Iterative reweighted linear least squares for accurate, fast, and robust estimation of diffusion magnetic resonance parameters

FULL PAPER

Magnetic Resonance in Medicine 00:00–00 (2014)

Iterative Reweighted Linear Least Squares for Accurate, Fast, and Robust Estimation of Diffusion Magnetic Resonance Parameters

Quinten Collier,[1]* Jelle Veraart,[1] Ben Jeurissen,[1] Arnold J. den Dekker,[1,2] and Jan Sijbers[1]

Purpose: Diffusion-weighted magnetic resonance imaging suffers from physiological noise, such as artifacts caused by motion or system instabilities. Therefore, there is a need for robust diffusion parameter estimation techniques. In the past, several techniques have been proposed, including RESTORE and iRESTORE (Chang et al. Magn Reson Med 2005; 53:1088–1095; Chang et al. Magn Reson Med 2012; 68:1654–1663). However, these techniques are based on nonlinear estimators and are consequently computationally intensive.

Method: In this work, we present a new, robust, iteratively reweighted linear least squares (IRLLS) estimator. IRLLS performs a voxel-wise identification of outliers in diffusion-weighted magnetic resonance images, where it exploits the natural skewness of the data distribution to become more sensitive to both signal hyperintensities and signal dropouts.

Results: Both simulations and real data experiments were conducted to compare IRLLS with other state-of-the-art techniques. While IRLLS showed no significant loss in accuracy or precision, it proved to be substantially faster than both RESTORE and iRESTORE. In addition, IRLLS proved to be even more robust when considering the overestimation of the noise level or when the signal-to-noise ratio is low.

Conclusion: The substantially shortened calculation time in combination with the increased robustness and accuracy, make IRLLS a practical and reliable alternative to current state-of-the-art techniques for the robust estimation of diffusion-weighted magnetic resonance parameters. **Magn Reson Med 000:000–000, 2014.**

Key words: MRI; robust; outlier detection; weighted linear least squares; diffusion tensor imaging

INTRODUCTION

Because some widely used diffusion models are log-linear, the parameter estimation problem can be structured as a linear regression problem by taking the natural logarithm of the diffusion-weighted magnetic resonance (DW-MR) signals. Typical examples include the popular diffusion tensor imaging (DTI) model (1) and the diffusion kurtosis imaging model (2, 3). Thus, the linear least squares (LLS) estimator (4) and its weighted variants (WLLS) (1, 5, 6) pose an interesting alternative to the nonlinear least squares (NLS) estimator (7) for the estimation of the unknown model parameters. Indeed, unlike iterative nonlinear estimation strategies, the (W)LLS estimator can be expressed in a closed form and is, therefore, computationally efficient. Moreover, if the rice-distributed DW-MR signals have a signal-to-noise ratio (SNR) value higher than two, the LLS is unbiased because the expectation value of the error term is zero after the log-transformation (5).

Unfortunately, the performance of most diffusion parameter estimators is often degraded by physiological noise. Spatially and temporally varying artifacts, such as cardiac pulsation, motion, or system instabilities, will perturb the DW-MR signals; yet they are ignored by conventional estimators. More robust estimators, i.e., estimators that are less sensitive to outliers, are needed to address these perturbations. In the recent literature, several approaches to increase the robustness of diffusion parameter estimators have been proposed (8–15). One of the first approaches was introduced by Mangin et al. (16). Their method is based on the well-known Geman–McClure M-estimator (17). In 2005, Chang et al. introduced a more advanced robust estimator based on the work of Mangin et al., called RESTORE (18). To date, RESTORE is still one of the best known robust tensor estimation methods and is still commonly used in clinical settings (e.g., (19–21)). Recently, the RESTORE method was adapted to better address signal dropouts; this effort led to the introduction of iRESTORE (22). Both RESTORE and iRESTORE, however, are based on computationally expensive NLS estimation.

Given the strengths of the WLLS, our goal is to develop an improved, robust outlier detection method built on linear estimators to speedup the robust estimation process. This led to the development of our iterative reweighted linear least squares (IRLLS) estimator. In addition to exploiting the strengths of the WLLS, we also aim at improving, compared to RESTORE, the sensitivity of the detection method to signal dropouts that are shown to be more likely in DW-MRI, without becoming blind to signal overshoots (cf. iRESTORE).

In this work, IRLLS is compared to three commonly used diffusion parameter estimation methods, including the state-of-the-art robust approaches RESTORE and iRESTORE, and

[1]iMinds-Vision Lab, Department of Physics, University of Antwerp, Antwerp, Belgium.

[2]Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands.

Grant sponsor: Scientific Research-Flanders (FWO); Grant sponsor: Interuniversity Attraction Poles Program (IUAP); Grant number: P7/11.

*Correspondence to: Quinten Collier, M.S., iMinds-Vision Lab, Department of Physics, University of Antwerp, Universiteitsplein 1, N 1.16, 2610 Wilrijk, Belgium. E-mail: Quinten.Collier@uantwerpen.be

Received 20 February 2014; revised 20 May 2014; accepted 13 June 2014

DOI 10.1002/mrm.25351

Published online 00 Month 2014 in Wiley Online Library (wileyonlinelibrary.com).

a regular WLLS estimation. This choice is motivated by the fact that we wish to compare IRLLS to other voxel-based outlier detection methods that make minimal assumptions about the origin of the artifacts and impose minimal constraints on the data acquisition (which excludes, for instance, the need for high angular resolution diffusion imaging acquisition (23)), thus making them widely applicable. Among such methods, RESTORE and iRESTORE are the most common methods for outlier detection in DW-MRI. Furthermore, both RESTORE and iRESTORE are freely available as part of TORTOISE (24). Our algorithm can be downloaded from http://www.visielab.ua.ac.be/irlls.

METHODS

This section starts with a brief description of the diffusion tensor model. In the next subsection, some properties of the error terms are discussed. Knowledge of these error terms is important because they determine the properties of the least squares estimators (5). In the final subsection, the proposed IRLLS algorithm for robust diffusion parameter estimation is explained in detail. This last subsection is subdivided into four parts, each one elaborating on an important step of the proposed algorithm. For the sake of simplicity, only the DTI model is discussed. The presented methodology, however, can be easily extended to higher order log-linear models such as diffusion kurtosis imaging.

Diffusion Tensor Model

We assume that N DW-MR images were acquired, wherein each acquisition is characterized by its diffusion sensitizing gradient strength b and direction $\boldsymbol{g}$. This information is stored in an $N \times 7$ b-matrix $\boldsymbol{B} = \{\boldsymbol{b}_i = [1, -b_i g_{x,i}^2, -2b_i g_{x,i} g_{y,i}, -2b_i g_{x,i} g_{z,i}, -b_i g_{y,i}^2, -2b_i g_{y,i} g_{z,i}, -b_i g_{z,i}^2] : i = 1, \ldots, N\}$, which embodies the effects of the diffusion weighting. In DTI, the magnitude of a noise free DW-MR signal S_i is modeled as:

$$S_i(\boldsymbol{b}_i; \theta) = \exp(\boldsymbol{b}_i \boldsymbol{\theta}). \quad [1]$$

Here, $\boldsymbol{\theta} = [\log(S_0), D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}, D_{zz}]^{\mathrm{T}}$ denotes a 7×1 parameter vector with S_0—the nondiffusion-weighted signal intensity, and D_{jk}—the independent elements of the symmetric diffusion tensor $\boldsymbol{D}$. Using this model, the error term ϵ_i can now be introduced. It is defined as the difference between the measured signal $\tilde{S}_i$ and the true noise-free DW-MR signal S_i:

$$\epsilon_i \equiv \tilde{S}_i(\boldsymbol{b}_i; \theta) - \exp(\boldsymbol{b}_i \boldsymbol{\theta}). \quad [2]$$

Under some well-defined conditions (25–27), $\tilde{S}_i$ can be modeled as a Rician-distributed random variable. Furthermore, when looking at the natural log-transformation of the measured signal $\tilde{S}_i$ and its corresponding noise-free DW-MR signal $\ln(S_i)$, expression (2) can be expressed in linear form (28):

$$\epsilon_i^* \equiv \ln\left(\tilde{S}_i(\boldsymbol{b}_i; \theta)\right) - \boldsymbol{b}_i \boldsymbol{\theta}. \quad [3]$$

Here, ϵ_i^* denotes the error term in linear space. This model allows the use of linear estimators as will be described in the Weighted LLS Estimation subsection.

