Fiber Orientation Distribution Estimation Using a Peaceman–Rachford Splitting Method*

Yannan Chen†, Yu-Hong Dai‡, and Deren Han§

Abstract. In diffusion-weighted magnetic resonance imaging, the estimation of the orientations of multiple nerve fibers in each voxel (the fiber orientation distribution (FOD)) is a critical issue for exploring the connection of cerebral tissue. In this paper, we establish a convex semidefinite programming (CSDP) model for the FOD estimation. One feature of the new model is that it can ensure the statistical meaning of FOD since as a probability density function, FOD must be nonnegative and have a unit mass. To construct such a statistically meaningful FOD, we consider its approximation by a sum of squares (SOS) polynomial and impose the unit-mass by a linear constraint. Another feature of the new model is that it introduces a new regularization based on the sparsity of nerve fibers. Due to the sparsity of the orientations of nerve fibers in cerebral white matter, a heuristic regularization is raised, which is inspired by the Z-eigenvalue of a symmetric tensor that closely relates to the SOS polynomial. To solve the CSDP efficiently, we propose a new Peaceman–Rachford splitting method and prove its global convergence. Numerical experiments on synthetic and real-world human brain data show that, when compared with some existing approaches for fiber estimations, the new method gives a sharp and smooth FOD. Further, the proposed Peaceman–Rachford splitting method is shown to have good numerical performances comparing several existing methods.

1. Introduction. Diffusion-weighted magnetic resonance imaging (DW-MRI) is a tool that explores the microarchitecture of the white matter of biological tissues in a noninvasive and in vivo manner. Since the DW-MRI signal attenuation is sensitive to the directional diffusion process of water molecules, which are restricted by the complex local biological tissue environment, we can infer the detailed orientation information of nerve fibers by analyzing DW-MRI

signals. In clinic [14], the orientation information of nerve fibers provided by DW-MRI helps to investigate some neurological and psychiatric diseases such as stroke, epilepsy, neurodegenerative diseases, and spinal cord disorders. In neuroscience [35], by means of DW-MRI, an ambitious project named the human connectome project,[1] is raised and aims to reveal the anatomical connectivity of human brain in vivo which was never before possible.

Assuming the diffusion process of water molecules is Gaussian, Basser, Mattiello, and LeBihan [5] proposed the diffusion tensor model, where the diffusion tensor is a 3-by-3 symmetric positive definite matrix. Lin et al. [39] showed that the principle eigenvector of a diffusion tensor corresponds well with the orientation of a single fiber in cerebral white matter. However, due to the complexity of the human brain, the Gaussian assumption fails in regions that contain multiple nerve fibers, such as crossing and kissing [3, 57]. In fact, one third of the cerebral white matter regions contains more than one nerve fiber [6].

To characterize a non-Gaussian diffusion of water in complex biological tissue environment, Jensen et al. [32] presented the diffusion kurtosis imaging, which generalizes the diffusion tensor model by introducing an excess kurtosis term that approximates the higher order part of the diffusion displacement probability distribution. Qi et al. [48, 27, 47] gave some properties on the eigenvalue and the extremum of a fourth order kurtosis tensor. Hu et al. [31] studied the positive definiteness of diffusion kurtosis imaging. Diffusion spectrum imaging is another effective approach which was proposed by Wedeen et al. [58]. They used the Fourier transform of the DW-MRI signals to recover the spin displacement spectra, whose orientational maxima is corresponding to a fiber orientation. However, diffusion spectrum imaging requires hundreds of DW-MRI measurements with multiple diffusion weighted b-values, which are expensive in a clinical application. Tuch et al. [57] raised a cheap DW-MRI sampling method named the high angular resolution diffusion imaging (HARDI), which collects DW-MRI signals in gradient orientations located on a semisphere with a fixed b-value. Signals with no diffusion weighting are also obtained for a normalization.

