Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 182 (2006) 115–125
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging

Cheng Guan Koay [a,*], Lin-Ching Chang [a], John D. Carew [b], Carlo Pierpaoli [a], Peter J. Basser [a]

[a] National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA
[b] Department of Statistics, University of Wisconsin, Madison, USA

Received 1 March 2006; revised 13 June 2006
Available online 7 July 2006

Abstract

A unifying theoretical and algorithmic framework for diffusion tensor estimation is presented. Theoretical connections among the least squares (LS) methods, (linear least squares (LLS), weighted linear least squares (WLLS), nonlinear least squares (NLS) and their constrained counterparts), are established through their respective objective functions, and higher order derivatives of these objective functions, i.e., Hessian matrices. These theoretical connections provide new insights in designing efficient algorithms for NLS and constrained NLS (CNLS) estimation. Here, we propose novel algorithms of full Newton-type for the NLS and CNLS estimations, which are evaluated with Monte Carlo simulations and compared with the commonly used Levenberg–Marquardt method. The proposed methods have a lower percent of relative error in estimating the trace and lower reduced χ^2 value than those of the Levenberg–Marquardt method. These results also demonstrate that the accuracy of an estimate, particularly in a nonlinear estimation problem, is greatly affected by the Hessian matrix. In other words, the accuracy of a nonlinear estimation is algorithm-dependent. Further, this study shows that the noise variance in diffusion weighted signals is orientation dependent when signal-to-noise ratio (SNR) is low ($\leqslant 5$). A new experimental design is, therefore, proposed to properly account for the directional dependence in diffusion weighted signal variance.
Published by Elsevier Inc.

Keywords: Newton's method; Levenberg–Marquardt; DTI; Diffusion tensor; Tensor estimation; Hessian

1. Introduction

Diffusion tensor imaging (DTI) is a novel noninvasive technique capable of providing important information about biological structures in the brain [1–4]. This technique depends upon accurate and precise estimation of the diffusion tensor. The mathematical framework for diffusion tensor estimation is both elegant and simple [1,4]. Its simplicity is due in part to the fact that the model is transformably linear [5]. However, the diffusion tensor in its original form as derived from *first principles* is a nonlinear model. Recent DTI studies have used several different models—from linear to nonlinear, and from unconstrained to constrained [4,6–12].

In general, the methods of estimation in DTI can be classified as linear least squares (LLS), weighted linear least squares (WLLS), nonlinear least squares (NLS) and their corresponding constrained counterparts, which will be denoted as CLLS, CWLLS and CNLS, respectively [4,6–12]. The constraint employed in the CLLS, CWLLS, and CNLS estimations is generally the positive definite constraint [11,12], i.e., the requirement that every eigenvalue of the diffusion tensor estimate be positive. The statistical comparison among different methods of diffusion tensor estimation, both unconstrained and constrained, has been studied in [12]. In the present study, we present a

* Corresponding author. Fax: +1 301 435 5035.
E-mail address: guankoac@mail.nih.gov (C.G. Koay).

1090-7807/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jmr.2006.06.020

theoretical and algorithmic framework for methods of estimation in DTI by investigating the properties of various least squares objective functions.

There are several numerical methods for solving the NLS problem in DTI. Yet, the Levenberg–Marquardt's (LM) approach has been the method of choice, perhaps, due to its simple implementation. This simplicity is due in part to its approximation to the Hessian matrix of the NLS objective function. Another approach is Newton's method (or full Newton-type method) where the complete Hessian matrix is required in the estimation process. It is well known that Newton's method is more robust than the LM method and can speed up convergence in NLS problems [13,14], but the complete Hessian matrix is often not available or known for a given problem. Fortunately, a previous account has shown that this is not the case in DTI [15]. In this study, we will show that the Hessian matrices for various methods of estimation in DTI have simple and compact forms.

We first review the basic estimation problem in DTI and discuss various least squares approaches for solving the problem. We then establish theoretical connections among the LLS, WLLS and NLS methods and among their constrained counterparts. We also derive all the Hessian matrices for the methods of estimation discussed in this paper. We propose an efficient strategy, which will be called Modified Full Newton's method (MFN), for solving both the NLS and CNLS problem. This strategy entails using the WLLS solution as the initial guess, adjusting the LM parameter, and incorporating the full Hessian matrix of the NLS objective function. A similar strategy is also adapted for solving the CNLS problem in DTI.

The performance of the proposed method is compared with the LM method using Monte Carlo simulations. The robustness and accuracy of the MFN method is assessed with respect to the LM method in terms of percent relative error in the estimated trace and reduced χ^2 value. The simulations are also used to assess the validity of the assumption of constant noise variance in a single voxel. The analysis and the results of this study provide new insights in constructing more appropriate experimental designs in which the direction-dependent noise variance is taken into account in the diffusion tensor estimation.

2. Materials and methods

2.1. Review of DTI estimation

In a DT-MRI experiment, the measured signal in a single voxel has the following form [1,4,16]:

$$s = S_0 \exp(-b\mathbf{g}^{\mathrm{T}}\mathbf{D}\mathbf{g}), \tag{1}$$

where measured signal, s, depends on the diffusion encoding gradient vector, $\mathbf{g}$, of unit length, the diffusion weight, b, the reference signal, S_0, and the diffusion tensor, $\mathbf{D}$. The symbol "T" denotes the matrix or vector transpose. Given $m \geqslant 7$ sampled signals based on at least six noncollinear gradient directions and at least one sampled reference signal, the diffusion tensor estimate can be found by minimizing different objective functions. To facilitate our theoretical investigation, the objective functions for the LLS, WLLS and NLS problems are defined as follows:

$$f_{\mathrm{LLS}}(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{7} W_{ij}\gamma_j\right)^2 = \frac{1}{2}\sum_{i=1}^{m} F_i^2, \tag{2}$$

$$f_{\mathrm{WLLS}}(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{m}\omega_i^2\left(y_i - \sum_{j=1}^{7} W_{ij}\gamma_j\right)^2 = \frac{1}{2}\sum_{i=1}^{m}\omega_i^2 F_i^2, \tag{3}$$

and

$$\begin{aligned} f_{\mathrm{NLS}}(\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{i=1}^{m}\left(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{m}(s_i - \hat{s}_i(\boldsymbol{\gamma}))^2 = \frac{1}{2}\sum_{i=1}^{m} r_i^2. \end{aligned} \tag{4}$$

The various symbols shown above are defined as:

$i = 1,\ldots,$ m,
$s_i =$ the measured diffusion weighted signal with noise,
$\hat{s}_i(\boldsymbol{\gamma}) = \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right] =$ the diffusion weighted function at $\boldsymbol{\gamma}$,
$\omega_i =$ the weights for the WLLS objective function,
$F_i = y_i - \sum_{j=1}^{7} W_{ij}\gamma_j$
is the error term for the LLS objective function,
$r_i(\boldsymbol{\gamma}) = s_i - \hat{s}_i(\boldsymbol{\gamma})$
is the error term for the NLS objective function,
$y_i = \ln(s_i)$,

$$\mathbf{W} = \begin{pmatrix} 1 & -b_1 g_{1x}^2 & -b_1 g_{1y}^2 & -b_1 g_{1z}^2 & -2b_1 g_{1x}g_{1y} & -2b_1 g_{1y}g_{1z} & -2b_1 g_{1x}g_{1z} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -b_m g_{mx}^2 & -b_m g_{my}^2 & -b_m g_{mz}^2 & -2b_m g_{mx}g_{my} & -2b_m g_{my}g_{mz} & -2b_m g_{mx}g_{mz} \end{pmatrix}$$

is a $m \times 7$ design matrix, and
$\boldsymbol{\gamma} = [\ln(S_0), D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yz}, D_{xz}]^{\mathrm{T}}$ is the parameter vector.

