservable ground truth X .

Figure 2 shows the densities of two distributions with the same mean and variance: the $\mathcal{G}_A^0(-2.5, 7.0686/\pi, 1)$ and the Gaussian distribution $\mathcal{N}(1, 4(1.1781 - \pi/4)/\pi)$ in semilogarithmic scale, along with their mean value. The different decays of their tails in the logarithmic plot are evident: the former behaves logarithmically, while the latter decays quadratically. This behavior ensures the ability of the \mathcal{G}_A^0 distribution to model data with extreme variability but, at the same time, the slow decay is prone to producing problems when performing parameter estimation.

[Figure 2 about here.]

Since systems that employ coherent illumination are used to survey inaccessible and/or unobservable regions (the surface of Venus, the interior of the human body, the bottom of the sea, areas under cloud cover, etc.), it is of paramount importance to be able to make reliable inference about the kind of target under analysis.

This inference can be performed through the estimation of the parameter $(\alpha, \gamma) \in \Theta = (\mathbb{R}_- \times \mathbb{R}_+)$ from samples $\mathbf{z} = (z_1, \dots, z_n)$ taken from homogenous areas in order to grant that the observations come from identically distributed random variables. The bigger the sample, in principle, the more accurate the estimation but, also, the bigger the chance of including spurious

observations. Also, if the goal is to perform some kind of image processing or enhancement [3, 8], as is the case of filtering based on distributional properties, large samples are obtained using big windows that usually cause heavy blurring. Inference with small samples is gaining attention in the specialized literature (see [16] for instance), and reliable inference using small samples is the core contribution of this work.

Usual inference techniques include methods based on the analogy principle [9] (moment and order statistics estimators being the most popular members of this class) and maximum likelihood [2]. Moment estimators are favored in applications, since they are easy to derive and are, usually, computationally attractive. An estimator based on the median and on the first moment was successfully used in [3] as the starting point for computing maximum likelihood (ML therein) estimates.

Given the sample \mathbf{z} , and assuming that these observations are outcomes of independent and identically distributed random variables with common distribution $\mathcal{D}(\theta)$, with $\theta \in \Theta \subset \mathbb{R}^p$, $p \geq 1$, a ML estimator of θ is given by

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; \mathbf{z}), \quad (7)$$

where L is the likelihood of the sample \mathbf{z} under the parameter θ . Under very mild conditions it is equivalent (and many times easier) to work with the reduced log-likelihood $\ell(\theta; \mathbf{z}) \propto \ln L(\theta; \mathbf{z})$, where all the terms that do not depend on θ have been discarded.

Though direct maximization of equation (7) is possible (either analytically or using numerical tools), and many times desirable, quite often one finds ML estimates by solving the system of (usually non-linear) p equations given by

$$\nabla \ell(\hat{\theta}) = 0. \quad (8)$$

This system is referred to as *likelihood equations*. The choice between solving either equation (7) or equation (8) heavily relies on computational issues: availability of reliable algorithms, computational effort required to implement and/or to obtain the solution, etc. The MLE, in general, has no explicit solution.

In our case the likelihood function is $L((\alpha, \gamma); \mathbf{z}) = \prod_{i=1}^n f_{\mathcal{Z}}(z_i)$, with $f_{\mathcal{Z}}$ given in eq. (1). Therefore, the reduced log-likelihood can be written as

$$\ell((\alpha, \gamma); \mathbf{z}) = \ln \frac{L - \alpha}{\gamma^\alpha \Gamma(-\alpha)} - \frac{L - \alpha}{n} \sum_{i=1}^n \ln(\gamma + Lz_i^2). \quad (9)$$

The system given by eq. (8) is, in our case,

$$\begin{cases} n[\Psi(-\hat{\alpha}) - \Psi(L - \hat{\alpha})] + \sum_{i=1}^n \ln\left(\frac{\hat{\gamma} + Lz_i^2}{\hat{\gamma}}\right) = 0, \\ -\frac{n\hat{\alpha}}{\hat{\gamma}} - (L - \alpha) \sum_{i=1}^n \ln\left(\frac{1}{\hat{\gamma} + Lz_i^2}\right) = 0, \end{cases} \quad (10)$$

where $\Psi(\tau) = d \ln \Gamma(\tau) / d\tau$ is the digamma function. No explicit solution for this system is available in general and, therefore, numerical routines have to be used.

Figure 3 shows a typical situation. A sample from the $\mathcal{G}_A^0(-8, \gamma^*, 3)$ of size $n = 9$ was generated, and the log-likelihood function of this sample is shown. The parameter γ^* is chosen such that the expected value is one. It is noticeable that finding the maximum of this function (provided it exists) is not an easy task due to the almost flat area it presents around the candidates. The ML estimates for this sample were $(\hat{\alpha}, \hat{\gamma}) = (-1.84, 1.44)$.