Properties of the Error Terms

It is insightful to take a closer look at the statistical properties of the error terms in both linear and nonlinear space to better understand the performance of (W)LLS and NLS estimators in terms of accuracy and precision. It is well-known that when uncorrelated errors are assumed, the (W)LLS estimators correspond to maximum likelihood estimators if the following criteria are met: (1) the expectation value of the error terms is zero, (2) the weights equal the reciprocal of the variance of the error terms, and (3) the error terms are normally distributed. To verify the degree to which these conditions are met, the first three (standardized) moments of the error term distribution are discussed. These moments are also shown in Figure 1. For this figure, both Rician- and log-Rician-distributed variables were simulated using Monte Carlo techniques for SNR values ranging from 0 to 10, with 2.5×10^6 data realizations for each value of SNR. The SNR is defined as the ratio of the noise-free signal intensity η to the standard deviation of the noise in the original complex data σ_n: $\mathrm{SNR} \equiv \eta/\sigma_n$. During these simulations, the variance of the underlying Gaussian noise σ_n^2 was kept constant at unity, while the signal intensity was increased to achieve different levels of SNR. For both sets of distributions, the expectation value, variance, and skewness of the magnitude data are subsequently plotted as a function of SNR.

Mean

In magnitude MR images, the expectation value of the error term will have a positive bias, which increases with decreasing SNR (25). Only for SNR values larger than 10, the bias of the error term can be assumed to be zero. However, after a log-transformation, the expectation value of the error term will be virtually zero for SNR values higher than 2, making (W)LLS estimators unbiased (5), and thus, more accurate, compared to the NLS. The expectation value of the error term in linear and nonlinear space as a function of the SNR is shown in Figure 1a.

Variance

Knowledge of the (scaled) error variance allows one to determine the ideal weights to account for the potential heteroscedasticity of the measurements. Heteroscedasticity denotes the situation where the variance of measurements is not constant. In nonlinear space, heteroscedasticity is often ignored. However, in linear space, this cannot be done because here, the variance is more strongly dependent on the underlying DW signal (5). As shown in Figure 1b, this variance can be well-approximated by SNR^{-2} for SNR values larger than 2. In practice, the (scaled) variance, or equivalently the weights terms, can be computed using an LLS estimate of $S_i(\boldsymbol{b}_i; \boldsymbol{\theta})$ (5). The practical importance of using this estimate instead of the measured signal $\tilde{S}_i(\boldsymbol{b}_i; \theta)$ was again demonstrated by Veraart et al. (6). For a theoretical basis on the variance on the data distributions in the nonlinear and linear space, we refer the reader to (29) and (5), respectively.

Skewness

Finally, the third standardized moment of the error distributions, better known as the skewness, also provides

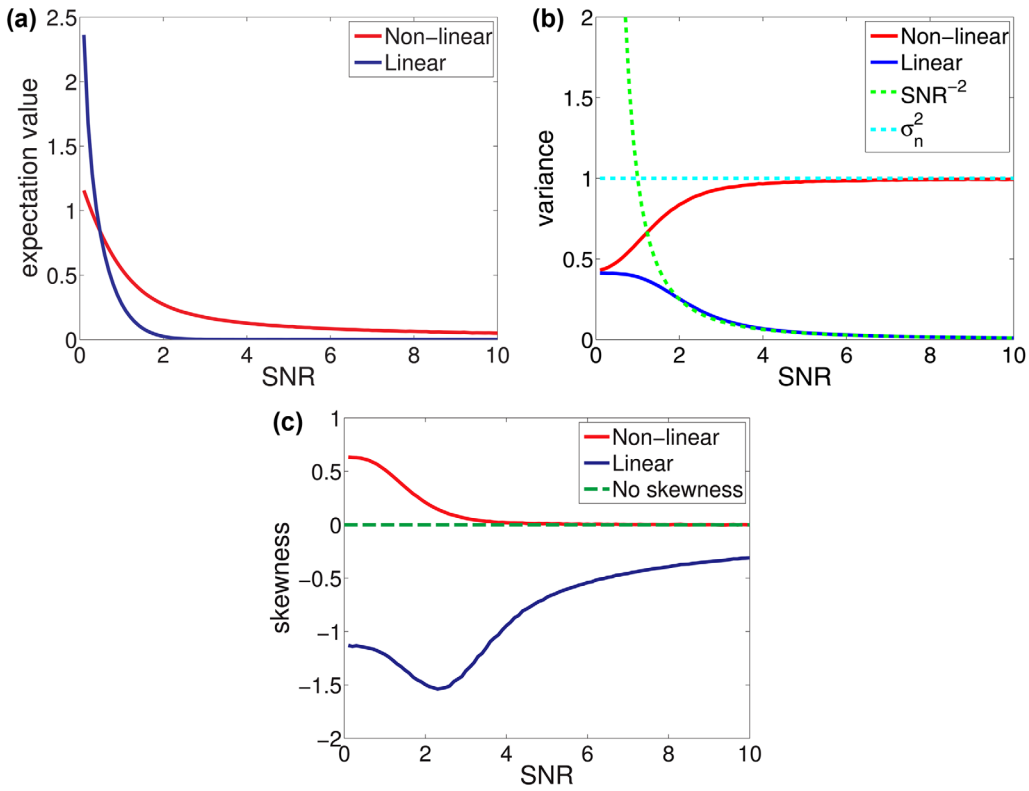

FIG. 1. The expectation value, variance, and skewness of the error terms in linear and nonlinear space are shown in figures (**a**), (**b**), and (**c**), respectively. Monte Carlo techniques were used to simulate both Rician- and log-Rician-distributed variables for SNR values ranging from 0 to 10 with 2.5×10^6 data realizations for each value of the SNR. Note: SNR was varied by keeping the variance in the signal constant, and modulating the signal intensity, not vice versa. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

some important insights. As shown in Figure 1c, for low SNR values, in both linear and nonlinear space, the skewness is nonzero with, respectively, a negative or positive sign. This nonzero skewness will have a significant impact on the parameter estimation as it is an indication that the error terms are not normally distributed. Consequently, the WLLS estimator is expected to have a lower precision than estimators that account for the proper data distribution, such as the maximum likelihood estimator. Another consequence of the nonzero skewness is a bias toward positive or negative residuals, depending on the sign of the skewness, for the detection of outliers due to the asymmetry of the distribution. This will be discussed in the Outlier Detection subsection.

The maximum likelihood estimator has asymptotically optimal properties with respect to accuracy and precision. However, the maximum likelihood estimator can only be derived if the true analytical expression of the data probability density function is known (30). Due to the manipulation of the magnitude images, such as motion or echo-planar imaging (EPI) correction, this is unfortunately rarely the case.

IRLLS estimator

In this subsection, our IRLLS algorithm will be explained in detail. The IRLLS framework is an improvement to the well-known RESTORE algorithm (18), which is a voxel-wise identification of outliers through the use of a robust iterative reweighting procedure. In addition, IRLLS aims at properly accounting for the data statistics after linearization of the diffusion parameter estimation. In Figure 2, both the RESTORE and IRLLS framework are shown. The main differences between both methods can be summarized as: (1) the use of linear estimators in IRLLS, (2) the introduction of a less biased way to identify positive and negative outliers, which is based on the data statistics, and (3) the use of studentized residuals. In the next four subsections, the main steps of the IRLLS algorithm will be discussed separately.

