Maximum Likelihood estimation of Rician distribution parameters

Jan Sijbers, Arnold J. den Dekker, Paul Scheunders and Dirk Van Dyck

***Abstract*— The problem of parameter estimation from Rician distributed data (e.g., magnitude Magnetic Resonance images) is addressed. The properties of conventional estimation methods are discussed and compared to Maximum Likelihood estimation which is known to yield optimal results asymptotically. In contrast to previously proposed methods, Maximum Likelihood estimation is demonstrated to be unbiased for high signal-to-noise ratio (SNR) and to yield physical relevant results for low SNR.**

***Index Terms*— Maximum Likelihood, Rician distribution, Parameter estimation, MR imaging**

I. INTRODUCTION

In Magnetic Resonance Imaging (MRI), the acquired complex valued data are corrupted by noise that is typically well described by a Gaussian probability density function (PDF) [1]. In case the MR data are acquired on a uniform Cartesian grid in K-space, after Fourier reconstruction, the real and imaginary data are still polluted by Gaussian noise. Although all information is contained in the real and imaginary images, it is common practice to work with magnitude and phase images instead as they have more physical meaning (proton density, flow, etc.). However, computation of a magnitude image is a nonlinear operation in which the Gaussian PDF of the pixels is transformed into a Rician PDF [2], [3]. In addition, Rician distributed data do not solely occur in conventional magnitude reconstructed images, they are also found in MR angiography imaging [4].

Knowledge of the data PDF is vital for image processing techniques based on parameter estimation such as, e.g., image restoration. These techniques usually assume the most general type of data PDF, which is Gaussian. Whenever other PDF's come into play, e.g., in magnitude MR images, one still tends to use parameter estimation techniques that are based on Gaussian distributed data [5], [6], [7]. The justification for this is that, when the signal-to-noise ratio (SNR) is high, the actual data PDF is very similar to a Gaussian one.

The authors are within the Department of Physics at the University of Antwerp, Belgium.

With magnitude MR images, the Rician data PDF deviates significantly from a Gaussian PDF when the SNR is low, leading to biased results. To reduce this bias, parameter estimation methods were proposed which exploit the knowledge of the Rician PDF [8], [9], [10], [11] [8],[9],[10],[11]. However, although the proposed estimators do reduce the bias, they are not able to remove it.

In this paper it is shown where the bias appears in the conventional estimation. In addition, a Maximum Likelihood (ML) estimator for Rician distributed data is constructed. The performance of the conventional estimator is compared to that of the ML estimator. The motivation for this is that, it is known that, if there exists an unbiased estimator of which the variance attains the lowest possible value, it is obtained by the ML method.

II. THE RICIAN DISTRIBUTION

If the real and imaginary data, with mean values respectively A_R and A_I, are corrupted by Gaussian, zero mean, stationary noise with standard deviation σ, it is easy to show that the magnitude data will be Rician distributed [12], with PDF:

$$p\left(M_i|A\right) = \frac{M_i}{\sigma^2} e^{-\frac{M_i^2+A^2}{2\sigma^2}} I_0\left(\frac{AM_i}{\sigma^2}\right) u\left(M_i\right) \qquad (1)$$

I_0 is the modified zeroth order Bessel function of the first kind, M_i denotes the ith data point of the magnitude image. The unit step function u is used to indicate that the expression for the PDF of M_i is valid for non-negative values of M_i only. Furthermore, A is given by:

$$A = \sqrt{A_R^2 + A_I^2} \qquad (2)$$

For further discussion, the moments of the Rician PDF are required. The μth moment of the Rician density function is given by:

$$E\left[M^\nu\right] = \int_0^\infty \frac{M^{\nu+1}}{\sigma^2} e^{-\frac{M^2+A^2}{2\sigma^2}} I_0\left(\frac{AM}{\sigma^2}\right) dM \qquad (3)$$

where $E[.]$ is the expectation operator. The previous equation can be analytically expressed as a function of the confluent hypergeometric function ${}_1F_1$:

$$E\left[M^{\nu}\right] = \left(2\sigma^2\right)^{\nu/2} \Gamma\left(1+\frac{\nu}{2}\right) {}_1F_1\left[-\frac{\nu}{2};1;-\frac{A^2}{2\sigma^2}\right] \quad (4)$$