[Figure 3 about here.]

3 Algorithms for Inference

The routines here reported were used as provided by the Ox platform, a robust, fast, free and reliable matrix-oriented language with excellent numerical capabilities. This platform is available for a variety of operational systems at [6].

Two categories of routines were tested: those devoted to direct maximization (or minimization), referred to as *optimization procedures*, and those that look for the solution of systems of equations. In the first category the Simplex Downhill, the Newton-Raphson and the Broyden, Fletcher, Goldfarb and Shanno (generally referred to as 'the BFGS method' [14]) algorithms were used to maximize eq. (9). In the second category the Broyden algorithm was used to find the roots of eq. (10).

These routines impose different requirements for their use. Some require or accept the derivatives of the function to be maximized, while others try to perform their tasks with mere evaluations of the target function. In our case, BFGS plays a central role, and it was used specifying the derivatives of the objective function.

Since the main goal of this work is to find suitable solutions, all routines were tested following the guidelines provided with the Ox platform: a variety of tuning parameters, starting points, steps and convergence criteria were tested. The results confirmed what is commented in the literature, namely, that inference for the \mathcal{G}_A^0 law requires *huge* samples in order to converge, let alone to produce sensible estimates.

An extensive analysis was performed in a variety of situations, namely using samples of size $n \in \{9, 25, 49, 81, 121\}$, roughness parameters $\alpha \in \{-1, -3, -5, -15\}$ and number of looks $L \in \{1, 2, 3, 8\}$. The scale parameter γ was chosen to yield unitary mean, so it was set to

$$\gamma^* = L \left(\frac{\Gamma(L)\Gamma(-\alpha)}{\Gamma(L+1/2)\Gamma(-\alpha-1/2)} \right)^2.$$

These sample sizes reflect the fact that most image processing techniques employ estimation in squared windows of even side and, therefore, samples are of size $n = s^2$, where s is the side of the window. Windows of sides 3, 5, 7, 9 and 11 are commonly used.

The roughness parameter describes regions with a wide range of smoothness, as discussed in section 2. The number of looks also reflects situations of practical interest, ranging from raw images ($L = 1$) to smoothed out data $L = 8$. It is convenient to note here that the bigger the number of looks the smoother the image, at the expense of loss of spatial resolution.

One thousand replications were performed for each of these eighty situations, generating samples with the specified parameters and, then, applying the four algorithms. Success (convergence to a point and numerical evidence of convergence to either a maximum or a root) or failure to converge was recorded, and specific situations of both outcomes were traced out.

Table 1 shows the percentage of times (in 1,000 independent trials) that the BFGS algorithm failed to converge in each of the eighty aforementioned situations. The bigger the sample size the better the performance, and the smoother the target the worse the convergence rate. In the worst case almost sixty percent of the samples were left unanalyzed, i.e., no sensible estimate was obtained. Similar (mostly worse) behaviour is observed in the other algorithms, and it is noteworthy that all of them were fine-tuned for the problem at hand.

The overall behaviour of these algorithms falls into one of three situations, namely

1. All of them converge to the same (sensible) estimate.
2. All of them converge, but not to the same value.
3. At least one algorithm fails to converge.

In order to illustrate these behaviors, two samples from the \mathcal{G}_A^0 were chosen, one leading to situation 1 above (denoted \mathbf{z}_1), and other to situation 2 (denoted \mathbf{z}_2). For each sample the likelihood function was computed and, in order to visualize and analyze the behavior of the algorithms, level curves of the likelihood and of the maximum likelihood equations were studied.

Situation 1 is illustrated in Figure 4, where it is noticeable that the point of convergence of the Broyden algorithm (denote as “*”) is in the interior of the highest level curve. This point coincides with the intersection of the curves corresponding to $\partial\ell/\partial\alpha = \partial\ell/\partial\gamma = 0$ and, regardless the precision of the estimation procedure, is an acceptable estimate.

[Figure 4 about here.]

Similarly, situation 2 is illustrated in Figure 5. In this case the point to which the Broyden algorithm converges is outside the highest level curve and, thus, does not correspond to the maximum of the likelihood function.

[Figure 5 about here.]

The Broyden algorithm *seemed* to have the best performance, since it reported convergence in many situations. But when at least two of the other algorithms converged, most of the time they did it to the same point whereas Broyden stopped very far from it. When checking the value of the likelihood in the solutions, the one computed by Broyden was orders of times smaller than the one found by maximization techniques. For this reason, though Broyden allegedly outperformed optimization procedures in terms of convergence, it was considered unreliable for this application.