Weighted LLS Estimation

In IRLLS, the WLLS estimation plays an important role. It is used for both the initial tensor fit and for the final tensor estimation during which the outliers are excluded. The WLLS estimator of the parameter vector θ can be written in closed form as:

$$\hat{\boldsymbol{\theta}}_{\mathrm{WLLS}} = \left(\boldsymbol{B}^{\mathbf{T}}\boldsymbol{W}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^{T}\boldsymbol{W}\tilde{\mathbf{S}}, \qquad [4]$$

where $\boldsymbol{W} = \{w_{ij}\}$ is an $N \times N$ weight matrix. Salvador et al. (5) showed that, assuming uncorrelated measurement errors, the optimal weighting matrix $\boldsymbol{W}$ (in terms of minimum variance and bias of the estimates) is given by a diagonal matrix, the elements of which are the reciprocals of the measurements' error variances. For moderate values of SNR and a constant variance of the noise in the original complex data,

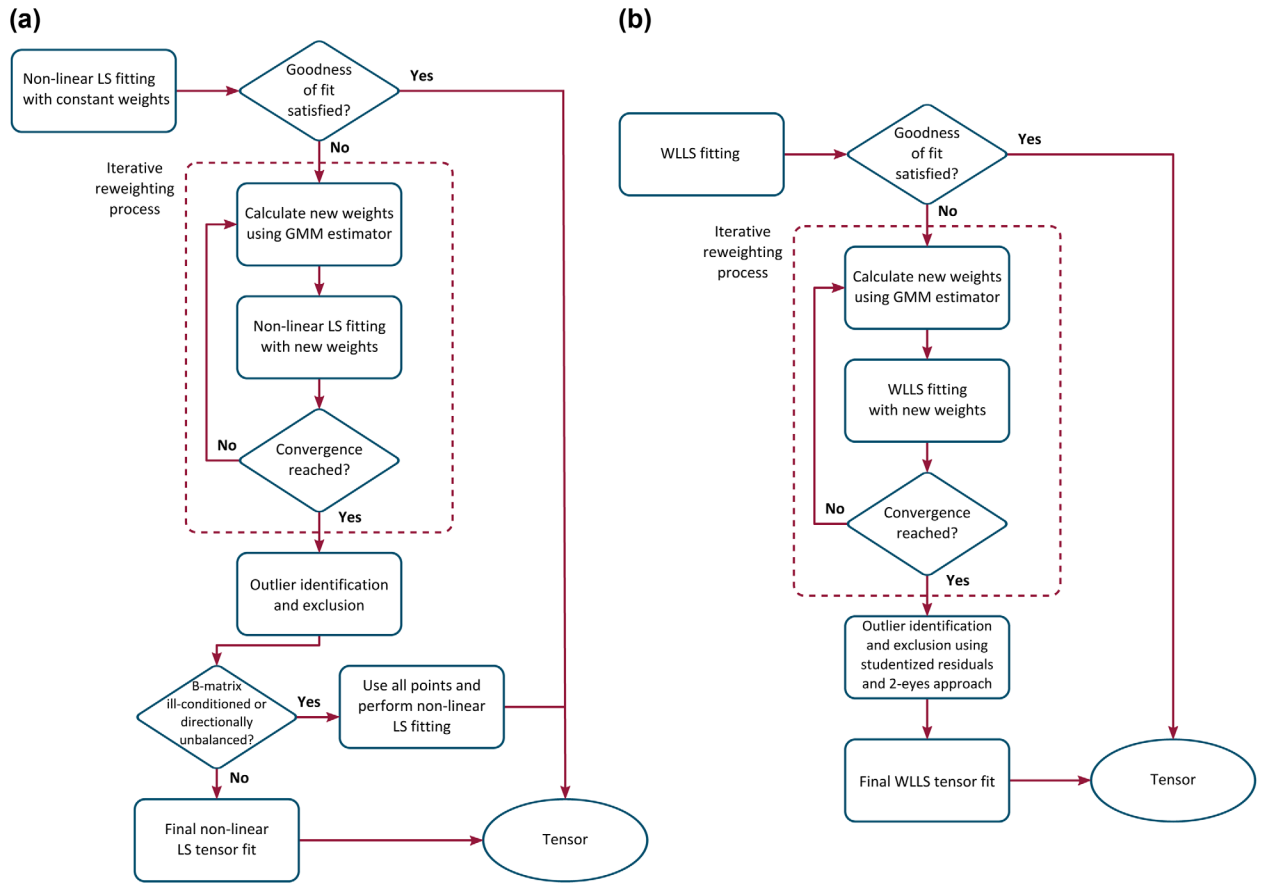

FIG. 2. Flow charts of RESTORE (**a**) and IRLLS (**b**). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

they have shown that the weights can be approximated by the squared LLS estimate of $S_i(\boldsymbol{b}_i;\boldsymbol{\theta})$ because it remains unbiased under inequality of variances (5):

$$w_{ii}=\exp\left(2\boldsymbol{b}_i\hat{\boldsymbol{\theta}}_{\mathrm{LLS}}\right), \quad [5]$$

where $\hat{\boldsymbol{\theta}}_{\mathrm{LLS}}$ is obtained from (4) by substituting the identity matrix for $\boldsymbol{W}$. From the estimated parameters $\hat{\boldsymbol{\theta}}$, the residuals e_i can be determined and are defined as the difference between a measured signal $\tilde{S}_i$ and its estimate $\hat{S}_i$. However, because the parameters can be estimated in two spaces, linear and nonlinear, the residuals can also be defined in both spaces:

$$e_i(\hat{\boldsymbol{\theta}}) = \tilde{S}_i(\boldsymbol{b}_i;\boldsymbol{\theta}) - \exp\left(\boldsymbol{B}\hat{\boldsymbol{\theta}}\right), \quad [6]$$

and

$$e_i^*(\hat{\boldsymbol{\theta}}) = \log\left(\tilde{S}_i(\boldsymbol{b}_i;\boldsymbol{\theta})\right) - \boldsymbol{B}\hat{\boldsymbol{\theta}}. \quad [7]$$

Analogous to the error terms, the notation of the residuals in the nonlinear and linear spaces is denoted by e_i and e_i^*, respectively.

Goodness-of-Fit

The initial WLLS fit is subsequently evaluated against a goodness-of-fit criterion. If the hypothesis that the model describes the data well cannot be rejected, no further outlier detection is performed and the WLLS estimation is the final one for this voxel. However, if the hypothesis is rejected, an iterative reweighting procedure is initiated. The goodness-of-fit criterion used in IRLLS consists of the evaluation of the reduced χ^2-statistic (31):

$$\chi^2_{\mathrm{red}} = \frac{1}{\nu}\sum_i \frac{e_i^2}{\hat{\sigma}^2}, \quad [8]$$

with $\nu = N - 7$ being the number of degrees of freedom of the distribution. Hereby 7 is the number of parameters in the DTI model and N the number of measurements. $\hat{\sigma}$ denotes the estimated standard deviation of the error terms:

$$\hat{\sigma} = 1.4826\sqrt{\frac{N}{\nu}}\mathrm{med}_i\left(\left|w_{ii}e_i^* - \mathrm{med}_j\left(w_{jj}e_j^*\right)\right|\right), \quad [9]$$

where the weights w_{ii} and residuals e_i^* are obtained from the WLLS estimation in Eq. (5). This method is largely based on the method proposed by Chang et al. (22). For normally distributed error terms, χ^2_{red} follows a χ^2-distribution with mean $\mu = 1$ and variance $\sigma^2 = 2/\nu$. By the central limit theorem, for large N, this distribution will approximate a normal distribution with the same mean and variance. The goodness-of-fit criterion consists of evaluating if the calculated reduced χ^2-statistic lies within the confidence interval $[1 - 3\sqrt{2/\nu}, 1 + 3\sqrt{2/\nu}]$.

