Majorization-Minimization Algorithms for Maximum Likelihood Estimation of Magnetic Resonance Images

Qianyi Jiang, Said Moussaoui, Jérôme Idier, Guylaine Collewet, Mai Xu

Majorization-Minimization Algorithms for Maximum Likelihood Estimation of Magnetic Resonance Images

Qianyi JIANG*†, Saïd MOUSSAOUI*, Jérôme IDIER*, Guylaine COLLEWET‡, Mai XU†
* Ecole Centrale Nantes, CNRS, LS2N, UMR 6004, 44321 Nantes, France
email: qianyi.jiang@eleves.ec-nantes.fr
email: said.moussaoui@ec-nantes.fr
email: jerome.idier@cnrs.fr
† Beihang University, Xueyuan Road, 100191 Beijing, China
email: MaiXu@buaa.edu.cn
‡ IRSTEA, UR OPAALE, CS 64427, 35044 Rennes, France
email: guylaine.collewet@irstea.fr

***Abstract*—This paper addresses maximum likelihood estimation of images corrupted by a Rician noise, with the aim to propose an efficient optimization method. The application example is the restoration of magnetic resonance images. Starting from the fact that the criterion to minimize is non-convex but unimodal, the main contribution of this work is to propose an optimization scheme based on the majorization-minimization framework after introducing a variable change allowing to get a strictly convex criterion. The resulting descent algorithm is compared to the classical MM descent algorithm and its performances are assessed using synthetic and real MR images. Finally, by combining these two MM algorithms, two optimization strategies are proposed to improve the numerical efficiency of the image restoration for any signal-to-noise ratio.**

***Index Terms*—Magnetic resonance imaging, Rician noise, maximum likelihood estimation, iterative optimization, majorization-minimization.**

I. Introduction

Noise in Magnetic Resonance Imaging (MRI) is a major concern which may affect image interpretation and exploitation. In order to obtain images with high signal-to-noise-ratio (SNR), long measurement duration is usually required. In practical use, it is necessary to reduce the acquisition time by associating some numerical noise reduction routines to the imaging system. Various kinds of denoising methods have been investigated (See for instance [1] which gives a thorough overview of existing methods) from which model-based statistical estimation is becoming increasingly prevalent. In the case where the MR data are acquired in the complex domain (K-space), the noise in both real and imaginary parts is independent and identically distributed and it is usually supposed to be Gaussian [2]. However, in most applications, the signal magnitude is processed since the phase is sensitive to many factors such as the magnetic field heterogeneity, the temperature of the sample or the motion of the analyzed object [3]. Accordingly, the computation of magnitude raw images yields Rician distributed data which makes the denoising techniques based on the hypothesis of additive Gaussian noise suboptimal.

978-1-5386-1842-4/17/$31.00

In the literature, several Rician denoising methods have been mentioned, including the so-called conventional approach, exploiting the properties of the second-order moment of the Rician distribution [4], the linear minimum mean square error (LMMSE) estimator [5] and the signal reconstruction method based on maximum likelihood (ML) estimation [6]. When the noisy data can be modeled accurately, the ML estimator is commonly applied since it tends to be consistent and asymptotically efficient under general conditions [7].

The negative log-likelihood (NLL) function for statistical estimation from Rician distribution is non-convex but unimodal. Although the conventional optimization methods can converge to the global minimum, their convergence efficiency depends severely on the initialization and the value to be estimated. In [8], Getreuer et al. proposed a maximum a posteriori (MAP) estimation with total variation prior and considered a somewhat complex convex approximation to overcome the non-convexity. Dong et al. [9] established another convex Rician model by adding a quadratic penalty term into the previous model under some certain conditions to ensure the existence and the uniqueness of the solution. [10] and [11] applied an Expectation-Maximization (EM) algorithm for Rician denoising estimation based on a convex tangent majorant function. Varadarajan [12] generalized this approach into the Majorization-Minimization (MM) framework for non-central chi (NCC) family and demonstrated the computational efficiency compared with generic numerical optimization techniques. However, as presented in the following section, this algorithm results in a poor convergence speed for low values of signal-to-noise ratio (SNR) which is partly due to the non-convexity of the NLL function.