Using HARDI, researchers proposed numerous methods in the last decade. Tuch [56] introduced the Q-ball imaging which uses a Funk-Radon transform to compute the linear radial projection of a diffusion probability density function onto a unit sphere and obtain the diffusion orientation distribution function. With the help of spherical harmonics, an analytical Q-ball imaging was presented [19]. When the volume element in the projection of a diffusion probability density function was considered, some variations of the Q-ball imaging were proposed [54, 1, 8, 42]. Recently, Wolfers, Schwab, and Vidal [60] studied an adaptive selection strategy for active constraints of convex optimization and produced a practically nonnegative diffusion orientation distribution function. Another kind of method is spherical deconvolution (SD). Taking the interesting assumption that all the nerve fibers share a common response function, Tournier et al. [53] expressed the DW-MRI signal attenuation as a convolution of the FOD and the common response function. By selecting a proper response function and performing SD numerically, they obtained the FOD. We note that SD helped Q-ball imaging to obtain a sharp diffusion orientation distribution function [20].

1.1. Estimation of fiber orientation distributions. As a probability density function of fiber orientations within one voxel, FOD must be nonnegative in all directions and integrate

---

[1]See http://www.neuroscienceblueprint.nih.gov/connectome/.

to one on a unit sphere. We remark that the nonnegative constraint is also considered in diffusion tensor imaging [40] and Q-ball imaging [26, 51, 60]. To construct a nonnegative FOD, SD methods could be classified into two sorts.

The first one is based on the spatial discretization of all the directions on a unit sphere. We discretize the unit sphere in a finite set of orientations, and rotate the fiber response function to them. In this way, we obtain approximately complete basis functions. Then, the remaining work is to determine the nonnegative weights corresponding to these basis functions. Here, different basis functions produce different approaches. Ramirez-Manzanares et al. [50] and Dell'Acqua et al. [16] rotated the single-fiber diffusion tensor to generate basis functions. Whereafter, Patel et al. [44] refined this method. Jian et al. [34] and Kaden, Knösche, and Anwander [36] used the Wishart distribution and the Bingham distribution, respectively. Weldeselassie, Barmpoutis, and Atkins [59] and Barmpoutis, Ho, and Vemuri [4] employed the even order power of predefined linear functions. Numerically, to estimate the nonnegative weights of basis functions, Jian and Vemuri [33] and Weldeselassie, Barmpoutis, and Atkins [59] employed the classical nonnegative least squares method. Dell'Acqua et al. [16, 17] used a Richardson–Lucy iterative algorithm, which comes from Bayesian statistics and converges to the nonnegative least squares solution in the context of Gaussian noise. Patel et al. [44] took a projected gradient descent algorithm since they introduced a nonlinear regularization.

The other sort of SD methods is based on the discretization in the functional space. Tournier et al. [52] represented FOD in a spherical harmonic basis and penalized the negative regions iteratively. However, the resulting FOD could not eliminate negative probability completely. Alexander [2] used the concept of entropy from information theory and expressed FOD as an exponential function of a linear form. Cheng et al. [13] approximated FOD by the square of a single homogeneous polynomial in an even order spherical harmonic basis. So the least squares fitting of magnetic resonance (MR) signals is a quartic polynomial that is obviously nonconvex from an optimization viewpoint. Moreover, they considered the unit-mass constraint and adopted a Riemannian gradient descent algorithm.

1.2. Peaceman–Rachford splitting methods. Over the past decade, Lagrangian-based methods found many applications in compressive sensing, signal processing, statistical learning, transportation research, and imaging science. In our last work [12], we employed the alternating direction method of multiplier (ADMM) to estimate the generalized diffusion tensors which have a solid physical significance. In fact, ADMM was originally proposed by Glowinski and Marrocco [25] in 1975 and it was summarized recently in reference [24]. Gabay [23] pointed out that ADMM is indeed the Douglas–Rachford splitting method [22] applied to the dual program of the convex separable minimization.