In general, the diffusion tensor is assumed to be symmetric positive definite—in other words the eigenvalues of the diffusion tensor have to be real and positive. By definition of the design matrix, $\mathbf{W}$, the diffusion tensor estimate is guaranteed to be symmetric but not positive definite. The positive definite condition requires more elaborate constraints on the diffusion tensor parameter vector, $[D_{xx},\cdots,D_{xz}]^{\mathrm{T}}$. A typical approach is to apply the Cholesky parametrization to $\mathbf{D}$ [17,11,12]. The Cholesky parametrization states that if $\mathbf{U}$ is an upper triangular matrix with nonzero diagonal elements

$$\mathbf{U} = \begin{pmatrix} \rho_2 & \rho_5 & \rho_7 \\ 0 & \rho_3 & \rho_6 \\ 0 & 0 & \rho_4 \end{pmatrix} \tag{5}$$

and $\mathbf{D} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ then $\mathbf{D}$ will be a symmetric positive definite matrix. Consequently, the parameter vector, $\boldsymbol{\gamma}$, may be

written as a vector-valued function of $\boldsymbol{\rho} = [\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_7]^{\mathrm{T}}$ so that:

$$\boldsymbol{\gamma}(\boldsymbol{\rho}) = [\rho_1, \rho_2^2, \rho_3^2 + \rho_5^2, \rho_4^2 + \rho_6^2 + \rho_7^2, \rho_2\rho_5, \rho_3\rho_6 + \rho_5\rho_7, \rho_2\rho_7]^{\mathrm{T}} \tag{6}$$

Rewriting Eqs. (2)–(4) in terms of $\boldsymbol{\rho}$, we have,

$$f_{\mathrm{CLLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \frac{1}{2}\sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\rho})\right)^2, \tag{7}$$

$$f_{\mathrm{CWLLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \frac{1}{2}\sum_{i=1}^{m}\omega_i^2\left(y_i - \sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\rho})\right)^2 \tag{8}$$

and

$$f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \frac{1}{2}\sum_{i=1}^{m}\left(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\rho})\right]\right)^2, \tag{9}$$

respectively, for the constrained estimations. Note that the CLLS and the CWLLS objective functions are no longer linear with respect to the new variables $\boldsymbol{\rho}$. The naming convention adopted here for the constrained LLS and WLLS methods is for convenience rather than technical correctness.

2.1.1. Theoretical connections among the least squares methods: zeroth order

Without loss of generality, we will focus on the unconstrained methods of estimation in this section. The goal of this section is to establish connections among the LLS, WLLS and NLS objective functions via the error terms defined above, and to understand the assumptions needed to arrive at the LLS and WLLS objective functions from the NLS objective function. It can be shown that Eq. (4) can be written as:

$$\begin{aligned} f_{\mathrm{NLS}}(\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{i=1}^{m} s_i^2(1 - \exp[-F_i])^2 \\ &= \frac{1}{2}\sum_{i=1}^{m}\hat{s}_i^2(\exp[+F_i] - 1)^2. \end{aligned} \tag{10}$$

The derivation of Eq. (10) is shown in Appendix A. Eq. (10) exhibits a certain symmetry when the error term, F_i, is small. Assuming $|F_i| \ll 1$, we take the first order Taylor expansion of $\exp[-F_i] \cong 1 - F_i$ and of $\exp[+F_i] \cong 1 + F_i$, so Eq. (10) can be approximated as

$$f_{\mathrm{WLLS}}(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{m} s_i^2 F_i^2 = \frac{1}{2}\sum_{i=1}^{m}\hat{s}_i^2 F_i^2, \tag{11}$$

which gives us two formulae analogous to the WLLS objective function in Eq. (3). Eq. (11) indicates that the weights, s_i and $\hat{s}_i$, are equally appropriate when the error F_i is small. Therefore, the observed diffusion weighted signals can be used as weights for the WLLS method. The use of diffusion weighted signals as the weights for the WLLS objective function has been previously proposed on different theoretical grounds by Salvador et al. [10] and by Basser et al. [4].

If we assume s_i's in Eq. (11) are approximately equal to some constant, C, then the WLLS objective function can be reduced to the LLS objective function by setting the constant to unity. Therefore,

$$f_{\mathrm{LLS}}(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{m} F_i^2. \tag{12}$$

The restrictive and physically implausible assumption needed to arrive at Eq. (12) from Eq. (11) clearly shows the inadequacy of the LLS method.

2.1.2. The theoretical connections among the least squares methods: higher order

In this section, we will present a higher order expression of the objective function for all the methods of estimation discussed above. Explicit expressions for the Hessian matrix, the Jacobian matrix, and the gradient vector for the NLS method will be presented first but the derivations of these expressions will be provided in Appendix B. Expressions for the Hessian matrix and the gradient vector of the NLS objective function have simple connections to those of the WLLS and LLS objective functions based on the analysis presented in Section 2.1. In the NLS method, the Hessian matrix, the transpose of the Jacobian matrix, and the gradient vector can be written as:

$$\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \mathbf{W}^{\mathrm{T}}(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}, \tag{13}$$

$$\mathbf{J}^{\mathrm{T}}(\boldsymbol{\gamma}) = -(\hat{\mathbf{S}}\mathbf{W})^{\mathrm{T}}, \tag{14}$$

and

$$\nabla f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = -(\hat{\mathbf{S}}\mathbf{W})^{\mathrm{T}}\mathbf{r}(\boldsymbol{\gamma}), \tag{15}$$

respectively; where the Hessian matrix is defined as $[\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})]_{ij} \equiv \frac{\partial^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j}$ and the matrix, $\mathbf{S}$, is a diagonal matrix whose nonzero elements are the measured diffusion weighted signals:

$$\mathbf{S} = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_m \end{pmatrix}. \tag{16}$$

Similarly, $\mathbf{R}$ and $\hat{\mathbf{S}}$ are diagonal matrices whose nonzero elements are the diffusion weighted functions and the error terms evaluated at $\boldsymbol{\gamma}$, respectively:

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{s}_1(\boldsymbol{\gamma}) & & \\ & \ddots & \\ & & \hat{s}_m(\boldsymbol{\gamma}) \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} r_1(\boldsymbol{\gamma}) & & \\ & \ddots & \\ & & r_m(\boldsymbol{\gamma}) \end{pmatrix}. \tag{17}$$

We shall derive the same higher order information for the WLLS and LLS methods from the NLS Hessian matrix as follows:

(I) $\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \mathbf{W}^{\mathrm{T}}(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}$;
(II) $\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) \cong \mathbf{W}^{\mathrm{T}}\hat{\mathbf{S}}^2\mathbf{W}$ if $\mathbf{R} \cong \mathbf{0}$;

(III) $\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) \cong \mathbf{W}^{\mathrm{T}}\mathbf{S}^2\mathbf{W}$ if $\mathbf{S} \cong \hat{\mathbf{S}}$, similar to the assumption used in Eq. (11);
(IV) $\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) \cong \mathbf{W}^{\mathrm{T}}\mathbf{W}$ if $\mathbf{S} \cong \mathbf{I}$, similar to the assumption used in Eq. (12).

In deriving (II) from (I), we have assumed that the error matrix, $\mathbf{R}$, is close to zero. If we further assume that $\mathbf{S} \cong \hat{\mathbf{S}}$, then we have the Hessian matrix for the WLLS method as is shown in (III). Pushing a step further by assuming $\mathbf{S} \cong \mathbf{I}$, we then arrive at the Hessian matrix of the LLS method, which is in (IV).