In this paper, we propose a novel MM optimization algorithm of the NLL function based on a transformed convex criterion that substantially accelerates the convergence speed at low SNR. By combining the conventional and our proposed MM approaches, we obtain two algorithms which improve the computational efficiency whatever the SNR level. The rest of the paper is organized as follows. Section II introduces the ML estimation in Rician noise. Section III overviews the principle of the existing non-convex model based MM algorithm. In Section IV, we introduce the novel MM algorithm and demonstrate the validation of our proposed majorant function. Section V compares the two algorithms and considers some combination strategies. Section VI illustrates the experimental results and Section VII presents some conclusions and perspectives of this work.

II. ML ESTIMATION FROM RICIAN DISTRIBUTION

Complex MRI signal is corrupted by an additive zero-mean Gaussian noise. Thus, its magnitude follows a Rician distribution:

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

where I_0 is the modified zeroth order Bessel function of the first kind, M_i represents the ith measurement with the corresponding true signal amplitude A [1], σ^2 is the variance of the Gaussian noise and u denotes the step function.

Under the assumption of independence of N measurements M_i, the estimation of the pixel intensity A can be obtained by maximizing the likelihood function or equivalently minimizing the NLL function, given as:

$$\begin{aligned} L(A) = -\log \prod_{i=1}^{N} p(M_i|A) = \sum_{i=1}^{N} \frac{A^2 + M_i^2}{2\sigma^2} \\ - \sum_{i=1}^{N} \log I_0\left(\frac{AM_i}{\sigma^2}\right) - \sum_{i=1}^{N} \log\left(\frac{M_i}{\sigma^2}\right). \end{aligned} \qquad (2)$$

In the case where a single image is acquired ($N = 1$) hereafter for each amplitude A, there is only one measurement denoted M. In the sequel, the general case will be considered, and a discussion on the impact of the number of image accumulations will be given in section VI-A.

As mentioned in [13], the objective function (2) has one or two stationary points where $A = 0$ is always a trivial solution.

III. MM DESCENT ON THE INITIAL CRITERION

The objective function cannot be minimized explicitly, therefore, an iterative optimization algorithm based on the MM framework can be applied [14]. The principle of MM is to iteratively construct and minimize a surrogate cost function $H_L(\cdot, A_k)$ known as *tangent majorant* instead of directly minimizing the initial objective function $L(\cdot)$. Each iteration k of the algorithm consists of:

$$\text{Majorization:} \qquad H_L(A, A_k) \geqslant L(A), \quad \forall A \geqslant 0 \qquad (3)$$

$$\text{Minimization:} \qquad A_{k+1} = \arg\min_{A}\, H_L(A, A_k) \qquad (4)$$

The main feature of this descent method is that $H_L(\cdot, A_k)$ admits a closed-form solution which is easier to compute than that of $L(\cdot)$.

The MM algorithm proposed in [12] takes into account that

$$f_i(A) = -\log I_0\big(c_i A\big)$$

with $c_i = \frac{M_i}{\sigma^2}$, is a concave function. It is majorized at any point A_k by an affine function:

$$f_i(A) = -\log I_0\big(c_i A\big) \leqslant g_i(A_k)(A - A_k) + f_i(A_k), \qquad (5)$$

where $g_i(A_k)$ is the gradient of $f_i(\cdot)$ at A_k:

$$g_i(A_k) = -c_i\, r_1(c_i A_k), \qquad (6)$$

with $r_1 = \frac{I_1}{I_0}$ and I_1 the modified first order Bessel function of the first kind.