[Table 1 about here.]

This behaviour motivated the proposal of an algorithm able to converge to sensible estimates. This is done in the next section.

4 Proposal: alternated optimization

Since simultaneous optimization is not reliable enough, an analysis of the marginal functions to be maximized was conducted for a variety of situations. In all these situations it was checked that, whereas the reduced log-likelihood showed flat regions, where simultaneous optimization may get lost or stuck, these surfaces could be sliced in order to yield better-behaved functions. This motivated the proposal of an *alternated* algorithm that consists of writing two equations out of eq. (9): one depending on α , given γ fixed, and the other depending on γ , given a fixed α . Provided a starting point for γ , say $\hat{\gamma}(0)$, one maximizes the first equation on α to find $\hat{\alpha}(0)$.

One can now use this crude estimate of α , solve again the first equation on γ and continue until evidence of convergence is achieved. Formally, the equations to be maximized are

$$\ell_1(\alpha; \gamma(j), \mathbf{z}) = \ln \frac{\Gamma(L - \alpha)}{(\gamma(j))^\alpha \Gamma(-\alpha)} + \frac{\alpha}{n} \sum_{i=1}^n \ln(\gamma(j) + Lz_i^2), \quad (11)$$

$$\ell_2(\gamma; \alpha(j), \mathbf{z}) = -\alpha(j) \left(\ln \gamma + \frac{1}{n} \sum_{i=1}^n \ln(\gamma + Lz_i^2) \right). \quad (12)$$

Algorithm 4.1 *Alternated optimization for parameter estimation.*

1. Fix the smallest acceptable variation to proceed (typically $\epsilon = 10^{-4}$) and the maximum number of iterations (typically $M = 10^3$).

2. Compute an initial estimate of γ , for example

$$\hat{\gamma}(0) = L \left(\frac{\widehat{m}_1 \Gamma(L)}{\Gamma(L + 1/2)} \right)^2, \quad (13)$$

where $\widehat{m}_1 = n^{-1} \sum_{i=1}^n z_i$ is the first sample moment.

3. Set the values needed to execute step (4c) for the first time $\epsilon = 10^3$ and $\widehat{\alpha}(0) = -10^6$, and start the counter $j = 1$

4. While $\epsilon \geq \epsilon \vee j \leq M$ do

(a) Find $\widehat{\alpha}(j) = \arg \max_{\alpha \in \mathbb{R}_-} \ell_1(\alpha; \gamma(j-1), \mathbf{z})$ given in eq. (11).

(b) Find $\widehat{\gamma}(j) = \arg \max_{\gamma \in \mathbb{R}_+} \ell_2(\gamma; \alpha(j), \mathbf{z})$ given in eq. (12).

(c) Compute $\epsilon = \left| \frac{\alpha(j+1) - \alpha(j)}{\alpha(j+1)} \right| + \left| \frac{\gamma(j+1) - \gamma(j)}{\gamma(j+1)} \right|$ the absolute value of the relative inter-iteration variation.

(d) Update the counter $j = j + 1$.

5. If $\epsilon > \epsilon$ return anything with a message of error, else return $(\widehat{\alpha}(j-1), \widehat{\gamma}(j-1))$ and a message of success.

Equation (13) is derived using $r = 1$ and discarding the dependence of α on eq. (2). In this manner, it is a crude estimator of γ based on the first sample moment \widehat{m}_1 . Other starting points were checked, and the effect on the algorithm convergence was negligible.

It was chosen to work with the BFGS algorithm in steps (4a) and (4b) since, for the considered univariate equations, it outperformed the other methods in terms of speed and convergence. The BFGS is generally regarded as the best performing method [12] for multivariate non-linear optimization. In our case the explicit analytical derivatives of the objective function were provided, a desirable information whenever available.

This alternated algorithm can be easily generalized to obtain parameters with as many components as desired, and its implementation in any computational platform is immediate, provided reliable univariate optimization routines.

Using this algorithm in the same 80,000 samples analyzed in Table 1, in only six of them there was no convergence; in these cases the problem was with step (4b), i.e., when trying to find an estimate for the scale parameter. This represents a notorious improvement with respect to classical algorithms.

5 Application

Using Algorithm 4.1 it was possible to conduct a Monte Carlo experience in order to evaluate the bias and mean square error of the MLE for a variety of situations that had to be left unexplored when using classical procedures. These results on the bias of $\hat{\alpha}$ are shown in Figure 6. These values can be huge, confirming previous results. Efforts to reduce this undesirable behavior of ML estimators are reported in [4].

[Figure 6 about here.]