The even moments of the Rician distribution (i.e., when ν is even) are simple polynomials. E.g:

$$E\left[M^2\right] = A^2 + 2\sigma^2 \quad (5)$$

$$E\left[M^4\right] = A^4 + 8\sigma^2 A^2 + 8\sigma^4 \quad (6)$$

III. Parameter estimation

Given the Rician distribution and its moments, both the conventional approach (section III-A) and the Maximum Likelihood approach (section III-B) to estimate a locally constant signal A from magnitude data points $\{M_i\}$, are discussed.

A. Conventional approach

1) Conventional estimator: Commonly, Eq. (5) is exploited for estimation of the underlying signal A. Thereby, $E\left[M^2\right]$ is estimated from a simple local spatial average [9],[10],[13],[14]:

$$\widehat{E[M^2]} = \langle M^2 \rangle = \frac{1}{N}\sum_{i=1}^{N} M_i^2 \quad (7)$$

Note that this estimator is unbiased since $E[\langle M^2\rangle] = E[M^2]$. Consequently, an unbiased estimator of A^2 is given by:

$$\widehat{A_c^2} = \langle M^2 \rangle - 2\sigma^2 \quad (8)$$

Taking the square root of Eq. (8) gives the conventional estimator of A [9],[10],[13],[14]:

$$\hat{A}_c = \sqrt{\langle M^2 \rangle - 2\sigma^2} \quad (9)$$

2) Discussion: The parameter to be estimated is the signal A. Obviously, A is a priori known to be real valued and non-negative. However, this a priori knowledge has not been incorporated into the conventional estimation procedure. Consequently, the conventional estimator $\hat{A}_c$, given in Eq. (9), may reveal estimates that violate the a priori knowledge and are therefore physically meaningless. This is the case when $\widehat{A_c^2}$ becomes negative. Therefore, $\hat{A}_c$ can not be considered a useful estimator of A if the probability that $\widehat{A_c^2}$ is negative differs from zero significantly. It can be shown that the PDF of $\widehat{A_c^2}$ is a noncentral χ^2 distribution [4], given by:

$$p_{\widehat{A_c^2}}\left(\widehat{A_c^2}\right) = \frac{N}{2\sigma^2}\left(\frac{\widehat{A_c^2}+2\sigma^2}{A^2}\right)^{\frac{N-1}{2}} e^{-N\frac{\widehat{A_c^2}+2\sigma^2+A^2}{2\sigma^2}} \times I_{N-1}\left(\frac{NA\sqrt{\widehat{A_c^2}+2\sigma^2}}{\sigma^2}\right) u\left(\widehat{A_c^2}+2\sigma^2\right) \quad (10)$$

In Fig. 1, $\Pr[\widehat{A_c^2}<0]$ is drawn as a function of the local SNR for several values of N, where the local SNR is defined as:

$$\text{SNR} = \frac{A}{\sigma} \quad (11)$$

Fig. 1 $\Pr\left[\widehat{A_c^2}<0\right]$ as a function of the SNR for various N.

From the figure one can conclude that for low SNR $\hat{A}_c$ cannot be a valid estimator of A unless a large amount of data points is used for the estimation. Therefore, in practice $\hat{A}_c$ will only be a useful estimator if the local SNR is high.