Accordingly, the tangent majorant function $H_L(\cdot, A_k)$ of L is given as:

$$H_L(A, A_k) = \frac{N}{2}\frac{A^2}{\sigma^2} + \sum_{i=1}^{N} g_i(A_k)(A - A_k) + f_i(A_k), \qquad (7)$$

and its gradient $\nabla H_L(\cdot, A_k)$ is expressed as :

$$\nabla H_L(A, A_k) = \frac{NA}{\sigma^2} - \sum_{i=1}^{N} c_i\, r_1(c_i A_k). \qquad (8)$$

The value of A_{k+1} canceling this gradient is directly derived as:

$$A_{k+1} = \frac{\sigma^2}{N} \sum_{i=1}^{N} c_i\, r_1(c_i A_k). \qquad (9)$$

It can be seen that this update scheme corresponds to a classical gradient descent strategy:

$$A_{k+1} = A_k - \frac{\sigma^2}{N} \nabla L(A_k), \qquad (10)$$

with a fixed stepsize $\alpha_k = \dfrac{\sigma^2}{N}$.

In the rest of this paper, this algorithm is referred to as *MM-χ* as the Rician distribution is a special case of the non-central χ distribution.

IV. MM DESCENT ON A TRANSFORMED CRITERION

According to Proposition 3 in [15], by introducing a change of variable $y = A^2$, the transformed objective function becomes convex and unimodal with respect to the variable y. Therefore, we propose to construct a quadratic majorant function for the convex criterion and we will apply the descent scheme MM for the optimization with respect to y and deduce the value of A.

[1] The corresponding pixel index is omitted to simplify the notations.

As illustrated by Figure 1, such scheme consists on constructing a non quadratic majorant in the initial space defined with respect to A which turns out to be more efficient than the quadratic surrogation around the concave zone.

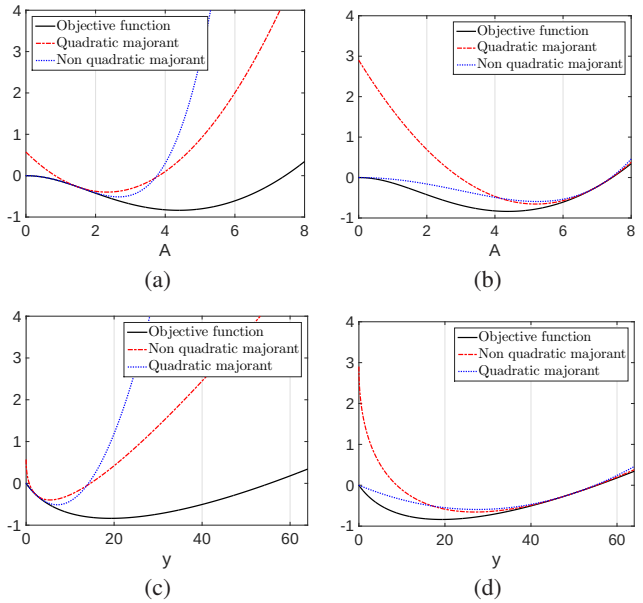

Fig. 1. Illustration of the initial non-convex criterion and the transformed convex criterion and the tangent majorant functions in the initial space for (a) $A_k = 1.5$, (b) $A_k = 7$ and (b, d) after transformation $y = A^2$.

Indeed, after the transformation $y = A^2$, the objective function is rewritten as:

$$\widetilde{L}(y) = \frac{Ny}{2\sigma^2} - \sum_{i=1}^{N} \log I_0(c_i\sqrt{y}), \tag{11}$$

Its gradient and Hessian are given by:

$$\nabla\widetilde{L}(y) = \frac{N}{2\sigma^2} - \sum_{i=1}^{N} \frac{c_i}{2\sqrt{y}} r_1(c_i\sqrt{y}), \tag{12}$$

$$\nabla^2\widetilde{L}(y) = \sum_{i=1}^{N} \frac{c_i^2}{4y}\left(r_1^2(c_i\sqrt{y}) + \frac{2}{c_i\sqrt{y}} r_1(c_i\sqrt{y}) - 1\right). \tag{13}$$

Finally, as noted in section (2.2) of [16],

$$r_1^2(x) + \frac{2}{x} r_1(x) - 1 > 0 \quad \forall x \geqslant 0, \tag{14}$$

and hence the positivity of the Hessian $\nabla^2\widetilde{L}(y)$ can be deduced.