Two applications were devised to show the applicability of the alternated algorithm: one with simulated data and the other with a real SAR image. The former consists of generating samples from the $\mathcal{G}_d^0(\alpha, \gamma^*, L)$ law, for fixed L .

Two-hundred and fifty samples of size $n = 121$ were generated, being fifty from the $\mathcal{G}_d^0(-5, \gamma^*, L)$, fifty from the $\mathcal{G}_d^0(-1, \gamma^*, L)$, fifty from the $\mathcal{G}_d^0(-15, \gamma^*, L)$ and the remaining 100 samples from the $\mathcal{G}_d^0(\alpha_j, \gamma^*, L)$ where $\alpha_j = 0.14j - 15$ and $1 \leq j \leq 100$ is the integer index.

For each of these samples two algorithms were employed to obtain the MLE, namely the BFGS and alternated algorithms. The procedure was repeated for each sample, but using 81, 49, 25 and 9 observations out of the complete data set.

In every situation the alternated algorithm achieved convergence, and the same does not hold for the BFGS algorithm. The percentage of situations for which BFGS did not converge is presented in Table 2. Again, the classical procedure is unreliable.

[Table 2 about here.]

Figure 7 shows, for $n = 25$, the true value of $-\alpha$ (in semilogarithmic scale) along with the estimates: "x" for the alternated algorithm and "o" for the one obtained with the BFGS procedure. The missing circles correspond to those situations where BFGS failed to converge (roughly 20% of the samples). It can be checked that when they both converge, they converge to similar values, and that there are many situations for which BFGS was unable to return an estimate. Similar behaviour is exhibited for other sample sizes, the smaller the sample the less reliable BFGS.

[Figure 7 about here.]

Figure 8 shows a SAR image obtained by the sensor E-SAR, managed by the German Aerospace Center DLR. This is an airborne sensor with polarimetric and high spatial resolution capabilities. The scene was taken over the surroundings of München, and typical classes are marked as "U" (Urban), "F" (Forest) and "C" (Crops). An hypothesized flight track is marked with the NW-SE white arrow, where small samples are being collected at every passage point.

[Figure 8 about here.]

One thousand samples were collected, and they were divided into four groups of the same size for the sake of simplicity. The analysis of these on-flight samples was performed with both the BFGS and the alternated algorithms. The latter always returned estimates while the number of samples for which the former failed to converge is reported in Table 3. Even with windows of size 11 almost a third of the coordinates would be left unanalyzed by the classical algorithm.

[Table 3 about here.]

Figures 9, 10, 11 and 12 show the values of $\hat{\alpha}$ in two-hundred and fifty sites using $n = 121, 49, 25$ and 9 observations, corresponding to groups 1, 2, 3 and 4, respectively. It can be seen that the bigger the window the smoother the analysis, leading to the conclusion that most sites correspond to heterogeneous or extremely heterogeneous spots (since $\hat{\alpha} > -7$). When the window is smaller, more heterogeneous areas appear ($\hat{\alpha} < -10$). The sensed area is suburban, and typical spots consist of scattered houses and small buildings (extremely heterogeneous return) with trees and gardens in between, where SAR will return heterogeneous and homogeneous clutter, respectively. The only exception is Group 3 (Figure 11), for which the estimated roughness at all window sizes is consistent.

The ground resolution of this sensor can be of less than one meter, so minute features of about two meters of side can be detected with the use of the alternated algorithm and the \mathcal{G}_A^0 model.

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

6 Conclusions and future work

The numerical problems that arise when estimating the parameters of the universal model for speckled data using maximum likelihood are alleviated by the use of an alternated optimization procedure.

An analysis of the performance of MLE estimators for the \mathcal{C}_A^0 distribution in the presence of small samples was conducted with the alternated algorithm. This study would not be possible with conventional techniques, since they fail to converge and/or to provide sensible estimates in as many as 60% of the situations.

The same technique was employed to analyze simulated and real data. In the latter sound information about minute features in the ground was retrieved in a SAR image.

This alternated technique is being employed to compute ML estimates of the parameters of polarimetric distributions for SAR data. These distributions are indexed by matrices of complex values, and their computation is prone to severe numerical instabilities.

Acknowledgements

The authors are grateful to CNPq for the partial support to this research.

References

- [1] M. Abramowitz and I. Stein. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.
- [2] P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*, volume 1. Prentice-Hall, NJ, 2 edition, 2001.