However, even if the condition of high SNR is met, the use of $\hat{A}_c$ as an estimator of A should still not be recommended since the results obtained are biased because of the square root operation in Eq. (9). This becomes more clear when $E[\hat{A}_c]$ is expanded about the unbiased value A, yielding:

$$E\left[\hat{A}_c\right] \approx A\left(1-\frac{\sigma^2}{2NA^2}\right) \quad (12)$$

Eq. (12) is valid for high SNR. The bias appears in the second term of Eq. (12). Note that it decreases with increasing SNR and increasing number of data points N.

B. Maximum Likelihood estimation

In this section the ML method is introduced into the problem of the estimation of Rician distribution parameters. The ML estimator exploits the a priori knowledge of the data statistics in an optimal way. Concerning the accuracy and precision of the ML estimator, it is known that, under very general conditions, the ML estimator is consistent and asymptotically most precise [15]. In addition, it is known that if the number of data points increases, the distribution of the ML estimator approaches the normal distribution with mean A and variance equal to the so-called Minimum Variance Bound (MVB), which is a lower bound on the variance of any unbiased estimator [16]. Furthermore, it is known that if there exists an unbiased estimator having the MVB as variance, it is the ML estimator [15].

1) ML estimator: The joint PDF of a sample of N independent observations $\{M_i\}$ is called the likelihood function of the sample, and is written as:

$$L = \prod_{i=1}^{N} p\left(M_i|A\right) \quad (13)$$

where $p(M_i|A)$ is given in Eq. (1). The ML estimator can be constructed directly from the likelihood function L. Once the observations have been made and numbers can be substituted for $\{M_i\}$, L is a function of the unknown parameter A only. The ML estimator of A is now defined as the estimator maximizing L, or equivalently $\log L$, as a function of A. Hence, using Eq. (1) it follows that:

$$\log L = \log\prod_{i=1}^{N}\frac{M_i}{\sigma^2}e^{-\frac{M_i^2+A^2}{2\sigma^2}} I_0\left(\frac{AM_i}{\sigma^2}\right) \quad (14)$$

or only as a function of A:

$$\log L \sim \sum_{i=1}^{N}\log I_0\left(\frac{AM_i}{\sigma^2}\right) - \sum_{i=1}^{N}\frac{A^2}{2\sigma^2} \quad (15)$$

Since I_0 is symmetric about $A = 0$, L as well as $\log L$ are also symmetric about $A = 0$. The ML estimate is the global maximum of $\log L$:

$$\hat{A}_{ML} = \arg\left\{\max_{A}\left(\log L\right)\right\} \tag{16}$$

2) Discussion: It is not possible to find the maximum of the $\log L$ function directly because the parameter A enters that function in a nontrivial way. Therefore, finding the maximum of the $\log L$ function will in general be an iterative numerical process.

In order to get some insight into the properties of the ML estimator, the structure of the $\log L$ function is now studied. This structure is established by the number and nature of the stationary points of the function. Stationary points are defined as points where the gradient vanishes:

$$\frac{\partial}{\partial A}\log L = 0 \tag{17}$$

Substituting Eq. (15) into Eq. (17) along with some rearrangements yields the condition for the stationary points

$$\hat{A} = \frac{1}{N}\sum_{i=1}^{N} M_i \frac{I_1\left(\frac{\hat{A}M_i}{\sigma^2}\right)}{I_0\left(\frac{\hat{A}M_i}{\sigma^2}\right)} \tag{18}$$

It follows from Eq. (18) that $A = 0$ is a stationary point of $\log L$, independent of the particular data set. The nature of a stationary point is determined by the sign of the second order derivative of the function in that point. From this derivative it follows whether a stationary point is a minimum or a maximum and whether or not it is degenerate. From Eq. (15) the second order derivative of the $\log L$ function can be computed to yield:

$$\frac{\partial^2 \log L}{\partial A^2} = \sum_{i=1}^{N}\frac{M_i^2}{\sigma^4}\left[1 - \frac{\sigma^2}{AM_i}\frac{I_1\left(\frac{AM_i}{\sigma^2}\right)}{I_0\left(\frac{AM_i}{\sigma^2}\right)} - \frac{I_1^2\left(\frac{AM_i}{\sigma^2}\right)}{I_0^2\left(\frac{AM_i}{\sigma^2}\right)}\right] - \frac{N}{\sigma^2} \tag{19}$$

It is then easy to verify that $A = 0$ is a minimum of $\log L$ whenever:

$$\frac{1}{N}\sum_{i=1}^{N} M_i^2 > 2\sigma^2 \tag{20}$$

If this condition is met, the $\log L$ function will have two further stationary points, being maxima.