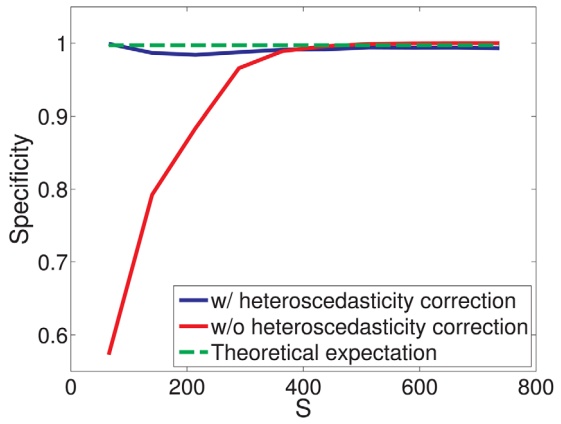

FIG. 3. Results of a Monte Carlo simulation of diffusion signals disturbed by Rician noise. No outliers were introduced. A simple outlier detection with and without a heteroscedasticity correction was performed. The specificity is shown for both cases in function of the signal strength S. Note that without a heteroscedasticity correction, the specificity of the outlier detection is substantially lower than the theoretical expectation for low signal strengths. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

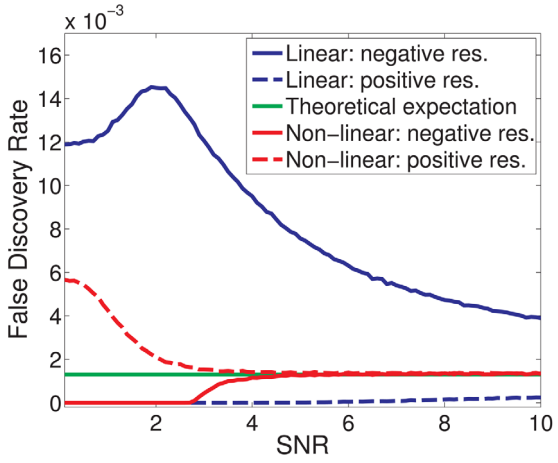

FIG. 4. The false discovery rate of a simple outlier detection method in linear (blue) and nonlinear (red) space is shown in function of the SNR. A distinction is made between false discoveries of negative (dashed line) and positive (full line) residuals. It is clear that depending on the choice of estimator, for low values of SNR, a significant bias will be introduced towards either negative or positive outliers. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Iterative Geman–McClure M-Estimation

When the goodness-of-fit criterion is not met, an iterative reweighting procedure is initiated. This procedure consists of minimizing the sum of function values $\rho(e_i^*)$, where the residuals e_i^* can be iteratively recalculated. $\rho(e_i^*)$ is chosen in such a way that it reduces the influence of potential outliers. This function should be symmetric, positive-definite and have a unique minimum at zero. One of the best suited functions is the Geman–McClure function (17):

$$\rho(e_i^*) = \frac{e_i^{*^2}}{\hat{\sigma}_i^{*^2} + e_i^{*^2}}, \qquad [10]$$

where $\hat{\sigma}_i^*$ denotes the estimate of the standard deviation of the distribution of the ith error term in the linear space:

$$\hat{\sigma}_i^* = \frac{\hat{\sigma}}{\hat{S}_i}. \qquad [11]$$

The factor $1/\hat{S}_i$ is needed to correct for the heteroscedasticity introduced using the log-transformation. The importance of this factor is visualized in Figure 3, where the specificities of a simple outlier detection method, with and without heteroscedasticity correction, are plotted in function of the signal strength S. When the sum over all $\rho(e_i^*)$ values is minimized with respect to the residuals e_i^*, it can be written as a weighted least squares estimation with weights defined as (32):

$$w_i = \frac{\hat{\sigma}_i^{*^2}}{(\hat{\sigma}_i^{*^2} + e_i^{*^2})^2}. \qquad [12]$$

These weights are now used in the WLLS estimation. Once the fit is performed, new residuals can be calculated to determine the new weights for the next iteration. In IRLLS, the L^2-norm of the difference vector of the unknown coefficients' vector $\hat{\mathbf{\theta}}$ between successive iterations is used to evaluate the convergence of the algorithm, unless a predefined maximum number of iterations get exceeded. When the L^2-norm reaches a predetermined threshold, the iterative procedure is assumed to have converged. In our work, we set these convergence criteria to a maximum of 25 iterations and the L^2-norm of the difference vector to be smaller than 0.1% of the L^2-norm of the estimated parameter vector $\hat{\mathbf{\theta}}$. Although there is no analytical proof of convergence, both simulations and real data experiments revealed that the maximum number of iterations criterion is rarely needed because only approximately 10 iterations are needed on average to reach convergence.

Outlier Detection

Once the iterative reweighting procedure has converged, the final residuals are examined. Each individual residual is verified as to whether it lies within a confidence interval of $[-3\sigma_i, 3\sigma_i]$, where σ_i denotes the noise level in either the nonlinear space ($\sigma_i = \hat{\sigma}$) or in the linear space ($\sigma_i = \hat{\sigma}_i^*$). If a residual lies outside this interval, the measurement associated with it is considered to be an outlier and is consequently excluded from the final standard WLLS tensor fit. An important issue that needs to be addressed is whether the residuals have to be examined in the linear space ($\boldsymbol{e}^*$) or in the nonlinear space ($\boldsymbol{e}$) because both will have different distributions. If these residuals are examined in the linear space, their distribution will be negatively skewed, causing the algorithm to be more sensitive to negative outliers if the SNR is low. In contrast, in the nonlinear space, it will be positively skewed and an outlier detection algorithm will consequently be more sensitive to positive outliers. This is also shown in Figure 4.

To exploit these properties for the detection of outliers, we propose a so-called two-eyes approach. This

approach distinguishes between positive and negative residuals during outlier detection. For the negative residuals, the outlier identification is performed on $\boldsymbol{e}^*$, while for positive residuals, $\boldsymbol{e}$ is examined. As a result, this combined detection method is sensitive to both negative and positive outliers because of the natural skewness of both distributions.

Finally, we also propose to studentize the residuals before outlier detection. This is motivated by the fact that because of the model fit, it is likely that the variance of the residuals varies significantly from one measurement to another, even though their error variance is the same. This is because some measurements have a higher influence on the model fit than others. Consequently, their residual variance will be smaller, and outliers on these measurements will become harder to distinguish from regular noise. Conversely, residuals having a larger variance will manifest more often as false discoveries when identifying outliers. To compensate for this effect, one has to studentize the residuals. This is performed by dividing the residuals e_i or e_i^* by an estimate of their standard deviation (33):

$$t_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}} \quad [13]$$

and

$$t_i^* = \frac{e_i^*}{\hat{\sigma}_i^*\sqrt{1 - h_{ii}}} \quad [14]$$

where t_i and t_i^* are the studentized residuals in nonlinear and linear space, respectively, and h_{ii} is called the leverage. The leverage of the ith residual can be obtained by calculating the ith diagonal element of the hat matrix $\boldsymbol{H}$ for uncorrelated errors (33):

$$\boldsymbol{H} = \boldsymbol{B}\left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}} \quad [15]$$

where $\boldsymbol{W}$ denotes the weight matrix used in the last Geman–McClure M-estimation iteration. The hat matrix maps the measured signal values to the fitted values.

Another notable advantage of having determined the leverages is that they provide an easy way to verify whether removing a certain measurement will result in an ill-conditioned or directionally unbalanced b-matrix. This is because measurements that are more important to the model fit (e.g., the nondiffusion-weighted measurement b_0 if only one is present) will have a much larger leverage. This way, measurements with leverage above a predetermined threshold (e.g., 0.9 in our implementation of IRLLS) can be excluded from the outlier identification process. This obviates the need for extra tests, as with RESTORE and iRESTORE, to check whether the b-matrix is ill-conditioned or directionally unbalanced.

Experiments

Simulations

To investigate the performance of IRLLS, Monte Carlo simulations of single diffusion tensors were performed under different conditions. In each simulation, IRLLS was compared to RESTORE, iRESTORE, WLLS, and an outlier free WLLS reconstruction, that is, regarded as the best achievable solution. Within the category of voxel-based outlier detection methods that make minimal assumptions about the origin of the artifacts and impose minimal constraint on the data acquisition, RESTORE and iRESTORE remain the default choices, making them our state-of-the-art reference. We implemented RESTORE and iRESTORE, and verified their performance by comparing their results with those from the original TORTOISE implementation (24). This way, a fairer comparison of calculation time could be made, and outlier maps could be generated. Furthermore, this allowed us to keep all internal parameters exactly the same for all methods, ensuring objective comparison.