- [3] O. H. Bustos, M. M. Lucini, and A. C. Frery. M-estimators of roughness and scale for GA0-modelled SAR imagery. *Applied Signal Processing*, 2002(1):105–114, Jan. 2002.
- [4] F. Cribari-Neto, A. C. Frery, and M. F. Silva. Improved estimation of clutter properties in speckled imagery. *Computational Statistics and Data Analysis*, in press.
- [5] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.
- [6] J. A. Doornik. *Ox: An Object-Oriented Matrix Programming Language*. Timberlake Consultants & Oxford, London, 4 edition, 2001. <http://www.nuff.ox.ac.uk/Users/Doornik>.
- [7] A. C. Frery, H.-J. Müller, C. C. F. Yanasse, and S. J. S. Sant’Anna. A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):648–659, May 1997.
- [8] A. C. Frery, S. J. S. Sant’Anna, N. D. A. Mascarenhas, and O. H. Bustos. Robust inference techniques for speckle noise reduction in 1-look amplitude SAR images. *Applied Signal Processing*, 4:61–76, 1997.
- [9] C. F. Manski. *Analog Estimation Methods in Econometrics*, volume 39 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York, 1988. Available at <http://elsa.berkeley.edu/books/analog.html>.
- [10] M. Mejail, J. Jacobo-Berles, A. C. Frery, and O. H. Bustos. Parametric roughness estimation in amplitude SAR images under the multiplicative model. *Revista de Teledetección*, 13:37–49, 2000.
- [11] M. E. Mejail, A. C. Frery, J. Jacobo-Berles, and O. H. Bustos. Approximation of distributions for SAR images: proposal, evaluation and practical consequences. *Latin American Applied Research*, 31:83–92, 2001.
- [12] R. C. Mittelhammer, G. G. Judge, and D. J. Miller. *Econometric Foundations*. Cambridge University Press, New York, 2000.
- [13] C. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. Artech House, Boston, 1998.
- [14] W. H. Press, B. P. Flannery, S. A. Teulosky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University, 2 edition, 1992.

- [15] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York, 2000.
- [16] P. J. Rousseeuw and S. Verboven. Robust estimation in very small samples. *Computational Statistics and Data Analysis*, in press.

List of Figures

1	Densities of the $\mathcal{G}_A^0(\alpha, 10, 1)$ distribution, with $\alpha \in \{-5, -2, -1\}$ (solid line, dash-point, dashes)	16
2	Densities of the $\mathcal{G}_A^0(-2.5, 7.0686/\pi, 1)$ (solid line) and the $\mathcal{N}(1, 4(1.1781 - \pi/4)/\pi)$ (dashes) distributions in semilogarithmic scale.	17
3	Log-likelihood function of a sample of size $n = 9$ of the $\mathcal{G}_A^0(-8, \gamma^*, 3)$ distribution.	18
4	Log-likelihood function for \mathbf{z}_1 : contour plots (solid lines), $\partial\ell/\partial\alpha$ (black dashes) and $\partial\ell/\partial\gamma$ (grey dashes).	19
5	Log-likelihood function for \mathbf{z}_2 : contour plots (solid lines), $\partial\ell/\partial\alpha$ (black dashes) and $\partial\ell/\partial\gamma$ (grey dashes).	20
6	Estimated bias of the MLE estimator of α for one look.	21
7	MLE estimates of α with $n = 25$ and $L = 1$	22
8	E-SAR synthetic aperture image with $L = 1$	23
9	Estimates of α in 250 sites with different window sizes: Group 1.	24
10	Estimates of α in 250 sites with different window sizes: Group 2.	25
11	Estimates of α in 250 sites with different window sizes: Group 3.	26
12	Estimates of α in 250 sites with different window sizes: Group 4.	27

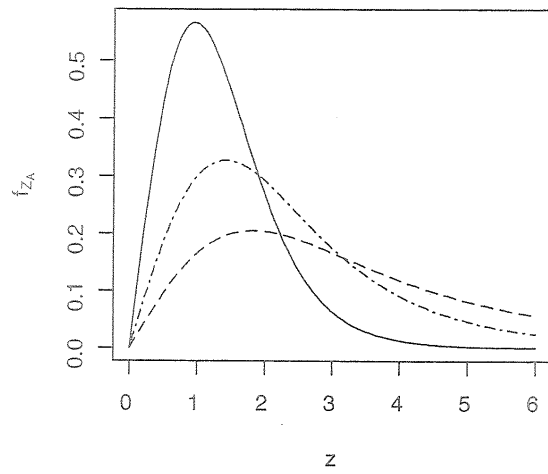


Figure 1: Densities of the $\mathcal{G}_4^{\alpha}(\alpha, 10, 1)$ distribution, with $\alpha \in \{-5, -2, -1\}$ (solid line, dash-point, dashes).