For completeness, the Hessian matrices and the gradient vectors for the WLLS and LLS methods are:

$$\nabla^2 f_{\mathrm{WLLS}}(\boldsymbol{\gamma}) = \mathbf{W}^{\mathrm{T}}\mathbf{S}^2\mathbf{W}, \text{ and} \tag{18}$$

$$\nabla^2 f_{\mathrm{LLS}}(\boldsymbol{\gamma}) = \mathbf{W}^{\mathrm{T}}\mathbf{W}, \tag{19}$$

$$\nabla f_{\mathrm{WLLS}}(\boldsymbol{\gamma}) = -(\mathbf{SW})^{\mathrm{T}}\mathbf{S}(\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}), \tag{20}$$

$$\nabla f_{\mathrm{LLS}}(\boldsymbol{\gamma}) = -\mathbf{W}^{\mathrm{T}}(\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}). \tag{21}$$

Despite the additional information required to specify the CNLS objective function, its Jacobian, Hessian, and gradient vector are remarkably similar to its unconstrained counterparts; these higher order structures are listed below:

$$\nabla^2 f_{\mathrm{CNLS}}(\boldsymbol{\rho}) = \mathbf{J}_{\boldsymbol{\rho}}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{W}^{\mathrm{T}}(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma}) + \sum_{i=1}^{m} r_i \hat{s}_i \mathbf{P}_i \tag{22}$$

$$\nabla f_{\mathrm{CNLS}}(\boldsymbol{\rho}) = -\mathbf{J}_{\boldsymbol{\rho}}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{W}^{\mathrm{T}}\hat{\mathbf{S}}\mathbf{r} \tag{23}$$

where $[\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma})]_{ij} \equiv \frac{\partial \gamma_i}{\partial \rho_j}$,

$$\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\rho_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\rho_3 & 0 & 2\rho_5 & 0 & 0 \\ 0 & 0 & 0 & 2\rho_4 & 0 & 2\rho_6 & 2\rho_7 \\ 0 & \rho_5 & 0 & 0 & \rho_2 & 0 & 0 \\ 0 & 0 & \rho_6 & 0 & \rho_7 & \rho_3 & \rho_5 \\ 0 & \rho_7 & 0 & 0 & 0 & 0 & \rho_2 \end{pmatrix}, \tag{24}$$

and

$$\mathbf{P}_i = (-1)\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2W_{i2} & 0 & 0 & W_{i5} & 0 & W_{i7} \\ 0 & 0 & 2W_{i3} & 0 & 0 & W_{i6} & 0 \\ 0 & 0 & 0 & 2W_{i4} & 0 & 0 & 0 \\ 0 & W_{i5} & 0 & 0 & 2W_{i3} & 0 & W_{i6} \\ 0 & 0 & W_{i6} & 0 & 0 & 2W_{i4} & 0 \\ 0 & W_{i7} & 0 & 0 & W_{i6} & 0 & 2W_{i4} \end{pmatrix}. \tag{25}$$

The derivations of the above equations are provided in Appendix C.

If the NLS estimate is positive definite then this estimate is equivalent to the CNLS estimate. This result can be obtained by replacing the map $\boldsymbol{\gamma}(\boldsymbol{\rho})$ with the identity map so that the Jacobian matrix $\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma})$ in Eqs. (22)–(24) reduces to the identity matrix. Therefore, the gradient vector and the Hessian matrix of the CNLS method reduce to that of the NLS method. The reduction from the CNLS method to the CWLLS and CLLS methods can be analogously established.

2.2. The modified full Newton's method

In this section, we present the basic idea of modified full Newton's (MFN) method; the specific algorithm is in Appendix D in a format that allows for ready implementation of the NLS and CNLS methods. Before presenting the algorithm, we would like to give a brief introduction to the LM and the proposed methods in the context of modified full Newton's method of function minimization.

Define the least squares objective function,

$$f(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{m} r_i(\boldsymbol{\gamma})^2 = \frac{1}{2}\mathbf{r}^{\mathrm{T}}\mathbf{r}, \tag{26}$$

where $\mathbf{r}(\boldsymbol{\gamma}) = [r_1(\boldsymbol{\gamma}) \cdots r_m(\boldsymbol{\gamma})]^{\mathrm{T}}$. The equation to be solved in the kth iteration in MFN method can be written as [13,14]:

$$\mathbf{H}(\boldsymbol{\gamma}_k)\boldsymbol{\delta}_k = -\nabla f(\boldsymbol{\gamma}_k), \tag{27}$$

where $\boldsymbol{\delta}_k$ is known as the search step vector, $\mathbf{H}(\boldsymbol{\gamma}_k)$ is the generalized Hessian matrix and $\nabla f(\boldsymbol{\gamma}_k)$ is the gradient vector. The gradient vector is written as:

$$\nabla f(\boldsymbol{\gamma}) = \mathbf{J}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{r}(\boldsymbol{\gamma}), \tag{28}$$

where $\mathbf{J}(\boldsymbol{\gamma})$ is known as the Jacobian matrix with $[\mathbf{J}(\boldsymbol{\gamma})]_{ij} = \partial r_i(\boldsymbol{\gamma})/\partial \gamma_j$.

It is interesting to note that the key difference in various approaches of function minimization lies in the expression of the generalized Hessian matrix. For example, the generalized Hessian matrix for the MFN, Gauss Newton's, Newton's, and LM methods can be written as:

$$\mathbf{H}_{MFN}(\boldsymbol{\gamma}) \equiv \mathbf{J}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{J}(\boldsymbol{\gamma}) + \sum_{i=1}^{N} r_i(\boldsymbol{\gamma})\nabla^2 r_i(\boldsymbol{\gamma}) + \lambda\mathbf{I}, \tag{29}$$

$$\mathbf{H}_{GN}(\boldsymbol{\gamma}) \equiv \nabla^2 f(\boldsymbol{\gamma}) = \mathbf{J}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{J}(\boldsymbol{\gamma}), \tag{30}$$

$$\mathbf{H}_{N}(\boldsymbol{\gamma}) \equiv \nabla^2 f(\boldsymbol{\gamma}) = \mathbf{J}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{J}(\boldsymbol{\gamma}) + \sum_{i=1}^{N} r_i(\boldsymbol{\gamma})\nabla^2 r_i(\boldsymbol{\gamma}), \tag{31}$$

and

$$\mathbf{H}_{LM}(\boldsymbol{\gamma}) \equiv \mathbf{J}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{J}(\boldsymbol{\gamma}) + \lambda\mathbf{I}, \tag{32}$$

respectively; where $\mathbf{I}$ is the identity matrix and λ is the Levenberg–Marquardt parameter, which is always assumed to be a nonnegative real number.

In the MFN algorithm, we take Eq. (29) as our generalized Hessian matrix. In addition to that, the parameter λ will be set to zero initially and will remain so during the iterative process until a higher objective function value is encountered. This is done so that a full Newton step can be taken at the first iteration since the WLLS estimate can be used as a reasonable initial

guess. For completeness, the algorithm for the MFN method is shown in Appendix D.

2.3. Methods of comparison and numerical simulations

Monte Carlo simulations similar to those of Pierpaoli and Basser [18] were carried out to analyze the MFN and the LM methods by comparing the percent relative error in estimating the trace, where the percent of relative error of an estimate $\hat{\psi}$ of a known parameter is defined as $|\frac{\hat{\psi}-\psi}{\psi}| \times 100\%$. Further, we used the reduced χ^2, χ^2_ν, value as another measure to gauge the accuracy and the goodness-of-fit among these methods [19].

Since the theoretical variance for a given simulation is known a priori, comparing the normalized histograms of the χ^2_ν estimates to the theoretical distribution provides an excellent measure for goodness of fit. Briefly, let $\hat{\gamma}$ be the NLS (or CNLS) estimate of the objective function f_{NLS} (or f_{CNLS}), then $\frac{2f_{\mathrm{NLS}}(\hat{\gamma})}{\nu}$ $\left(\text{or } \frac{2f_{\mathrm{CNLS}}(\hat{\gamma})}{\nu}\right)$ is an unbiased variance estimate of the DW signals where $\nu = m - p = m - 7$ is the number of degrees of freedom; m is the number of sampled signals; and p the number of parameters. We shall denote $\frac{2f_{\mathrm{NLS}}(\hat{\gamma})}{\nu}$ or $\frac{2f_{\mathrm{CNLS}}(\hat{\gamma})}{\nu}$ as σ^2_{DW}. The χ^2_ν value can be computed by dividing the variance estimate with the known variance, that is, $\frac{\sigma^2_{\mathrm{DW}}}{\sigma^2_{\mathrm{Rician}}}$, where $\sigma^2_{\mathrm{Rician}}$ is the known variance of the noise based on the Rician probability density [20,21]. Intuitively, $\frac{\sigma^2_{\mathrm{DW}}}{\sigma^2_{\mathrm{Rician}}} \approx 1$ indicates a good estimate of σ^2_{DW}.