In general, the following parameters were used to simulate measurements; however, in each simulation, one of the parameters was varied. Noise-free mean diffusivity (MD) and fractional anisotropy (FA) values were set to 0.8×10^{-3} mm^2/s and 0.85, respectively, to correspond with values measured in the corpus callosum. The acquisition design that was used included 5 non-DW signal (b_0) and 30 DW-MR signals with a b-value of 1000 s/mm^2 (b_{1000}). For the b_{1000} voxels, 30 different gradient directions were determined using electrostatic repulsion (34). For every simulation, 20,000 repetitions were performed. Six outliers were randomly introduced in the b_{1000} shell, each one being 50% of its true value. This choice was motivated by the fact that a loss of signal is more frequent than signal overshoots. Additionally, this makes it fairer to compare IRLLS to iRESTORE because the latter only assumes signal dropouts. Nevertheless, the number of outliers and their strengths were varied in simulations 1 and 3, respectively. After the introduction of the outliers, Rician noise was added with an SNR of 20, which was defined on the b_0-images.

In the first simulation, the total number of outliers was varied between 0 and 6 for both positive (150%) and negative (50%) outliers, while in simulation 3, the outlier strengths ranged from 0% (complete signal dropouts) up to 300% (extreme hyperintensities) of the true signal value. The effect of different SNR values was evaluated in simulation 2, where the SNR varied from 10 to 30. All the aforementioned outlier detection algorithms were highly dependent on a correct estimation of the noise level. In the first three simulations, the correct value of the noise level was used as opposed to an estimated value. However, in simulation 4, a bias on the true noise level was introduced. This factor ranged from 1 up to 3 times the true value. No negative bias was simulated because when outliers are present, an overestimation of the noise level is expected. The final tensor was estimated using WLLS, ignoring the outliers identified by the respective algorithms. All differences that were found between IRLLS and the other methods were evaluated using a Wilcoxon signed rank test.

Real Data Experiments

The outlier detection performance of IRLLS, RESTORE, and iRESTORE was tested on real data. A dataset of a neonate was studied knowing that such data often

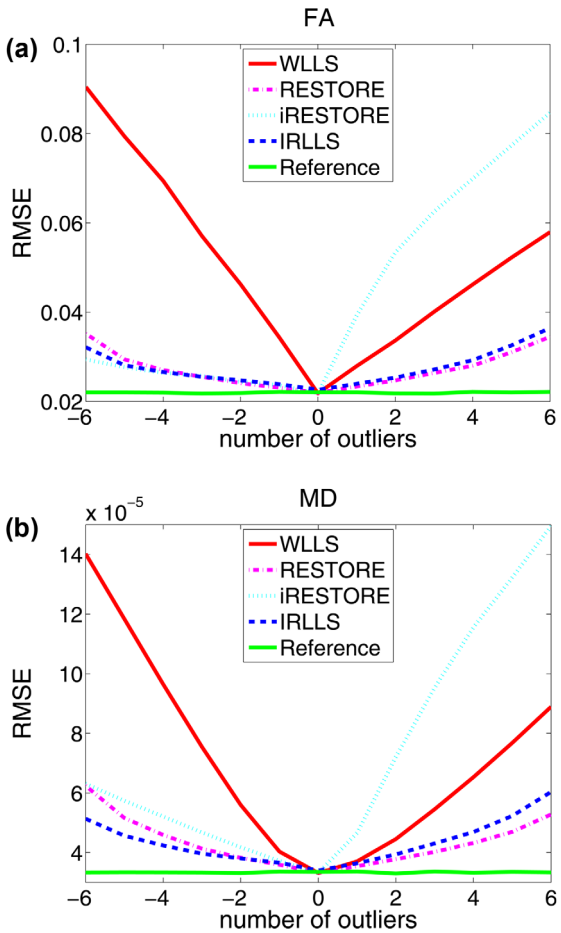

FIG. 5. The root mean squared error of FA (**a**) and MD (**b**) estimators in function of the number of the outliers for IRLLS, RESTORE, iRESTORE, WLLS, and WLLS estimation of the outlier free signal (reference). A negative number represent negative outliers (50% of true value), while positive numbers represent positive outliers (150% of true value). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

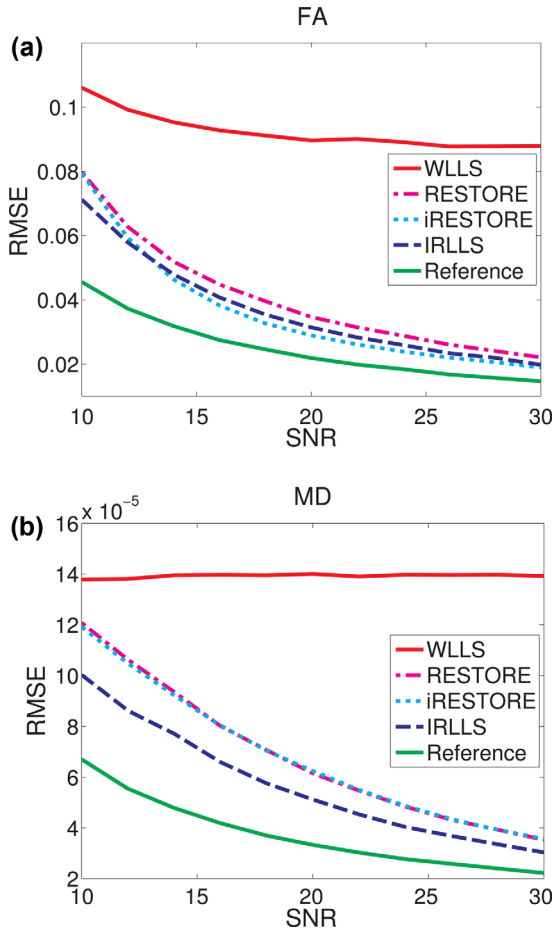

FIG. 6. The root mean squared error of FA (**a**) and MD (**b**) estimators in function of the SNR for IRLLS, RESTORE, iRESTORE, WLLS, and WLLS estimation of the outlier free signal (reference). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

suffers from various artifacts. The data were acquired on a 1.5T MRI Signa scanner using a Neonate Head Coil for reception. An EPI\spin echo (SE) DW pulse sequence was used with acquisition matrix 64 × 128, number of slices = 29, an interslice gap of 3 mm, echo time (TE) = 89 ms, pulse repetition time (TR) = 11,725 ms, averages = 1, and specific absorption rate (SAR) = 0.1172. The DW gradient settings that were used consisted of 1 b_0 image and 25 images with a b-value of 750 s/mm^2. The raw data were upsampled in k-space into a 256 × 256 matrix that resulted in a final voxel size of 0.625 × 0.625 × 3 mm^3. Homodyne reconstruction was used to reconstruct the partially Fourier-encoded data. The subject was a healthy female, nonsedated, 16-days-old neonate. Parental informed written consent was obtained.

RESULTS

Simulation Experiments

The results of the simulation experiments are presented in Figures 5–8. For all experiments, the root mean squared error (RMSE) of FA and MD measurements are plotted in subfigures (a) and (b), respectively. All of the reported differences were significant, with a significance level of 0.001 when using the Wilcoxon signed rank test. A Bonferroni correction for multiple comparisons was applied. The only exceptions were the two cases where no outliers were present (cf. in simulations 1 and 3 where the number of outliers is zero and the outlier strength is 1, respectively).

Simulation 1: Variable Number of Outliers

The results of the first simulation are presented in Figure 5. For any number of negative outliers, IRLLS, RESTORE, and iRESTORE, perform the same in terms of RMSE of FA, while IRLLS demonstrates a slightly lower RMSE of MD. In this range, they all perform substantially better than the standard WLLS in terms of RMSE of both FA and MD. For positive outliers, RESTORE appears to handle the outliers slightly better than IRLLS. iRESTORE on the contrary, performs even worse than the WLLS estimation. Also important to notice is the simulation where the number of outliers is zero. Here, none of the robust techniques show significant differences compared to the standard WLLS.