This can be seen by studying the possible structures of the $\log L$ function using catastrophe theory. Catastrophe theory is concerned with the structural change of a parametric function under influence of its parameters [17]. It tells us that a structural change of the function is always preceded by a degeneracy of one of its stationary points. In order to analyze such a structural change, the parametric function can be replaced by a Taylor expansion in the essential variables about the latter stationary point. The essential variables correspond to the directions in which degeneracy may occur. According to the catastrophe theory the global structure of a parametric function with only one essential variable is completely set by its Taylor expansion up to the degree of which the coefficient cannot vanish under the influence of its parameters. The function studied in this paper is the $\log L$ function as a function of A. Its parameters are the observations. Thus, the structural change of the $\log L$ function under the influence of the observations has to be studied. The only essential variable is the signal parameter A. The stationary point that may become degenerate is the point $A = 0$ (degeneracy occurs whenever (19) becomes equal to zero). If the $\log L$ function is Taylor expanded about the stationary point $A = 0$, we yield:

$$\log(L) = a + \frac{b}{2!}A^2 + \frac{c}{4!}A^4 + O\left(A^6\right) \tag{21}$$

with

$$a = \sum_{i=1}^{N}\frac{M_i^2}{\sigma^4}\left[1 - \frac{1}{2}\frac{\sigma^2}{M_i}\right] - \frac{N}{\sigma^2} \tag{22}$$

$$b = \sum_{i=1}^{N}\frac{M_i^2}{2\sigma^4} - \frac{N}{\sigma^2} \tag{23}$$

$$c = -\frac{3}{8}\sum_{i=1}^{N}\frac{M_i^2}{\sigma^8} \tag{24}$$

and $O(.)$ is the order symbol of Landau. Notice that since the $\log L$ function is symmetric about $A = 0$, the odd terms are absent in Eq. (21). In order to investigate if the expansion up to the quartic term in Eq. (21) is sufficient, it has to be determined if the coefficients may change sign under influence of the observations. It is clear from Eq. (23) that the coefficient b may change sign. The coefficient c, however, will always be negative, independent of the particular set of observations. This means that the expansion (21) is sufficient to describe the possible structures of the $\log L$ function. Consequently, the study of the $\log L$ function as a function of the observations can be replaced by a study of the following quartic Taylor polynomial in the essential variable A:

$$\frac{b}{2!}A^2 + \frac{c}{4!}A^4 \tag{25}$$

where the term a has been omitted since it does not influence the structure. The polynomial (25) is always stationary at $A = 0$. This will be a minimum, a degenerate maximum or a maximum when b is positive, equal to zero, or negative, respectively. It follows directly from (25) that $\log L$ has two additional stationary points (being maxima) if b is positive, that is, if Eq. (20) is met. Notice that condition (20) is always met for noise free data. However, in practice the data will be corrupted by noise and for particular realizations of the noise, condition (20) may not be met. Then $A = 0$ will be a (possibly degenerate) maximum. Moreover, if condition (20) is not met, b in (25) is negative and thus $\log L$ is convex, which means that $A = 0$ will be the only, and therefore, the global maximum of the $\log L$ function. This implies that under the influence of noise the two maxima and one minimum have merged into one single maximum at $A = 0$. This maximum then corresponds to the ML estimate. Note that, since condition (20) is identical to (and therefore can be replaced by) the condition $\widehat{A_c^2} > 0$, the probability that the ML estimate is found at $A = 0$ is equal to the probability that $\widehat{A_c^2} \leq 0$. This probability can be computed from the PDF given in Eq. (10).