A. Construction of the majorant function

The proposed quadratic tangent majorant is given by:

$$\widetilde{H}_L(y, y_k) = \frac{1}{2} B_k (y - y_k)^2 + \nabla\widetilde{L}(y_k)(y - y_k) + \widetilde{L}(y_k), \tag{15}$$

with a curvature defined by:

$$B_k = \frac{2\left(\nabla\widetilde{L}(y_k) y_k - \widetilde{L}(y_k)\right)}{y_k^2}. \tag{16}$$

a) Discussion.: This curvature results from the specification to get a majorant function that passes through the same point as $\widetilde{L}$ at $y = 0$.
This construction makes it possible to obtain a tangent quadratic majorant with the smallest curvature and thus allowing larger steps. The property of majorization can be deduced from the fact that the third derivative of the $\widetilde{F}$ is negative and therefore its curvature is decreasing. Demonstrations are omitted here due to the lack of space and we only show the proof of convexity of the majorant function $H_L(\cdot, y_k)$.

Lemma 1: The function $h(x) = x r_1(x) - 2\log I_0(x)$ is negative for all $x > 0$.

Proof 1: Taking the derivative of this function

$$h'(x) = x - x r_1(x)^2 - 2 r_1(x)$$

and according to (14), one can deduce that $h(x)$ is a decreasing function. Therefore $x r_1(x) - 2\log I_0(x) < h(0) = 0, \forall x > 0$, which completes the proof.

Theorem 1: The curvature B_k of the tangent majorant function $\widetilde{H}(\cdot, y_k)$ defined by Equation (16) is positive for any y_k non-negative.

Proof 2: Note that $B_k(y)$ can be formulated as the first derivative of $b(y) = 2\frac{\widetilde{L}(y)}{y}$, which can be expressed as

$$b(y) = \frac{N}{\sigma^2} - 2\sum_{i=1}^{N} \frac{\log I_0(c_i\sqrt{y})}{y},$$

with $c_i = \frac{M_i}{\sigma^2}$.

Its derivative is expressed by:

$$b'(y) = -\sum_{i=1}^{N} \frac{c_i\sqrt{y}\, r_1(c_i\sqrt{y}) - 2\log I_0(c_i\sqrt{y})}{y^2}.$$

Lemma 1 allows to deduce the positivity of the curvature B_k by setting $x = c_i\sqrt{y}$.

B. Majorant function minimization

The value of y_{k+1} canceling the gradient of $\widetilde{H}_L(\cdot, y_k)$ is expressed as:

$$y_{k+1} = y_k - {B_k}^{-1} \nabla\widetilde{L}(y_k). \tag{17}$$

This algorithm is denoted as *MM-*χ^2 since it corresponds to applying the transformation $y = A^2$. It can be noted that this update scheme is similar to a quasi-Newton algorithm where the pseudo-Hessian matrix is deduced from the Majorization principle.

V. Performance analysis

This section is dedicated to an empirical analysis of the performances of the two MM optimization schemes in terms of convergence speed and the proposal of a new descent strategy combining the two algorithms.

A. Comparison of the two MM algorithms

In order to assess the performances for different noise levels, we first consider a simulation with the measured signal module to noise ratio (M/σ) varying from 0 to 8. The number of measurements is set to $N = 1$. A discussion on this choice will be conducted in section VI-A. The optimization is performed using the two MM algorithms. The measurement is taken as the initial value and the stopping condition is based on a gradient norm value less than 10^{-5}.