To facilitate the comparison between the normalized histogram and the theoretical density curve, we will need the reduced χ^2 probability density. We provide here an outline of this derivation. Let the χ^2_ν probability density be $g_{\chi^2_\nu}$, and the Chi-square χ^2 probability density be g_{χ^2}. Then, the χ^2 density can be written as [19]

$$g_{\chi^2}(x) = \frac{2^{-\nu/2}}{\Gamma(\frac{\nu}{2})} x^{(\nu/2)-1} \mathrm{e}^{-x/2}. \tag{33}$$

The χ^2_ν probability density can be obtained by making a linear transformation on the random variable, x, so that the new random variable, y, can be written as $y = x/\nu$:

$$g_{\chi^2_\nu}(x) = \nu g_{\chi^2}(\nu x). \tag{34}$$

The expected value and variance of a random variable with χ^2_ν density are:

$$E_{\chi^2_\nu}[x] = 1 \quad \text{and} \quad Var_{\chi^2_\nu}[x] = 2/\nu. \tag{35}$$

The plot of the χ^2_ν density with different numbers of degrees of freedom is shown in Fig. 1.

The magnitude MR image is derived from the complex signals and is used for diffusion tensor estimation; therefore, noise characteristics of the magnitude MR signal will affect the accuracy of the tensor estimate. It is well known that noise in MR magnitude signals follows the Rician distribution [20–22]. Therefore, the theoretical variance used to generate Gaussian noise $\sigma^2_{\mathrm{Gaussian}}$ for each of the real and complex components will have to be transformed appropriately with respect to Rician density when the noise variance in the magnitude image is of interest. Provided here is an exact formula taken from Koay and Basser [22] for expressing the variance in magnitude MR signal in terms of the variance of the Gaussian noise in the two quadrature channels and a correction factor, ξ. This correction factor is written in terms of SNR in order to facilitate simulation studies. Let $\theta = \mathrm{SNR}$, the noise variance in magnitude MR signal can be expressed as [22]:

$$\sigma^2_{\mathrm{Rician}} = \xi(\theta)\sigma^2_{\mathrm{Gaussian}}, \tag{36}$$

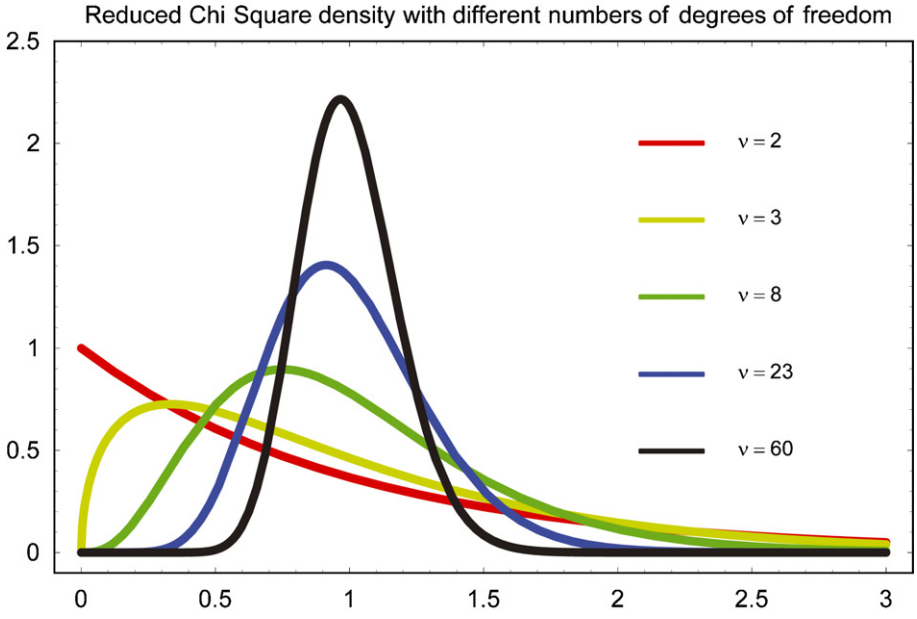

Fig. 1. Reduced χ^2 probability density curves with different numbers of degrees of freedom.

where

$$\xi(\theta) = \left(2 + \theta^2 - \frac{\pi}{8} e^{-\theta^2/2}\left[(2+\theta^2) I_0(\theta^2/4) + \theta^2 I_1(\theta^2/4)\right]^2\right) \tag{37}$$

and I_0 and I_1 are the modified Bessel functions of order zero and one, respectively.

Two different simulations are carried out in this work. The first simulation focuses on the distributional properties of the χ_v^2 estimate and of the trace estimate as obtained by various nonlinear LS algorithms. In this type of computationally expensive simulation, we have to be selective in the choice of physiologically relevant tensors in order to reduce computation cost and, more importantly, to make the simulation results concise and representative. Therefore, we have chosen two specific tensors for this simulation: two cylindrically symmetric tensors with the same trace value of 2.190×10^{-3} mm^2/s, but different FA values of 0.5398 and 0.8643. Other relevant simulation parameters are listed here, diffusion weight ($b = 1000$ s/mm^2), the reference signal ($S_0 = 1000$ a.u.) and the parameter vectors $\boldsymbol{\gamma}$, ([ln(1000) s/mm^2, 1.236×10^{-3}, 4.765×10^{-3}, 4.765×10^{-3}, 0, 0, 0]$^{\mathrm{T}}$ mm^2/s and [ln(1000) s/mm^2, 1.758×10^{-3}, 2.158×10^{-3}, 2.158×10^{-3}, 0, 0, 0]$^{\mathrm{T}}$ mm^2/s).

The second simulation is based on simulated human brain tensor data. Its goal is to complement the first simulation by accounting for a wide range of tensor shapes. In this simulation, we focus on the human brain map of the mean value of the relative error in estimated trace. The clinical DT-MRI human brain images were acquired from a healthy volunteer using a high angular scheme [27,28]. All images were co-registered [23] and robust tensor estimation [7] was used to eliminate "outliers" from the data. The computed tensors, combined with the relevant parameters mentioned above, were then used to create the simulated diffusion weighted signals and one non-diffusion weighted signal using the single diffusion tensor model of Basser [4]. Gaussian noise was added in quadrature [18] so as to simulate images with a signal-to-noise ratio (in the non-diffusion weighted image) of 5 in each pixel. This particular approach allows us to investigate the response of anatomically specific tensors in the brain under the same simulation conditions, which would otherwise be quite difficult experimentally. In this way, we are able to identify regions in the brain where the constrained methods are likely to be useful, i.e., in regions where negative eigenvalues are more prevalent.

We shall adopt the following convention on the algorithms mentioned above when discussing the results: NLS-LM (NLS estimation using the LM method), NLS-MFN (NLS estimation using the MFN method), CNLS-LM (CNLS estimation using the LM method) and, CNLS-MFN (CNLS estimation using the MFN method). Finally, the LM method used in this study was taken from a routine in JMSL of Visual Numerics® called NonlinLeastSquare which is based on MINPACK routine LMDIF by Moré et al. [24]. The MFN routine for the NLS and CNLS methods was developed in-house using the Java programming language together with the QR decomposition routine from JAMA [25].

3. Results and discussion

The results on the distributional properties of the χ_v^2 estimate and of the trace estimate are summarized in Figs. 2 and 3. The results on the average value of the relative error in estimating the trace in the simulated human brain map are shown in Fig. 4. Fig. 5 shows the difference in these average values among various methods considered in this paper. The results of Figs. 2 and 3 are computed from a collection of 50,000 simulated tensors. In Figs. 4 and 5, the results on each pixel are computed from a collection of 10,000 simulated tensors. The histograms of the χ_v^2 estimate and of the trace estimate are plotted in Figs. 2 and 3, respectively. Each histogram in the panel is computed using different methods, i.e. the NLS-LM, the NLS-MFN, the CNLS-LM or the CNLS-MFN method.