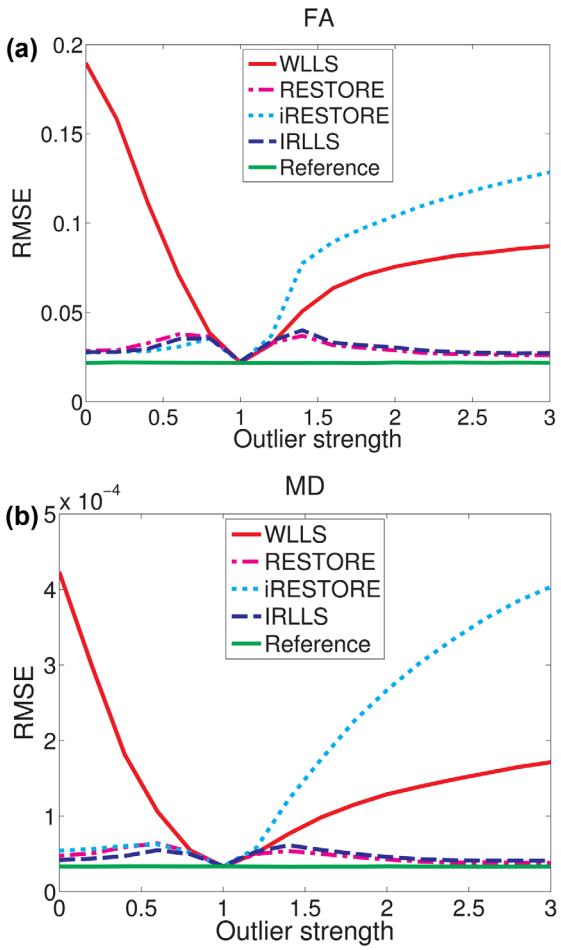

FIG. 7. The root mean squared error of FA (**a**) and MD (**b**) estimators in function of the outlier strength for IRLLS, RESTORE, iRESTORE, WLLS, and WLLS estimation of the outlier free signal (reference). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

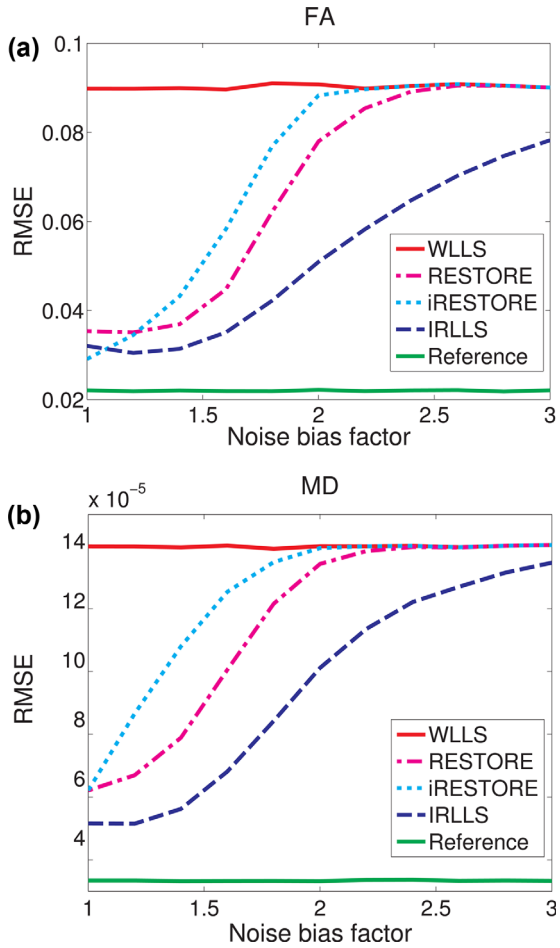

FIG. 8. The root mean squared error of FA (**a**) and MD (**b**) estimators in function of the noise bias factor for IRLLS, RESTORE, iRESTORE, WLLS, and WLLS estimation of the outlier free signal (reference). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Simulation 2: Variable SNR

In Figure 6, the results of simulation 2 are presented. In terms of the RMSE of the FA, all of the robust techniques perform approximately equally well and show a great improvement compared to a standard WLLS estimation. In terms of the MD estimation, however, IRLLS clearly performs best, especially for low SNR values.

Simulation 3: Variable Outlier Strength

Figure 7 shows the results of simulation 3. Six outliers were introduced, with strengths ranging from 0% up to 300% of the true value. As expected, iRESTORE demonstrates extremely bad results for positive outliers, while IRLLS and RESTORE perform very well over the entire range of outlier strengths.

Simulation 4: Variable Bias on Noise Level Estimation

The aforementioned robust techniques require an estimate of the noise standard deviation to exclude outliers. Although robust methods for the estimation of the noise variance have been proposed (35), most noise estimation methods are vulnerable to outliers as well. It is clear from the results in Figure 8 that over almost the entire range of noise level offsets, the RMSE of the IRLLS estimations of FA and MD are considerably lower than those of RESTORE or iRESTORE.

Real Data Experiments

The results from the real data experiments are shown in Figure 9. Because motion cannot be properly dealt with when scanning newborns and homodyne detection (36) was used to reconstruct the partially Fourier-encoded data, this dataset suffers from various artifacts, such as signal dropouts and worm-artifacts. Both artifacts may originate from the combination of homodyne reconstruction and a rolling motion of the head (37). From the acquired datasets, three interesting slices were selected, each one being represented in a different row in Figure 9. The first slice shows a signal increase on the left side of the brain image. In the second row of Figure 9, a magnitude image is shown with the aforementioned typical worm-artifacts, spread out over the entire image. The last artifact that will be discussed is a signal dropout. In the