Figure 2 shows the required number of iterations for the convergence of the two MM algorithms with different SNR values. We can observe that the necessary number of iterations for the convergence of *MM*-χ^2 is much smaller when $\frac{M}{\sigma}$ is less than a value in the order of 1.8, which demonstrates the interest of the variable change especially in the context of a low SNR. However, *MM*-χ algorithm requires fewer iterations when the SNR increases. This can be explained by the proximity of the Rician distribution to a Gaussian for high SNR.

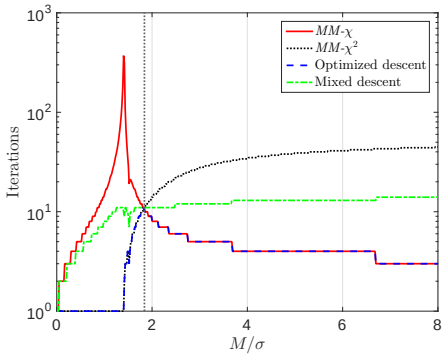

Fig. 2. The number of iterations of the MM descent algorithms for different levels of the signal to noise ratio in the case of a single measurement.

B. Combination of the two MM algorithms

In order to obtain a more efficient algorithm for any value of the SNR, we firstly adopt a linear combination of these two descent algorithms either in the initial space A or in y. The optimization strategy will be referred to hereafter as a *mixed descent* whose main steps are summarized in Algorithm 1.

An *optimized approach* consists in selecting at each iteration the MM solution which induces the largest descent of the criterion (Algorithm 2). As illustrated by Figure 2, the proposed *optimized descent* and *mixed descent* make it possible to take advantage of the two algorithms by significantly reducing the number of iterations whatever the value of SNR. It is worth noting that the *optimized descent* turns out to take automatically the fewest number of iterations while the *mixed descent* results, for $\alpha = 0.5$, in an intermediate convergence speed.

C. Performance analysis

A second experiment is performed to discuss the impact of the number of measurements on the convergence rate of the algorithms. A Monte-Carlo simulation with 10^5 realizations is performed with the true signal to noise ratio A/σ varies from 0 to 8 in the case of measurements number $N = 1$ or $N = 10$. Figure 3 illustrates the average number of iterations using the different variants of the MM algorithms. It can be seen that for both single and multiple measurements cases, the three proposed algorithms outperform the classical MM algorithm when the SNR is lower than a certain level and the *Optimized descent* improves significantly the convergence efficiency. Also, a rise in the number of iterations for A is noted for *MM*-χ when the number of measurements increases.

Algorithm 1 Iterative *mixed descent*

Initialize $A_0 = M$
Choose $Descent$ in A or y
Set α between $(0, 1)$
for $k = 1, 2, 3, ...$ until convergence **do**
  Set $y_k = A_k^2$
  Compute g_k and A_{k+1} according to Eq (6) and (9)
  Compute B_k and y_{k+1} according to Eq (16) and (17)
  switch $Descent$ **do**
    case A
      Set $D_A = A_{k+1} - A_k$
      Set $D_y = \sqrt{y_{k+1}} - \sqrt{y_k}$
      $A_{k+1} = A_k + \alpha D_A + (1 - \alpha) D_y$
    case y
      Set $\widetilde{D}_A = A_{k+1}^2 - A_k^2$
      Set $\widetilde{D}_y = y_{k+1} - y_k$
      $y_{k+1} = y_k + \alpha \widetilde{D}_A + (1 - \alpha) \widetilde{D}_y$
      $A_{k+1} = \sqrt{y_{k+1}}$
end for

Algorithm 2 Iterative *optimized descent*

Initialize $A_0 = M$
for $k = 1, 2, 3, ...$ until convergence **do**
  Set $y_k = A_k^2$
  Compute g_k and A_{k+1} according to eq.(6) and (9)
  Set $L_A = L(A_{k+1})$ as defined in eq.(2)
  Compute B_k and y_{k+1} according to eq.(16) and (17)
  Set $L_y = L(\sqrt{y_{k+1}})$
  if $L_y < L_A$ **then** $A_{k+1} = \sqrt{y_{k+1}}$
  end if
end for

VI. Application to MRI

This section is dedicated to the application to MRI data and a discussion on the relevance of the proposed algorithm.