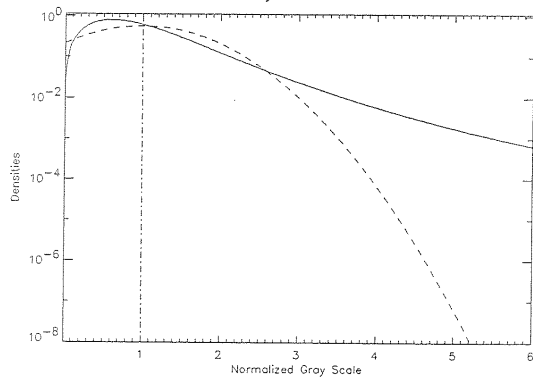


Figure 2: Densities of the $G_A^0(-2.5, 7.0686/\pi, 1)$ (solid line) and the $\mathcal{N}(1, 4(1.1781 - \pi/4)/\pi)$ (dashes) distributions in semilogarithmic scale.

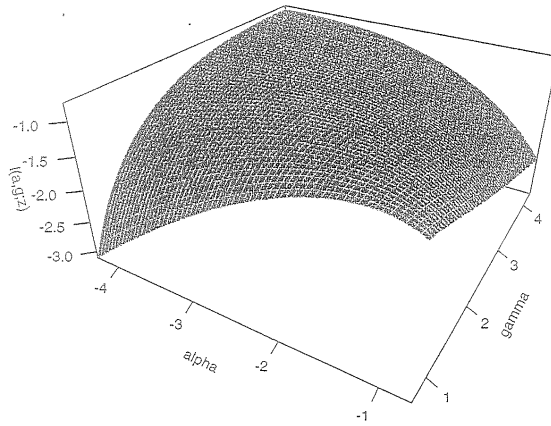


Figure 3: Log-likelihood function of a sample of size $n = 9$ of the $G_{\alpha}^0(-8, \gamma^*, 3)$ distribution.

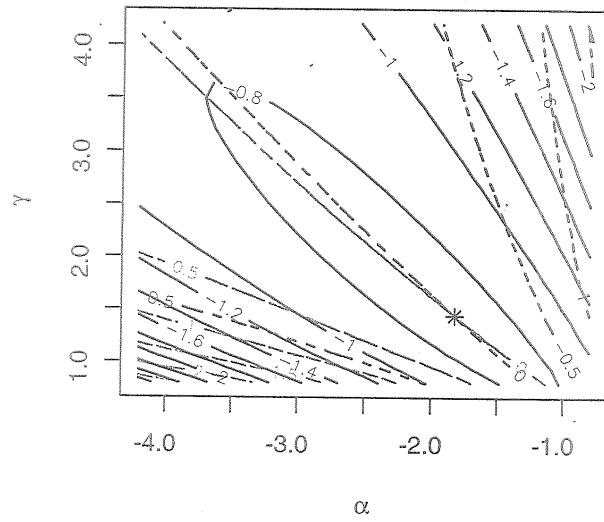


Figure 4: Log-likelihood function for z_1 : contour plots (solid lines), $\partial\ell/\partial\alpha$ (black dashes) and $\partial\ell/\partial\gamma$ (grey dashes).

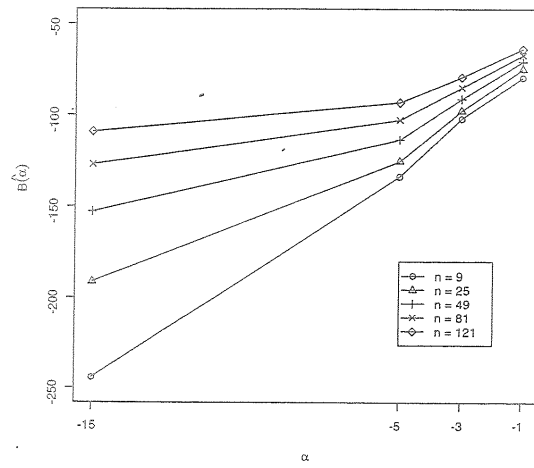


Figure 6: Estimated bias of the MLE estimator of α for one look.

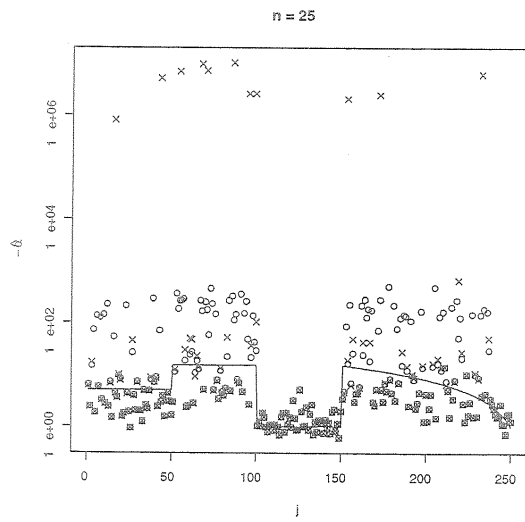


Figure 7: MLE estimates of α with $n = 25$ and $L = 1$.