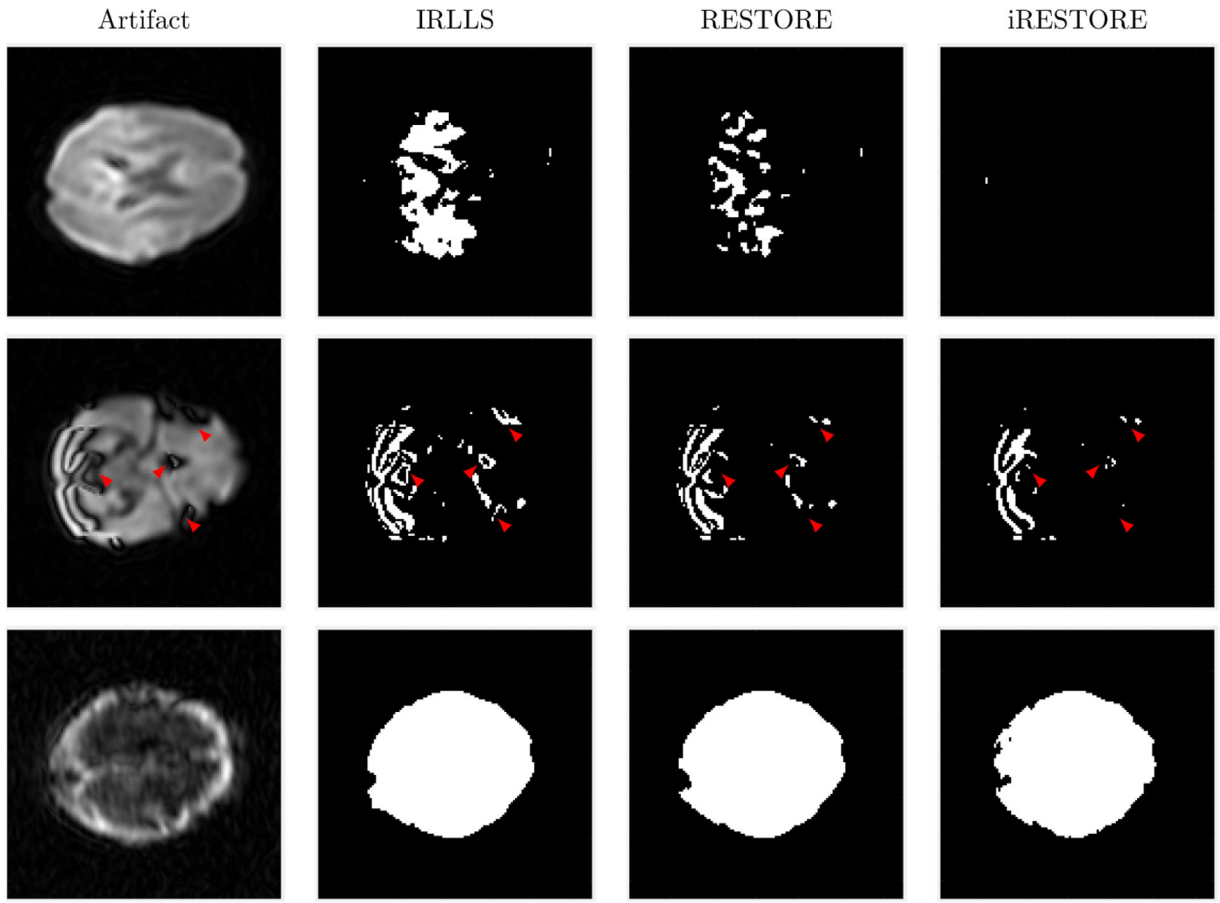

FIG. 9. The results of real data experiments are shown. Three artifacts from the real data experiments are selected, each of which is represented in one row. The first column shows the magnitude data, while the second, third, and fourth column show the outliers detected by IRLLS, RESTORE, and iRESTORE, respectively. The red arrows in the images of the second artifact denote four areas where IRLLS is able to better resolve an artifact when compared to RESTORE and iRESTORE. The background of all of the images was cropped to show the relevant data more clearly.

second column, all outliers detected by IRLLS are visualized, whereas in the third and fourth column the outliers detected by the RESTORE and iRESTORE algorithm, respectively, are shown.

For the first artifact, one of iRESTORE's limitations is immediately clear. The large signal intensity in the left side of the image was detected by both IRLLS and RESTORE, while iRESTORE did not classify it as an outlier region. It is also interesting to note that IRLLS has a higher outlier detection rate compared to RESTORE. In the second magnitude image, a worm-artifact is present. All techniques are rather successful in detecting this artifact, although, as indicated by the red arrows in the second row Figure 9, IRLLS succeeds in identifying noticeably more erroneous data points compared to both RESTORE and iRESTORE. In the bottom row, a complete signal dropout is shown. In this case, all of the techniques are equally successful in detecting this artifact. One of the main advantages of IRLLS becomes clear when comparing calculation times. Thanks to the use of linear estimators, the calculation time of IRLLS was around an order of magnitude faster compared to RESTORE and iRESTORE for this particular dataset.

DISCUSSION

In this work, we introduced IRLLS, a method based on the well-known RESTORE algorithm, which is substantially accelerated by the use of the fast WLLS estimator. While RESTORE is based on a Geman–McClure weighted NLS estimation to identify outliers, IRLLS uses an iterative WLLS estimation with the same Geman–McClure weights. The motivation behind this adaptation is the large gain in calculation speed and an increase in the accuracy of the parameter estimators. It has been empirically shown that the iterative process converges and rarely requires the maximum number of iterations criteria. To develop a method that is sensitive to artifacts that manifest either as a loss or gain in signal intensity, a two-eyes approach was introduced. This approach involves evaluating negative and positive residuals in either the linear or nonlinear space, respectively. In our method, we decided not to rely on the actual probability density function of the data but instead to use a more pragmatic approach. This is motivated by the fact that the data probability density function is generally unknown because of, for example, motion correction, or the use of parallel imaging techniques such as

generalized autocalibrating partially parallel acquisition (38). Finally, we also introduced studentized residuals to compensate for differences in residual variances that could have been introduced through the model fit and to ensure that the *b*-matrix cannot be ill-conditioned or directionally unbalanced.

IRLLS was tested in both an extensive Monte Carlo simulation framework and on a real dataset. Simulations 1 and 3 show that over a broad range of different numbers and strengths of outliers, IRLLS, RESTORE, and iRESTORE perform more or less equally well for negative outliers and demonstrated great improvement compared to a WLLS estimation. For positive outliers, IRLLS and RESTORE present better results than iRESTORE, which because of its design, performed significantly worse than the other estimators. It is also important to note that the use of IRLLS has no harmful effects when no outliers are present. Consequently, IRLLS can also be used in datasets that are considered "outlier free" without any significant loss in quality of the derived diffusion parameters. In simulation 2, all estimation methods were tested for different values of the SNR. At low SNR, IRLLS outperforms the other estimators, showing a significantly lower RMSE on MD estimations while the RMSE of FA estimations remains approximately the same for all methods. This can be explained by the fact that for low SNR values, the data distributions will be more skewed. Because IRLLS operates in the log-Rician space for negative residuals, it partially compensates for this positive skewness in the nonlinear space.

A downside of the previously mentioned robust estimation techniques is that they all rely on a correct estimation of the noise level (39–45). Obtaining good estimates, however, is not obvious, and an overestimation is to be expected, especially in the presence of artifacts. Simulation 4 inspects the effect of an overestimation of the noise level on the performance of the aforementioned estimators. The results from this simulation indicate that IRLLS is far more robust, compared to RESTORE or iRESTORE, when large biases of the noise level are present. Both for the RMSE of FA and MD, the difference between IRLLS and the other estimators are substantial. We believe the increased robustness to erroneous noise level estimates and low SNR, together with the substantially reduced calculation time, are the biggest incentives to favor IRLLS over RESTORE and iRESTORE.

Judging from the results of the real data experiments, our conclusions from the simulations seem to be confirmed. Both the simulation and real data experiments demonstrate that IRLLS outperforms RESTORE and iRESTORE in the outlier detection success rate. Specifically, by design, iRESTORE cannot detect erroneous signal increases, while our experiment did not show an improved sensitivity to signal dropouts. The improved sensitivity of IRLLS to signal dropouts compared to RESTORE roots in the skewness of the error terms resulting from the log-transformation needed to linearize the DTI (or diffusion kurtosis imaging) model. The linearization of those diffusion models allows the estimation of the model parameters using a (W)LLS estimator that has a closed form solution. Therefore, unlike iterative nonlinear strategies, the linear estimators are computationally efficient and not prone to getting stuck in a local optimum. Although it is clear that IRLLS outperforms RESTORE and iRESTORE in computation time, a quantitative comparison in calculation time between IRLLS and RESTORE is beyond the scope of this work as it includes many implementation variables such as convergence criteria and optimization algorithms.

We would like to stress that the IRLLS is basically an outlier detection algorithm rather than a robust estimator. The resulting outlier map is used to detect the voxels that should be discarded in the final tensor estimation step. We propose to use the WLLS estimator as described by Salvador et al. (5) and Veraart et al. (6) as a final estimation strategy because of its inherent high accuracy given Rician-distributed data. However, other strategies, such as the NLS estimator, constrained NLS estimator (7) (e.g., to impose the positive definite constraint on the diffusion tensor), or maximum likelihood estimator (30), can be applied as well. For comparison of the different estimation approaches, we would like to refer the interested reader to the literature (4, 46).

In conclusion, a novel method for the robust estimation of diffusion parameters, called IRLLS, was introduced. IRLLS uses the features of the data distributions which led to a two-eyes approach for the detection of outliers. Furthermore, studentized residuals were introduced to compensate for differences in residual variances and to ensure that the *b*-matrix does not become ill-conditioned or directionally unbalanced through the removal of outliers. IRLLS was evaluated and compared to WLLS, RESTORE, and iRESTORE in both a Monte Carlo simulation framework and real data experiments. These experiments showed that our proposed method is substantially faster, more robust against inaccuracies of the noise level estimate and provides more accurate estimates of MD for low values of SNR. Moreover, in every simulation, IRLLS proved to be at least as reliable as the other methods in terms of RMSE of FA and MD estimations. These combined features make IRLLS a practical and reliable tool for the robust estimation of DW-MR parameters.