A. Multiple or single measurements

Considering the dataset used for ML estimation in practical applications, two different raw images can be obtained according to the acquisition set-up. A common one is to collect N

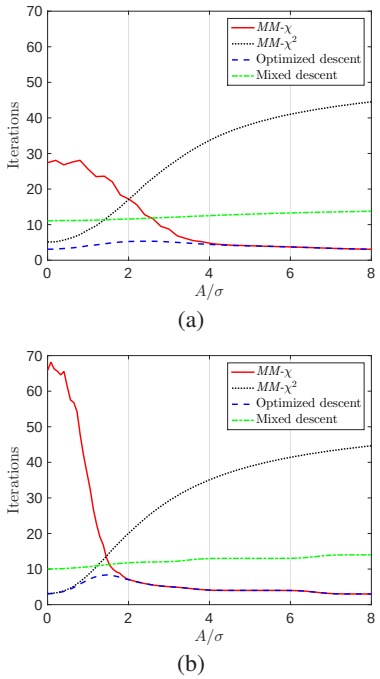

Fig. 3. Number of iterations of the MM descent algorithms for different original signal amplitude and (a) $N = 1$ measurement or (b) $N = 10$ measurements.

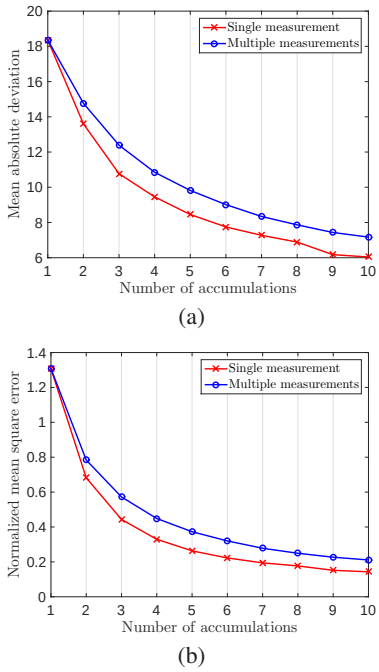

Fig. 4. The mean absolute deviation and normalized mean square error of ML estimation using a single measurement and multiple measurements.

independent complex-valued images then convert them into N magnitude images. An alternative processing consists in accumulating the N complex-valued k-space images into a single observation by computing their average then processing the magnitude image [17]. The question is therefore whether the ML estimation should be performed by taking multiple measurements with single accumulation or a single measurement with multiple accumulations.

We compare these two approaches by an experiment with real MRI data of a cylinder phantom filled with oil. We collected ten repeated measurements of independent complex-valued observations, from which we can compute a number of magnitude measurements or alternatively taking the averaged magnitude image.

Figure 4 presents the mean absolute deviation and normalized mean square error of ML estimation given a number of independent measurements (from 1 to 10) compared with a given single image obtained by averaging the same number of accumulations. It can be observed that the second approach has a better performance in terms of deviation for some different number of accumulations. This demonstrates the superiority of ML estimation using the averaged image.

B. Application to a synthetic data sets

The proposed approach was evaluated using the freely available brain database [18]. A set of realistic MRI data volumes with T_1 pulse sequence and $1mm$ slice thickness produced by the MRI simulator are taken. The noise variance is computed using a method of local noise variance distribution [5].