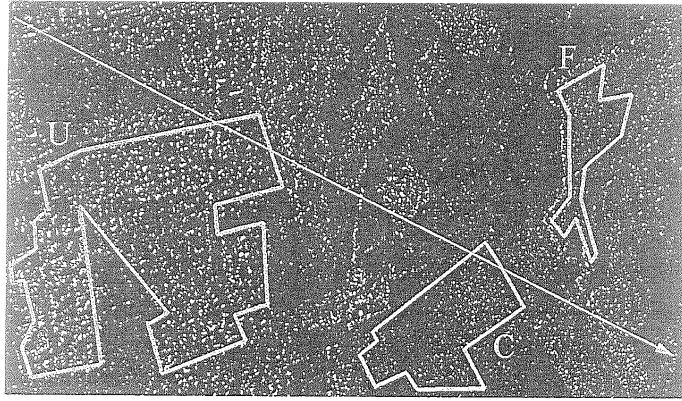


Figure 8: E-SAR synthetic aperture image with $L = 1$.

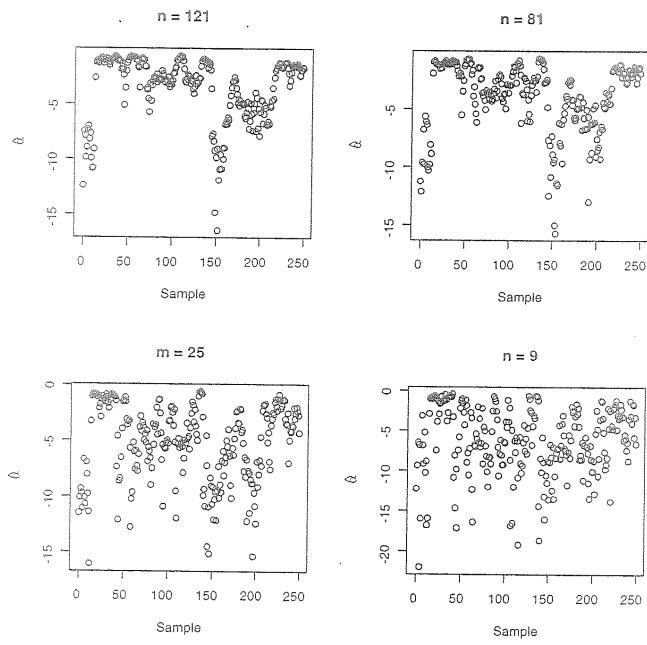


Figure 9: Estimates of α in 250 sites with different window sizes: Group 1.

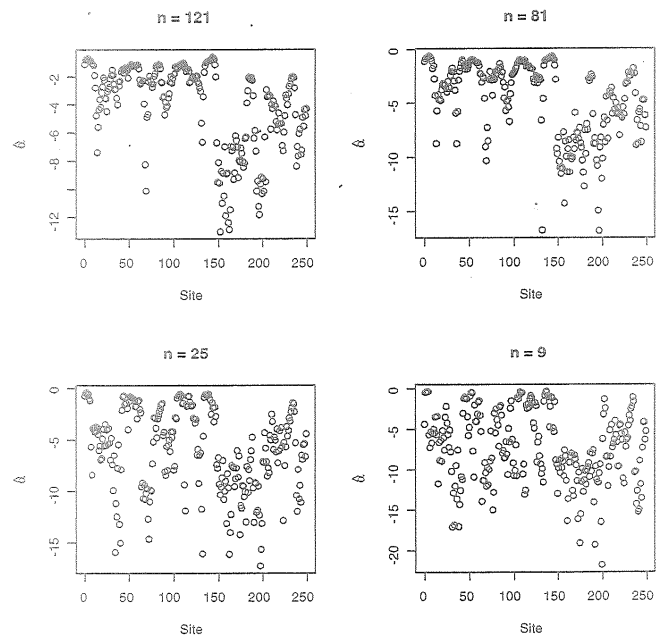


Figure 10: Estimates of α in 250 sites with different window sizes: Group 2.