It follows from these considerations that, when the conventional estimator becomes invalid, the ML estimator will still yield physically relevant results.

IV. Simulation experiments

In order to compare the conventional estimator $\hat{A}_c$ to the ML estimator $\hat{A}_{ML}$ described above, an experiment was simulated in which the underlying signal was estimated from 16 Rician distributed data points ($N = 16$) as a function of the noise standard deviation σ. The true value of A was 100. The ML estimate was obtained by maximization of the likelihood function using Brent's algorithm [18]. This is an efficient one-dimensional optimization method based on parabolic interpolation which converges rapidly as the likelihood function is well described by a parabola. The same experiment of determining $\hat{A}_c$ and $\hat{A}_{ML}$ was repeated 2.10^5 times after which the averages $\langle \hat{A}_c \rangle$ and $\langle \hat{A}_{ML} \rangle$ were computed. The results are shown in Fig. 2 along with the 95% confidence intervals.

Fig. 2 Comparison between the conventional and the ML estimator for $N = 16$. Each point denotes the average of 10^5 estimations. Also the 95% confidence interval is shown.

From that figure one can see that at high SNR ($\mathrm{SNR} > 3$) the ML estimator cannot be distinguished from an unbiased estimator, whereas the conventional estimator is clearly biased (Fig. 2a). As can also be observed, the experimental estimations $\hat{A}_c$ are in agreement with the expectation value of $\hat{A}_c$, predicted by Eq. (12).

At low SNR ($\mathrm{SNR} < 3$) the use of $\hat{A}_c$ is no longer justified because the probability that $\hat{A}_c^2$ is negative becomes too high. As to still compare the ML estimator with the conventional one, we modified the conventional estimator in these adverse cases to yield the same estimate as the ML estimator: $\hat{A}_c = 0$. From Fig. 2b one can observe that both estimators become biased though the bias of the ML estimator is significantly smaller compared to the modified conventional estimator. The bias of the ML estimator has to do with the increasing probability of a structural change of the likelihood function. For low SNR, simulation experiments have shown the occurrence of both structures of $\log L$, described above, i.e., only one maximum or two maxima and one minimum. Some $\log L$ functions obtained from simulation experiments are shown in Fig. 3 for high and low SNR.

Fig. 3 Likelihood functions for high (a) and low (b) SNR with $N = 16$. The different curves correspond to different realizations of the same experiment.

Up to now, no other structures were observed. Remark that the occurrence of only one maximum at positive A-values makes the computational requirements for the maximization of the $\log L$ function very low.

In this paper the true value of the noise variance σ^2 was assumed to be known. In practice however, σ^2 needs to be estimated from the background or from homogeneous signal regions [3]. Thereby, the accuracy of the σ^2 estimate is often influenced by systematic errors due to for example ghosting artefacts. This problem can be tackled by acquiring two realisations of the same image [19],[20]. However, if the noise variance estimate can not be prevented from degradation by systematic errors, σ^2 will automatically be over-estimated. An erroneous noise estimate will in turn influence the signal estimates discussed in this paper. Simulation experiments however showed that even with a 10% over- or under-estimated noise variance value ML estimation still yields significantly better results compared to conventional estimation.

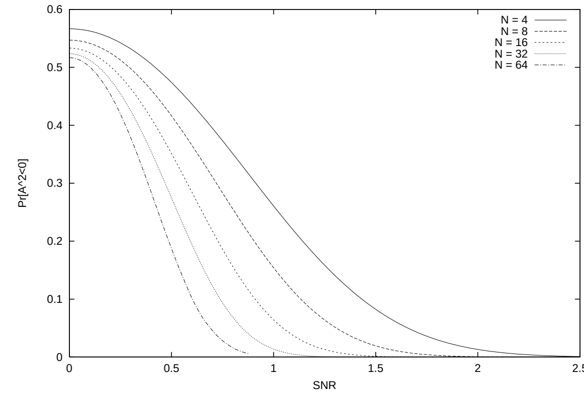


Fig. 1.