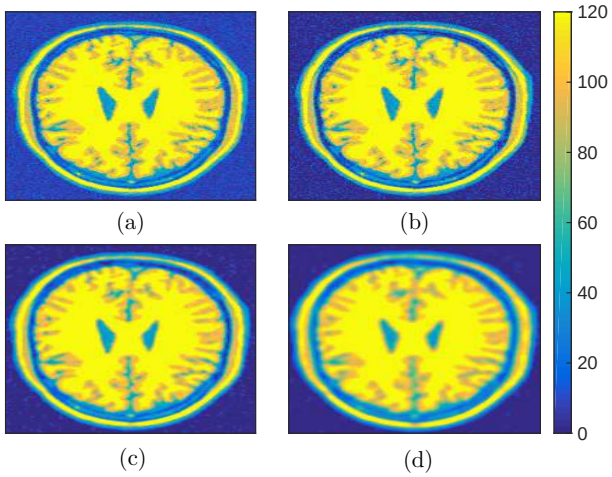

Fig. 5. (a) Original noisy data. (b) ML pixel-wise estimation. (c) ML estimation using 3×3 neighborhoods (d) ML estimation using 5×5 neighborhoods

In Figure 5, three different methods of ML estimation are applied to the noisy data: (b) represents the ML pixel-wise estimation where $N = 1$. As the ML estimation is originally designed to take multiple images and according to the results of section VI-A, a single averaged image is selected. We also perform two additional tests: (c) and (d) corresponds to ML estimation using 3×3 and 5×5 local neighborhoods where the number of measurements is 9 and 25 respectively. As mentioned by Santiago [5], taking into account the local statistics for the estimation yields a better noise reduction but causes edge-blurring. Since this paper discusses the efficiency of different MM optimization strategies, we focus on the

computational performance and not on the relevance of the ML estimation.

The *MM-*χ algorithm and the proposed three algorithms are evaluated with the same dataset and using the same stopping criterion in order to ensure getting identical solutions after convergence. To compare their convergence speed, their computation time (T), as well as the number of iterations (Iter) are reported in Tables I and II. It can be seen that the three proposed approaches substantially outperform *MM-*χ algorithm in terms of computational efficiency. Moreover, as the number of measurements increases, while *MM-*χ requires significantly more iterations, the *Mixed descent* remains a relatively moderate convergence speed while *Optimized descent* requires always the fewest iterations. In the case of pixel-wise estimation where $N = 1$, we obtain the same result as in the previous simulation experiment, and for multiple measurements, the *Mixed descent* turns to be the most efficient in terms of the computation time. It is worth noting that even though for each iteration, the *Mixed descent* and *Optimized descent* are somehow more time-consuming as two criteria evaluations are involved, the combination strategy still results in less overall computation time due to fewer iterations.

TABLE I
THE NECESSARY NUMBER OF ITERATIONS FOR THE FOUR MM ALGORITHMS FOR ML ESTIMATION FOR A SINGLE IMAGE USING PIXEL-WISE, 3×3 AND 5×5 LOCAL NEIGHBORHOODS.

Algorithm	Pixel-wise	3×3	5×5
MM-χ	58	482	459
MM-χ^2	47	47	46
Optimised descent	**9**	**11**	**11**
Mixed descent	14	14	14

TABLE II
THE COMPUTATION TIME (IN SECONDS) OF THE FOUR MM ALGORITHMS FOR ML ESTIMATION FOR SINGLE IMAGE USING PIXEL-WISE, 3×3 AND 5×5 LOCAL NEIGHBORHOODS.

Algorithm	Pixel-wise	3×3	5×5
MM-χ	5.8	308.1	839.7
MM-χ^2	3.1	19.2	51.2
Optimized descent	**1.6**	13.9	41.1
Mixed descent	1.7	**11.4**	**31.9**

VII. CONCLUSION

In this paper, we have investigated the problem of signal estimation for MRI with Rician noise. The main proposal is a novel MM framework based on a convex criterion which improves significantly the convergence efficiency in the context of a low SNR. The validation of its corresponding quadratic tangent majorant function has been demonstrated mathematically and empirically. Furthermore, we introduced two algorithms combining the two MM strategies in order to obtain the optimal performance.