In Fig. 2, the results of the χ_v^2 estimate associated with the first tensor with medium FA of 0.539 at SNR = 5 and SNR = 15 are shown in panels A and B, respectively. Similarly, the results associated with the second tensor with FA = 0.864 at SNR = 5 and SNR = 15 are shown in panels 2C and D, respectively. In each panel, the theoretical distribution is shown in gray. It is interesting to note that the χ_v^2 histogram of the NLS-MFN method is shifted to the left of the theoretical distribution in Figs. 2A and C, which implies that the χ_v^2 estimated by the NLS-MFN method is, in general, lower than the known distribution! Low χ_v^2 values do not necessarily indicate a better fit, but rather a problematic estimate of the variance, i.e., $\sigma^2_{\mathrm{Rician}}$. This anomaly of having a lower χ_v^2 value than expected might not have been noticed without the Newton-type method of optimization, i.e., the MFN method. More importantly, this anomaly suggests that the signal variance is orientation dependent, that is, the variance depends on the gradient direction. Therefore, a new experimental design capable of obtaining multiple replicates in each gradient direction is needed. This new experimental design would allow estimation of the mean signal and signal variance on each gradient direction, the analytically exact correction scheme proposed by Koay and Basser [22] can be used to estimate diffusion weighted signals that are Gaussian distributed. This approach reduces considerably the effects of the noise floor. This research topic is currently under investigation. Note that the pathologies of the rectified noise floor on tensor-derived quantities have been investigated by Jones and Basser [9].

The results in Fig. 3 are arranged similarly to those in Fig. 2. It is interesting to note here that a systematic shift in the distributions of the trace estimate as computed by the LM method, i.e., NLS-LM and CNLS-LM, can be seen

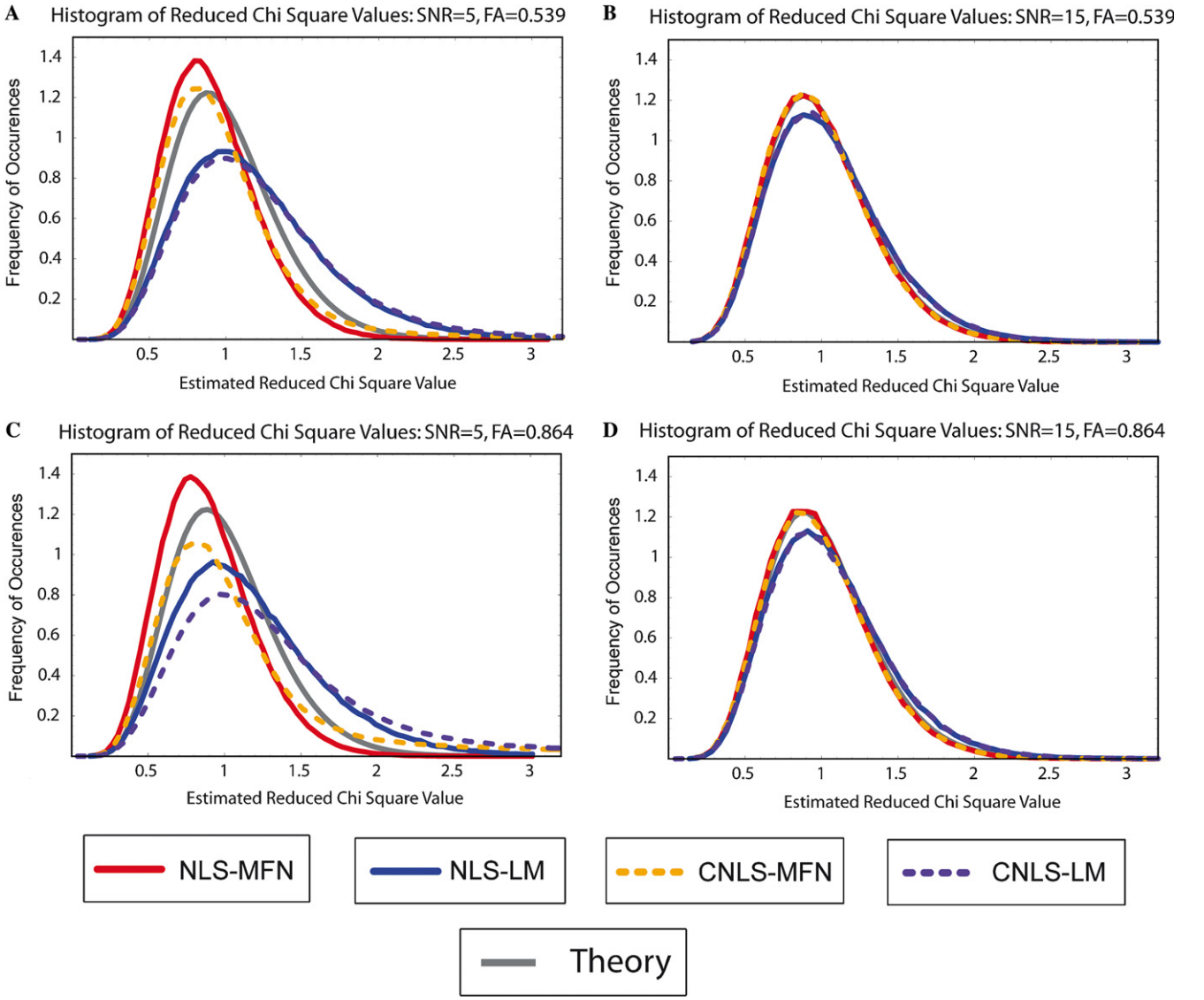

Fig. 2. Histogram of reduced χ^2 values for two different SNR levels and FA values calculated from 500,000 simulated tensors: (A) SNR = 5, FA = 0.539, (B) SNR = 15, FA = 0.539, (C) SNR = 5, FA = 0.864, and (D) SNR = 15, FA = 0.864. Note that the theoretical reduced χ^2 curve in (B) and in (D) is superimposed on that of MFN.

quite easily at SNR = 5. The quantitative information on these shifts is tabulated in Table 1 as the percent of relative error in estimating the trace. The results in Table 1 can be summarized as follows: (I) in NLS estimation, the MFN method has a lower relative error in estimating the trace than the LM method, (II) in CNLS estimation, the MFN method is also better than the LM method, and (III) the CNLS-MFN method has lower relative error in estimating the trace than other methods considered in this paper.

The results on the simulated human brain data are shown in Figs. 4 and 5. Fig. 4 is the whole brain map of the average value of the relative error in estimating trace. The results show that the CNLS-MFN method has the lowest relative error among the methods considered here. The images shown in Figs. 4 and 5 indicate that the MFN method has lower relative error in estimating trace than the LM method in almost every region of the brain except in the ventricles and in the sulci where the results between the methods are comparable. Further, the difference between the NLS and the CNLS estimations by the same method of optimization, the LM method or the MFN method, can also be discerned, particularly, in the genu of the internal capsule and in the Corpus callosum, Figs. 5B and D. An obvious feature of Figs. 5B and D is that the figures closely resemble the FA map! This shows that the constrained methods are most relevant in the white matter regions.

Analysis of the algorithms presented here is an interesting area of study and is under investigation. A detailed discussion of this topic is beyond the scope of this paper. It suffices to say that the computation time per estimation for the methods discussed in this paper was approximately 1 ± 0.5 ms on a Dell Precision 670 with dual Intel Xeon 3.5-GHz processors.