V. Conclusion

In this paper the problem of signal estimation from Rician distributed data was discussed. It has been shown that the conventional estimator may not be used at low SNR unless a large amount of data points is used, which is often not available in practice. Even at high SNR the use of this estimator is still not recommended since it is biased.
As an alternative, the Maximum Likelihood estimator was proposed because it outperforms the conventional one with respect to accuracy. The ML estimator yields physically relevant solutions for the whole range of SNR's. Moreover, it was shown that, unlike the conventional estimator, the ML estimator cannot be distinguished from an unbiased estimator at high SNR.

VI. Note added in proof

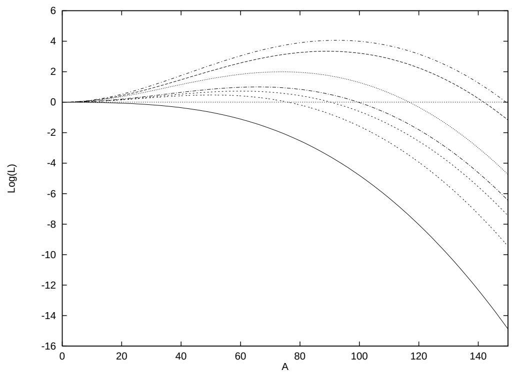
After completion of this manuscript, the authors discovered the existence of a paper by Bonny et al. [21] in which, independently, similar results on ML estimation of the signal parameter from magnitude MR data have been presented. However, the results of that paper show severe discrepancies with those presented in the present work. These discrepancies deserve further study.

References

[1] Y. Wang and T. Lei, "Statistical analysis of MR imaging and its applications in image modeling," in *Proceedings of the IEEE International Conference on Image Processing and Neural Networks*, 1994, vol. I, pp. 866–870.

[2] S. O. Rice, "Mathematical analysis of random noise," *Bell System of Technology*, vol. 23, pp. 282, 1944.

[3] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Medical Physics*, vol. 12, no. 2, pp. 232–233, 1985.



(a)

Fig. 2.

[4] A. H. Andersen and J. E. Kirsch, “Analysis of noise in phase contrast MR imaging,” *Medical Physics*, vol. 23, no. 6, pp. 857–869, 1996.

[5] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, “Nonlinear anisotropic filtering of MRI data,” *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, pp. 221–232, 1992.

[6] G. Z. Yang, P. Burger, D. N. Firmin, and S. R. Underwood, “Structure adaptive anisotropic filtering for magnetic resonance image enhancement,” in *Proceedings of CAIP: Computer Analysis of Images and Patterns*, 1995, pp. 384–391.

[7] S. J. Garnier and G. L. Bilbro, “Magnetic resonance image restoration,” *Journal of Mathematical Imaging and Vision*, vol. 5, pp. 7–19, 1995.

[8] M. A. Bernstein, D. M. Thomasson, and W. H. Perman, “Improved detectability in low signal-to-noise ratio magnetic resonance images by means of phase-corrected real construction,” *Medical Physics*, vol. 16, no. 5, pp. 813–817, 1989.

[9] G. McGibney and M. R. Smith, “An unbiased signal-to-noise ratio measure for magnetic resonance images,” *Medical Physics*, vol. 20, no. 4, pp. 1077–1078, 1993.

[10] A. J. Miller and P. M. Joseph, “The use of power images to perform quantitative analysis on low SNR MR images,” *Magnetic Resonance Imaging*, vol. 11, pp. 1051–1056, 1993.

[11] H. Gudbjartsson and S. Patz, “The Rician distribution of noisy MRI data,” *Magnetic Resonance in Medicine*, vol. 34, pp. 910–914, 1995.