ACKNOWLEDGMENTS

The authors thank Annemarie Plaisier and Jeroen Dudink of the Division of Neonatology, Erasmus MC Rotterdam (The Netherlands) for the acquisition of the research data.

REFERENCES

1. Basser PJ, Mattiello J, Le Bihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. J Magn Reson 1994;103: 247–254.
2. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. Magn Reson Med 2005;53: 1432–1440.
3. Tabesh A, Jensen JH, Ardekani BA, Helpern JA. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. Magn Reson Med 2011;65:823–836.
4. Koay CG. Least squares approaches to diffusion tensor estimation. In: Jones D, editor. Diffusion MRI. New York: Oxford University Press; 2011, pp. 281–343.
5. Salvador R, Peña A, Menon DK, Carpenter TA, Pickard JD, Bullmore ET. Formal characterization and extension of the linearized diffusion tensor model. Hum Brain Mapp 2005;24:144–155.

6. Veraart J, Sijbers J, Leemans A, Jeurissen B. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. NeuroImage 2013;81:335–346.
7. Koay CG, Chang LC, Carew JD, Pierpaoli C, Basser PJ. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. J Magn Reson 2006;182:115–125.
8. Pannek K, Raffelt D, Bell C, Mathias JL, Rose SE. HOMOR: higher order model outlier rejection for high b-value MR diffusion data. NeuroImage 2012;63:835–842.
9. Jiang H, Chou MC, van Zijl PC, Mori S. Outlier detection for diffusion tensor imaging by testing for ADC consistency. In Proceedings of the 17th Annual Meeting of ISMRM, Honolulu, Hawaii, USA, 2009. p. 1404.
10. Niethammer M, Bouix S, Aja-fernández S, Westin Cf, Shenton ME. Outlier rejection for diffusion weighted imaging. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), London, UK, 2007. pp. 161–168.
11. Sharman MA, Cohen-Adad J, Descoteaux M, Messé A, Benali H, Lehericy S. Impact of outliers on diffusion tensor and Q-ball imaging: clinical implications and correction strategies. J Magn Reson Imaging 2011;33:1491–1502.
12. Maximov II, Grinberg F, Shah NJ. Robust tensor estimation in diffusion tensor imaging. J Magn Reson 2011;213:136–144.
13. Zhou Z, Liu W, Cui J, et al. Automated artifact detection and removal for improved tensor estimation in motion-corrupted DTI data sets using the combination of local binary patterns and 2D partial least squares. Magn Reson Imaging 2011;29:230–242.
14. Morris D, Nossin-Manor R, Taylor MJ, Sled JG. Preterm neonatal diffusion processing using detection and replacement of outliers prior to resampling. Magn Reson Med 2011;66:92–101.
15. Zwiers MP. Patching cardiac and head motion artefacts in diffusion-weighted images. NeuroImage 2010;53:565–575.
16. Mangin JF, Poupon C, Clark C, Le Bihan D, Bloch I. Distortion correction and robust tensor estimation for MR diffusion imaging. Med Image Anal 2002;6:191–198.
17. Geman S, McClure DE. Statistical methods for tomographic image reconstruction. Bull Int Stat Inst 1987;52:5–21.
18. Chang LC, Jones DK, Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. Magn Reson Med 2005;53:1088–1095.
19. Forde NJ, Ronan L, Suckling J, et al. Structural neuroimaging correlates of allelic variation of the bdnf val66met polymorphism. NeuroImage 2014;90:280–289.
20. Johnson RT, Yeatman JD, Wandell BA, Buonocore MH, Amaral DG, Nordahl CW. Diffusion properties of major white matter tracts in young, typically developing children. NeuroImage 2014;88:143–154.
21. Travers BG, Bigler ED, Tromp DP, et al. Longitudinal processing speed impairments in males with autism and the effects of white matter microstructure. Neuropsychologia 2014;53:137–145.
22. Chang LC, Walker L, Pierpaoli C. Informed RESTORE: a method for robust estimation of diffusion tensor from low redundancy datasets in the presence of physiological noise artifacts. Magn Reson Med 2012;68:1654–1663.
23. Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn Reson Med 2002;48:577–582.
24. Pierpaoli C, Walker L, Irfanoglu MO, et al. Tortoise: an integrated software package for processing of diffusion MRI data. In ISMRM 18th annual meeting, Stockholm, Sweden, 2010. p. 1597.
25. Henkelman RM. Measurement of signal intensities in the presence of noise in MR images. Med Phys 1985;12:232–233.
26. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magn Reson Med 1995;34:910–914.
27. Dietrich O, Raya JG, Reeder SB, Ingrisch M, Reiser MF, Schoenberg SO. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. Magn Reson Imaging 2008;26:754–762.
28. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. Biophys J 1994;66:259–267.
29. Koay CG, Basser PJ. Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. J Magn Reson 2006;179:317–322.
30. Sijbers J, den Dekker AJ, Scheunders P, Van Dyck D. Maximum-likelihood estimation of Rician distribution parameters. IEEE Trans Med Imaging 1998;17:357–361.
31. Taylor JR. An introduction to error analysis. Sausalito, CA: University Science Books, 1997.
32. Li G. Robust regression. In: Hoaglin DC, Mosteller F, Tukey JW, editors. Exploring data tables, trends, and shapes. New York: Wiley; 1985, pp. 281–343.
33. Cook RD, Weisberg S. Residuals and influence in regression. New York: Chapman and Hall; 1982.
34. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn Reson Med 1999;42:515–525.
35. Sijbers J, Poot D, den Dekker AJ, Pintjens W. Automatic estimation of the noise variance from the histogram of a magnetic resonance image. Phys Med Biol 2007;52:1335–1348.
36. Noll DC, Nishimura DG, Macovski A. Homodyne detection in magnetic resonance imaging. IEEE Trans Med Imaging 1991;10:154–163.
37. Bammer R, Holdsworth SJ, Aksoy M, Skare ST. Phase errors in diffusion-weighted imaging. In: Jones DK, editor. Diffusion MRI, Oxford: Oxford University Press Inc.; 2011, pp. 232–235.
38. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002;47:1202–1210.
39. Chang LC, Rohde GK, Pierpaoli C. An automatic method for estimating noise-induced signal variance in magnitude-reconstructed magnetic resonance images. In SPIE Medical Imaging. International Society for Optics and Photonics, San Diego, California, USA, volume 5747, 2005. pp. 1136–1142.
40. Aja-Fernández S, Tristán-Vega A, Alberola-López C. Noise estimation in single- and multiple-coil magnetic resonance data based on statistical models. Magn Reson Imaging 2009;27:1397–1409.
41. Koay CG, Ozarslan E, Pierpaoli C. Probabilistic identification and estimation of noise (PIESNO): a self-consistent approach and its applications in MRI. J Magn Reson 2009;199:94–103.
42. Walker L, Chang LC, Koay CG, Sharma N, Cohen L, Verma R, Pierpaoli C. Effects of physiological noise in population analysis of diffusion tensor MRI data. NeuroImage 2011;54:1168–1177.
43. Rajan J, Veraart J, Van Audekerke J, Verhoye M, Sijbers J. Nonlocal maximum likelihood estimation method for denoising multiple-coil magnetic resonance images. Magn Reson Imaging 2012;30:1512–1518.
44. Maximov II, Farrher E, Grinberg F, Jon Shah N. Spatially variable rician noise in magnetic resonance imaging. Med Image Anal 2012;16:536–548.
45. Veraart J, Rajan J, Peeters RR, Leemans A, Sunaert S, Sijbers J. A comprehensive framework for accurate diffusion MRI parameter estimation. Magn Reson Med 2013;81:972–984.
46. Jones DK, Basser PJ. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn Reson Med 2004;52:979–993.