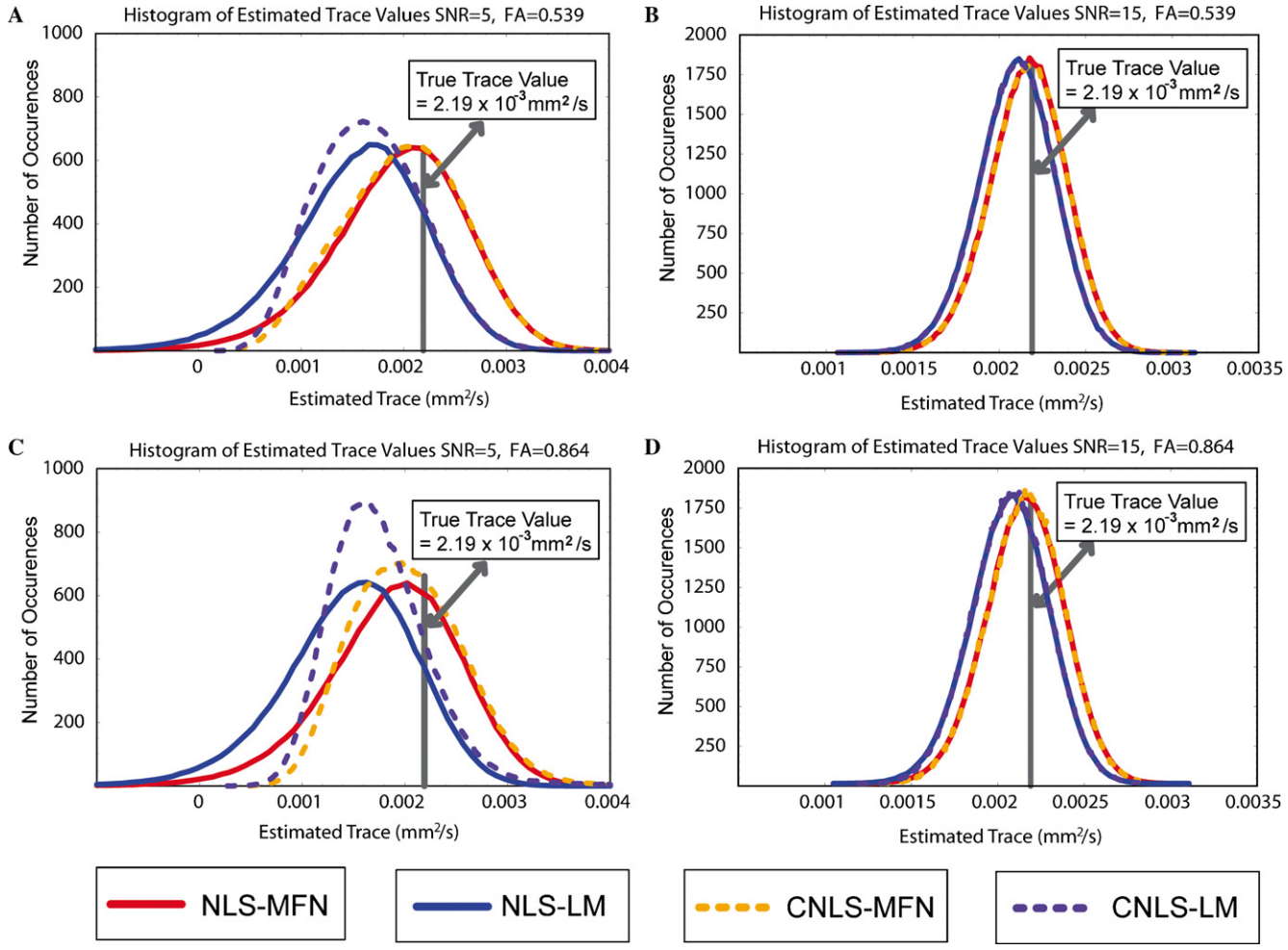

Fig. 3. Histogram of estimated trace values for two different SNR levels and FA values: (A) SNR = 5, FA = 0.539, (B) SNR = 15, FA = 0.539, (C) SNR = 5, FA = 0.864, and (D) SNR = 15, FA = 0.864.

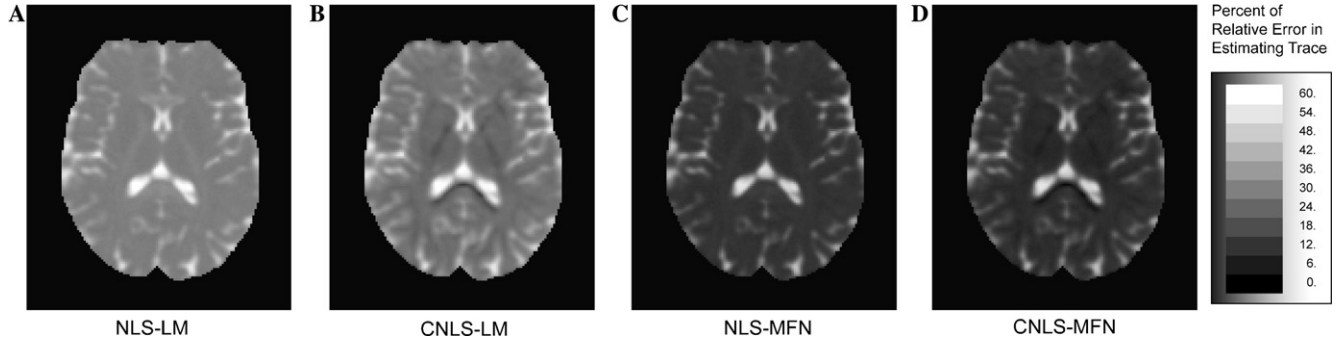

Fig. 4. The average value of the percent relative error in estimating trace by the (A) NLS-LM, (B) CNLS-LM, (C) NLS-MFN, and (D) CNLS-MFN methods based on simulated human brain data with SNR = 5, $b = 1000$ s/mm^2 and a 23 gradient direction set. These images show that the MFN method has lower relative error in estimating trace than does the LM method in almost every region of the brain except in the ventricles and sulci. Interestingly, the difference between the NLS and the CNLS estimations by the same method of optimization, the LM method or the MFN method, can readily be discerned in the genu of the internal capsule and in the Corpus callosum (B and D); these regions are known to have high FA values.

4. Conclusion

The Hessian matrices for various least squares problems are explicitly derived. Simulation results indicate that the accuracy of a diffusion tensor estimate can be substantially improved by explicitly including the Hessian matrix in the least squares estimation algorithm. The proposed constrained nonlinear least squares estimation based on the modified full Newton's method has lower relative error in estimating the trace than other methods discussed in this