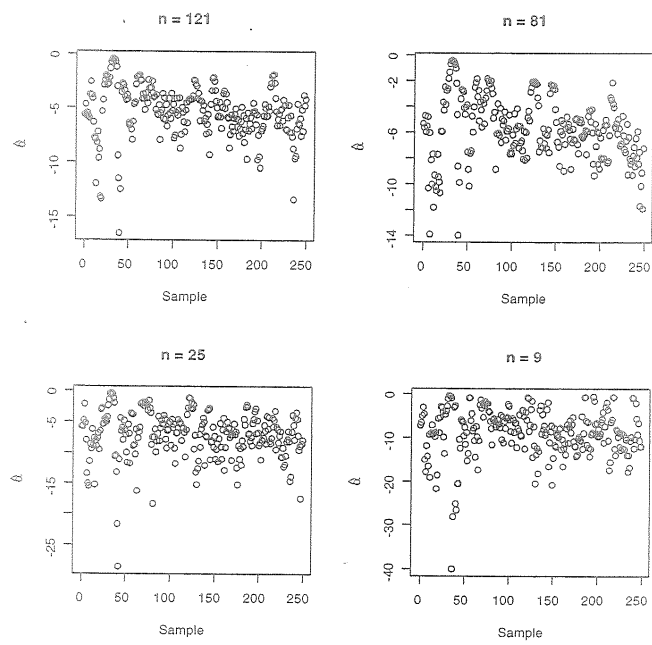


Figure 11: Estimates of α in 250 sites with different window sizes: Group 3.

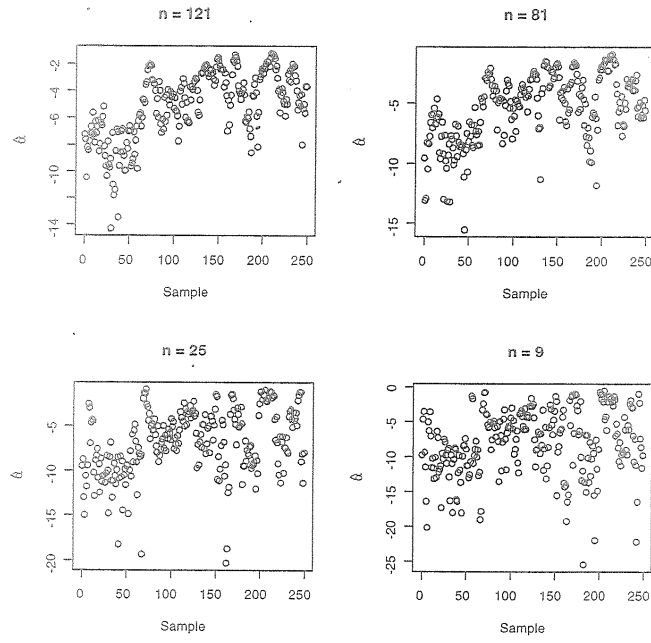


Figure 12: Estimates of α in 250 sites with different window sizes: Group 4.

List of Tables

1	Percentage of situations for which BFGS fails to converge in 1,000 replications. . . .	29
2	Situations where BFGS failed to converge.	30
3	Percentage of samples for which BFGS failed to converge in the four groups of real data using samples of size n	31

L	α	n				
		9	25	49	81	121
1	-15	59.9	48.2	36.2	27.8	25.2
	-5	52.6	30.1	14.5	8.6	3.9
	-3	42.3	19.1	6.1	1.5	0.4
	-1	17.6	1.0	0.1	0.0	0.0
2	-15	51.9	35.4	25.8	16.2	11.4
	-5	37.7	13.5	5.4	1.7	0.2
	-3	25.0	5.4	0.4	0.0	0.0
	-1	4.6	0.0	0.0	0.0	0.0
3	-15	46.5	28.7	16.6	9.9	7.1
	-5	28.1	7.9	1.4	0.1	0.0
	-3	17.4	2.3	0.0	0.0	0.0
	-1	2.1	0.0	0.0	0.0	0.0
8	-15	31.2	9.1	2.3	0.8	0.2
	-5	8.2	0.3	0.0	0.0	0.0
	-3	2.9	0.0	0.0	0.0	0.0
	-1	0.1	0.0	0.0	0.0	0.0

Table 1: Percentage of situations for which BFGS fails to converge in 1,000 replications.

n	121	81	49	25	9
%	1.6	4.8	10.8	19.2	41.2

Table 2: Situations where BFGS failed to converge.

n	Group			
	G1	G2	G3	G4
121	20.8	21.2	39.2	32.0
81	22.4	28.8	42.8	36.4
49	32.4	34.8	53.6	48.0
25	42.0	46.4	55.6	54.0
9	52.4	65.4	69.2	65.6

Table 3: Percentage of samples for which BFGS failed to converge in the four groups of real data using samples of size n .