From the simulation part and the real images analysis results, we can conclude that both algorithms have shown substantial improvement over *MM-*χ in terms of computational efficiency. A further theoretical analysis including the convergence rates of the algorithms is being investigated. The extension to penalized ML estimation is being explored in order to enhance the denoising performances.

ACKNOWLEDGMENT

This work was funded by the GDR 720 ISIS in the framework of the 2016 call for projects.

REFERENCES

[1] Mohan J, Krishnaveni V, and Yanhui Guo, "A survey on the magnetic resonance image denoising methods," *Biomedical Signal Processing and Control*, vol. 9, pp. 56–69, 2014.

[2] Hákon Gudbjartsson and Samuel Patz, "The Rician distribution of noisy MRI data," *Magnetic resonance in medicine*, vol. 34, no. 6, pp. 910–914, 1995.

[3] Peter B Roemer, William A Edelstein, Cecil E Hayes, Steven P Souza, and OM Mueller, "The NMR phased array," *Magnetic resonance in medicine*, vol. 16, no. 2, pp. 192–225, 1990.

[4] McGibney G and MR Smith, "An unbiased signal-to-noise ratio measure for magnetic resonance images," *Medical physics*, vol. 20, no. 4, pp. 1077–1078, 1993.

[5] Santiago Aja-Fernández, Carlos Alberola-López, and Carl-Fredrik Westin, "Noise and signal estimation in magnitude MRI and Rician distributed images: a LMMSE approach," *IEEE transactions on image processing*, vol. 17, no. 8, pp. 1383–1398, 2008.

[6] Jan Sijbers, Arnold Jan den Dekker, Paul Scheunders, and Dirk Van Dyck, "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 357–361, 1998.

[7] Karlsen Ole T, Rieko Verhagen, and Wim MMJ Bovee, "Parameter estimation from Rician-distributed data sets using a maximum likelihood estimator: Application to T1 and perfusion measurements," *Magnetic resonance in medicine*, vol. 41, no. 3, pp. 614–623, 1999.

[8] Pascal Getreuer, Melissa Tong, and Luminita Vese, "A variational model for the restoration of MR images corrupted by blur and rician noise," *Advances in Visual Computing*, pp. 686–698, 2011.

[9] Yiqiu Dong and Tieyong Zeng, "A convex variational model for restoring blurred images with multiplicative noise," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1598–1625, 2013.

[10] Victor Solo and Joonki Noh, "An EM algorithm for Rician fMRI activation detection," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*. IEEE, 2007, pp. 464–467.

[11] Hongtu Zhu, Yimei Li, Joseph G Ibrahim, Xiaoyan Shi, Hongyu An, Yashen Chen, Wei Gao, Weili Lin, Daniel B Rowe, and Bradley S Peterson, "Regression models for identifying noise sources in magnetic resonance images," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 623–637, 2009.

[12] Divya Varadarajan and Justin P Haldar, "A majorize-minimize framework for Rician and non-central Chi MR images," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 2191–2202, 2015.

[13] Sijbers Jan and Den Dekker AJ, "Maximum likelihood estimation of signal amplitude and noise variance from MR data," *Magnetic Resonance in Medicine*, vol. 51, no. 3, pp. 586–594, 2004.

[14] David R Hunter and Kenneth Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[15] Jérôme Idier and Guylaine Collewet, "Properties of Fisher information for Rician distributions and consequences in MRI," Tech. Rep., IRCCyN/IRSTEA, 2014.

[16] Andrea Laforgia and Pierpaolo Natalini, "Some inequalities for modified Bessel functions," *Journal of Inequalities and Applications*, vol. 2010, no. 1, pp. 253035, 2010.

[17] Chris D Constantinides, Ergin Atalar, and Elliot R McVeigh, "Signal-to-noise measurements in magnitude images from NMR phased arrays," *Magnetic Resonance in Medicine*, vol. 38, no. 5, pp. 852–857, 1997.

[18] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans, "Brainweb: Online interface to a 3D MRI simulated brain database," in *NeuroImage*. Citeseer, 1997.