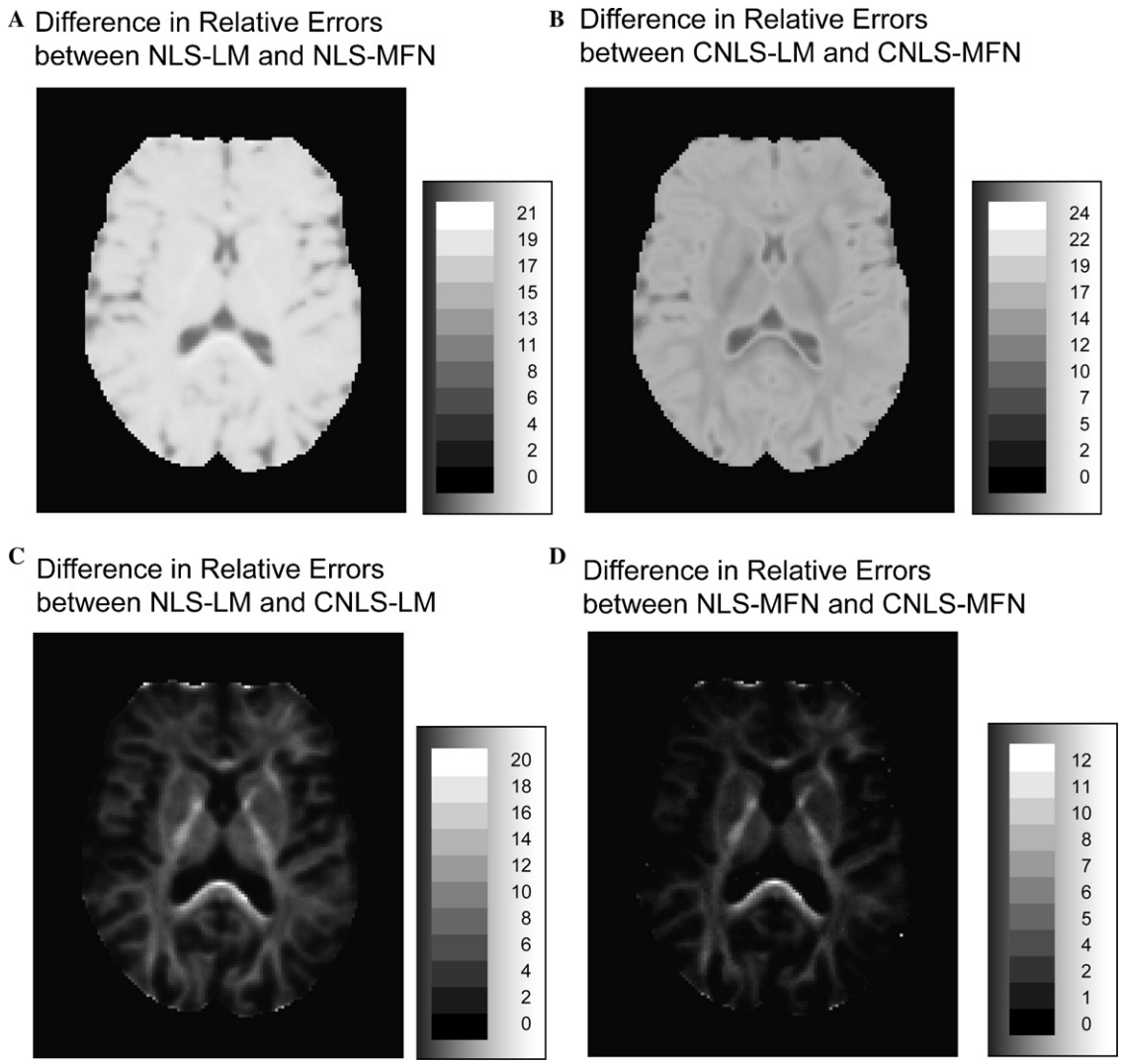

Fig. 5. The difference in the average percent of relative error in estimating trace between (A) NLS-LM and NLS-MFN, (B) CNLS-LM and CNLS-MFN, (C) NLS-LM and CNLS-LM, and (D) NLS-MFN and CNLS-MFN. These images again show that the MFN method has lower relative error in estimating trace than does the LM method. The differences between the NLS and the CNLS estimations by the same method of optimization is more readily discernible in (C and D) and, as commented on Fig. 4, these differences are most distinct in the genu of the internal capsule and in the Corpus callosum. It is interesting to note the similarity in features between these images (C and D) and a typical FA map.

Table 1
Percent of relative error in estimating the trace

	SNR 5 (%)		SNR 15 (%)	
	Medium FA	High FA	Medium FA	High FA
NLS-MFN	10.76	14.10	1.10	1.49
NLS-LM	29.22	33.39	4.21	5.44
CNLS-MFN	8.70	7.24	1.08	1.31
CNLS-LM	23.82	20.10	4.19	5.21

paper. The proposed method not only provides a more accurate tensor estimate but also a more accurate Hessian matrix. The importance of the Hessian matrix can be gleaned from recent works by Chang et al. [29], Carew et al. [30] and Koay et al. [31], where the inverse of the Hessian matrix is used for computing the variance–covariance matrix of the estimated DTI parameters. Therefore, the proposed framework will be very useful in testing optimal experimental designs in DTI as well as in fiber tractography where the variability in the major eigenvector can be accurately quantified.

Acknowledgments

C.G.K. thanks Dr. M. Elizabeth Meyerand and Dr. Andrew L. Alexander for the initial encouragement on this work. The authors thank Dr. Andrew L. Alexander for critically reading the early draft of this paper and Dr. Stefano Marenco for acquiring the human brain data set. We gratefully acknowledge Liz Salak for editing this paper. This research was supported by the Intramural Research Program of the National Institute of Child Health and Human Development (NICHD), National Institutes of Health, Bethesda, Maryland.

Appendix A

The derivation of Eq. (11) from Eq. (10) is shown below:

$$\begin{aligned} f_{\mathrm{NLS}}(\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{i=1}^{m}\left(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{m} s_i^2\left(1 - \frac{1}{s_i}\exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{m} s_i^2\left(1 - \exp\left[-\left(\ln(s_i) - \sum_{j=1}^{7} W_{ij}\gamma_j\right)\right]\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{m} s_i^2(1 - \exp[-F_i])^2 \end{aligned}$$

and

$$\begin{aligned} f_{\mathrm{NLS}}(\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{i=1}^{m}\left(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{m} \hat{s}_i^2\left(s_i \exp\left[-\sum_{j=1}^{7} W_{ij}\gamma_j\right] - 1\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{m} \hat{s}_i^2(\exp[+F_i] - 1)^2 \end{aligned}$$

Appendix B

In this appendix, we will derive the gradient vector, the Jacobian matrix and the Hessian matrix of the NLS objective function. Given the NLS objective function

$$f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{m}\left(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]\right)^2,$$

the derivative of $f_{\mathrm{NLS}}(\boldsymbol{\gamma})$ with respect to γ_l is

$$\begin{aligned} \frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_l} &= \sum_{i=1}^{m}\left[r_i(-\hat{s}_i)\left(\sum_{j=1}^{7} W_{ij}\frac{\partial \gamma_j}{\partial \gamma_l}\right)\right] = -\sum_{i=1}^{m} r_i\hat{s}_i W_{il} \\ &= -\sum_{i=1}^{m} W_{li}^{\mathrm{T}}\hat{s}_i r_i. \end{aligned}$$

In matrix notation, the gradient vector has the following form:

$$\nabla f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \begin{bmatrix} \frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_1} \\ \vdots \\ \frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_7} \end{bmatrix} = -\mathbf{W}^{\mathrm{T}}\hat{\mathbf{S}}\mathbf{r} = \mathbf{J}^{\mathrm{T}}\mathbf{r},$$

where the transpose of the Jacobian matrix is $\mathbf{J}^{\mathrm{T}} = -\mathbf{W}^{\mathrm{T}}\hat{\mathbf{S}}$.

The second order derivative of the NLS objective function will be established as follows:

$$\begin{aligned} \frac{\partial^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_k \partial \gamma_l} &= \sum_{i=1}^{m}\frac{\partial}{\partial \gamma_k}[r_i(-\hat{s}_i)W_{il}] \\ &= \sum_{i=1}^{m}\left[\hat{s}_i^2 W_{il}W_{ik} + r_i(-\hat{s}_i)W_{il}W_{ik}\right] \\ &= \sum_{i=1}^{m} W_{ki}^{\mathrm{T}}\left[\hat{s}_i^2 - r_i\hat{s}_i\right]W_{il}. \end{aligned}$$

In matrix notation, the full Hessian matrix is

$$\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \mathbf{W}^{\mathrm{T}}(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}.$$

Appendix C

In this appendix, we derive the gradient vector and the Hessian matrix of the constrained nonlinear least squares method.

$$f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \frac{1}{2}\sum_{i=1}^{m}(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\rho})\right])^2$$

$\frac{\partial f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))}{\partial \rho_l} = \sum_{i=1}^{7}\frac{\partial \gamma_i}{\partial \rho_l}\frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_i}$ by change of variables. In matrix notation, the gradient vector is

$$\nabla f_{\mathrm{CNLS}}(\boldsymbol{\rho}) = \begin{bmatrix} \frac{\partial f_{\mathrm{CNLS}}(\boldsymbol{\rho})}{\partial \rho_1} \\ \vdots \\ \frac{\partial f_{\mathrm{CNLS}}(\boldsymbol{\rho})}{\partial \rho_7} \end{bmatrix} = -\mathbf{J}_{\boldsymbol{\rho}}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{W}^{\mathrm{T}}\hat{\mathbf{S}}\mathbf{r}$$

$$\begin{aligned} \frac{\partial^2 f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))}{\partial \rho_k \partial \rho_l} &= \sum_{i=1}^{7}\frac{\partial^2 \gamma_i}{\partial \rho_k \partial \rho_l}\frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_i} + \frac{\partial \gamma_i}{\partial \rho_l}\frac{\partial^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_i \partial \rho_k} \\ &= \sum_{i=1}^{7}\frac{\partial^2 \gamma_i}{\partial \rho_k \partial \rho_l}\frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_i} + \frac{\partial \gamma_i}{\partial \rho_l}\frac{\partial}{\partial \gamma_i}\sum_{j=1}^{7}\frac{\partial \gamma_j}{\partial \rho_k} \\ &\quad \times \frac{\partial f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_j} \\ &= \sum_{i=1}^{7}\sum_{j=1}^{7}\frac{\partial \gamma_i}{\partial \rho_l}\frac{\partial^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\frac{\partial \gamma_j}{\partial \rho_k} \\ &\quad + \sum_{q=1}^{m} r_q\hat{s}_q \underbrace{\sum_{i=1}^{7}(-W_{qi})\frac{\partial^2 \gamma_i}{\partial \rho_k \partial \rho_l}}_{[\mathbf{P}_q]_{kl}}. \end{aligned}$$

In matrix notation, we have

$$\nabla^2 f_{\mathrm{CNLS}}(\boldsymbol{\gamma}) = \mathbf{J}_{\boldsymbol{\rho}}^{\mathrm{T}}(\boldsymbol{\gamma})\mathbf{W}^{\mathrm{T}}(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma}) + \sum_{i=1}^{m} r_i\hat{s}_i\mathbf{P}_i.$$

Appendix D

In this appendix, we provide the MFN algorithm for both the NLS and CNLS estimations. In the CNLS estimation, the initial guess has to be modified slightly before being used in the MFN algorithm.

MFN algorithm:

At the initial iteration, let γ_0 be the solution to the WLLS problem, $\lambda = 0$, and *flag* = *true*. When the Hessian and the gradient vector have to be evaluated at new γ the *flag* will be set *true*.

Then at the kth iteration,

1. if(*flag* == *true*) Evaluate $\mathbf{H}_{MFN}(\gamma_k)$ and $\nabla f(\gamma_k)$
2. Solve $(\mathbf{H}_{MFN}(\gamma_k) + \lambda \mathbf{I})\ \boldsymbol{\delta}_k = -\nabla f(\gamma_k)$ for $\boldsymbol{\delta}_k$
3. If $(f(\gamma_k + \delta_k) < f(\gamma_k))$ {
   $\lambda = 0.1 \times \lambda$
   Accept $\boldsymbol{\delta}_k$ by setting $\gamma_{k+1} = \gamma_k + \boldsymbol{\delta}_k$
   flag = *true*
   }Else {
   if ($\lambda == 0$), set $\lambda = 0.0001$
   else $\lambda = 10.0 \times \lambda$
   Reject $\boldsymbol{\delta}_k$ by setting $\gamma_{k+1} = \gamma_k$
   flag = *false*
   }
4. Repeat these steps, (1, 2 and 3), until $0 \leqslant -\boldsymbol{\delta}_k^{\mathrm{T}} \nabla f(\gamma_{k+1}) < \varepsilon_1$ and $|f(\gamma_{k+1}) - f(\gamma_k)| < \varepsilon_2$ where ε_1 and ε_2 are small positive numbers.

As mentioned in the text, a slight modification is needed for the CNLS method because the initial guess is taken from the WLLS method rather than the CWLLS method. Therefore, the parameter vector, $\boldsymbol{\rho}$, for the CNLS method has to be obtained from the *modified Cholesky* factor [26] derived from the diffusion tensor estimate of γ. The modified Cholesky factorization is one of the approaches to make a non positive definite symmetric matrix sufficiently positive definite [13,26].

References

[1] P.J. Basser, J. Mattiello, D. Lebihan, MR diffusion tensor and imaging, Biophys. J. 66 (1994) 259–267.

[2] P.J. Basser, Inferring microstructural features and the physiological state of tissues from diffusion-weighted images, Nmr Biomed. 8 (1995) 333–344.

[3] P.J. Basser, C. Pierpaoli, Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI, J. Magn. Reson. B 111 (1996) 209–219.

[4] P.J. Basser, D. LeBihan, J. Mattiello, Estimation of the effective self-diffusion tensor from the NMR spin echo, J. Magn. Reson. B 40 (1994) 247–254.

[5] D.M. Bates, D.G. Watts, Nonlinear Regression Analysis and Its Applications, Wiley, New York, 1988.

[6] A.W. Anderson, Theoretical analysis of the effects of noise on diffusion tensor imaging, Magn. Reson. Med. 46 (2001) 1174–1188.

[7] L.C. Chang, D.K. Jones, C. Pierpaoli, RESTORE: robust estimation of tensors by outlier rejection, Magn. Reson. Med. 53 (2005) 1088–1095.

[8] N.G. Papadakis, K.M. Martin, I.D. Wilkinson, C.L. Huang, A measure of curve fitting error for noise filtering diffusion tensor MRI data, J. Magn. Reson. 164 (2003) 1–9.

[9] D.K. Jones, P.J. Basser, "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data, Magn. Reson. Med. 52 (2004) 979–993.

[10] R. Salvador, A. Pe~na, D.K. Menon, T.A. Carpenter, J.D. Pickard, E.T. Bullmore, Formal characterization and extension of the linearized diffusion tensor model, Human Brain Mapp. 24 (2005) 144–155.

[11] Z. Wang, B.C. Vemuri, Y. Chen, T.H. Mareci, A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI, IEEE Trans. Med. Imaging 23 (8) (2004) 930–939.

[12] C.G. Koay, J.D. Carew, A.L. Alexander, P.J. Basser, M.E. Meyerand, Investigation of anomalous estimates in tensor-derived quantities in DTI, Magn. Reson. Med. 55 (2006) 930–936.

[13] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.

[14] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, Cambridge University Press, New York, 1992, 408p.

[15] C.G. Koay, Advances in data analysis of diffusion tensor imaging. PhD Dissertation at the University of Wisconsin-Madison, 2005. UMI Publication Number AAT 3186244. Link: http://www lib.umi.com/dissertations/fullcit/3186244.

[16] E. Stejskal, J. Tanner, Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient, J. Chem. Phys. 42 (1965) 288–292.

[17] J.C. Pinheiro, D.G. Bates, Unconstrained parametrizations for variance–covariance matrices, Stat. Comput. 6 (1996) 289–296.

[18] C. Pierpaoli, P.J. Basser, Toward a quantitative assessment of diffusion anisotropy, Magn. Reson. Med. 36 (1996) 893–906.

[19] P.R. Bevington, D.K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, second ed., McGraw-Hill, New York, 1992, 195p.

[20] R.M. Henkelman, Measurement of signal intensities in the presence of noise in MR images, Med. Phys. 12 (2) (1985) 232–233.

[21] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965, pp. 195–196.

[22] C.G. Koay, P.J. Basser, Analytically exact correction scheme for signal extraction from noisy magnitude MR signals, J. Magn. Reson. 179 (2006) 317–322.

[23] G.K. Rohde, A.S. Barnett, P.J. Basser, S. Marenco, C. Pierpaoli, Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI, Magn. Reson. Med. 51 (2004) 103–114.

[24] Moré, Jorge, Burton Garbow, Kenneth Hillstrom, *User guide for MINPACK-1*, Argonne National Laboratory Report ANL-80-74, Argonne, Illinois, 1980.

[25] http://math.nist.gov/javanumerics/jama/.

[26] P. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, New York, 1981.

[27] D.K. Jones, M.A. Horsfield, A. Simmons, Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging, Magn. Reson. Med. 42 (3) (1999) 515–525.

[28] C. Pierpaoli, D.K. Jones, Removing CSF Contamination in brain DT-MRIs by using a two-compartment tensor model, ISMRM 2004, Japan 2004.

[29] L.C. Chang, C. Pierpaoli, P.J. Basser, The variance of DTI-derived parameters via first-order perturbation methods, Proc. Intl. Soc. Mag. Reson. Med. 14 (2006).

[30] J.D. Carew, C.G. Koay, G. Wahba, A.L. Alexander, P.J. Basser, M.E. Meyerand, The asymptotic distribution of diffusion tensor and fractional anisotropy estimates, Proc. Intl. Soc. Mag. Reson. Med. 14 (2006).

[31] C.G. Koay, L.C. Chang, C. Pierpaoli, P.J. Basser, Error propagation framework for diffusion tensor imaging, Proc. Intl. Soc. Mag. Reson. Med. 14 (2006).