Here, we focus on another splitting method, the Peaceman–Rachford splitting method (PRSM) [45], which is similar to ADMM except that PRSM updates the multiplier twice in each iteration. Gabay [23] pointed out that PRSM is more efficient than ADMM if it converges. To enforce the convergence of PRSM, He et al. [28] introduced a common underdetermined relaxation factor in the two steps of updating the multiplier. This smart modification makes PRSM strictly contractive. Then, the global convergence and the local sublinear convergence rate of the strictly contractive PRSM were established.

1.3. Contributions. The contributions of this paper are as follows.

1. *We establish a novel model for the problem.* Our model is a convex semidefinite programming (CSDP); see (2.30); which consists of all key characters of the problem. First, based on the discretization in a functional space, we estimate a fiber orientation distribution with a clearly statistical meaning, i.e., we impose the nonnegativity of the function in the constraint. In other words, although the coefficients are completely undetermined, a nonnegative SOS homogeneous polynomial is employed to approximate the FOD. Hence, a basis of the SOS polynomial does not correspond to any preset fiber orientation, and the resulting homogeneous polynomial is more flexible to capture the contour profile of nerve fibers. Moreover, we consider the unit-mass constraint which plays an important role in our numerical experiments. Inspired by the Z-eigenvalue theory of a higher order symmetric tensor which is uniquely determined by the SOS homogeneous polynomial, we introduce a generalized eigenproblem. Under some conditions, we prove that the eigenvalue of this generalized eigenproblem is the Z-eigenvalue of a symmetric tensor. Then, we give a new heuristic regularization skill.

2. *We propose a new efficient splitting method to solve the CSDP.* Our model falls into the framework of convex optimization with the sum of separable convex functions and linear equality constraints. A state-of-the-art algorithm for solving such problems is the alternating direction method of multipliers, which is a dual application of the classic Douglas–Rachford splitting method. Numerically, another splitting method, PRSM, usually performs better. We thus customize it to our model and propose a new Peaceman–Rachford splitting algorithm. To accelerate the convergence of iterates, we enlarge the relaxation factor in the second update-step of multiplier and introduce an additional correction step. The global convergence of the new algorithm is analyzed in the framework of a general convex separable minimization. Preliminary numerical experiments confirm the efficiency of this new algorithm.

3. We perform numerical experiments to examine the effectiveness of the model and the efficiency of the new algorithm. We code our algorithm in MATLAB and do experiments on both synthetic data and real-world human brain data. For synthetic data, we approve the validity of the new spherical deconvolution method, which generates sharp and smooth FODs. For the real-world human brain data, using the new method, we reconstruct the region of corpus callosum crossing corticospinal tracts. The results are consistent with neuroanatomy.

1.4. Outline of the paper. The outline of this paper is as follows. In section 2, we establish the statistically meaningful spherical deconvolution method and obtain a novel convex semidefinite programming model. The heuristic regularization skill is also discussed in this section. To solve the optimization model, we propose and analyze a new Peaceman–Rachford splitting algorithm in section 3. Section 4 includes the details of the new Peaceman–Rachford splitting algorithm applying to the novel convex semidefinite programming model. Numerical experiments on synthetic and real-world DW-MRI data are reported in section 5. We complete the paper with section 6 by drawing some conclusions and remarks.

2. The novel spherical deconvolution method. We start from the SD model [33]. Given a pulsed magnetic field gradient sequence with a gradient orientation $\boldsymbol{g}$ and a diffusion-weighted b-value b, the measured HARDI signal is denoted by $S(\boldsymbol{g})$. S_0 is the zero-gradient DW-MRI

signal that is indeed a standard T2 image. Then, the SD model is expressed as

$$(2.1) \qquad S(\boldsymbol{g})/S_0 = \int_{\mathbb{S}^2} f(\boldsymbol{v}) R(\boldsymbol{g}, \boldsymbol{v}) \, \mathrm{d}\boldsymbol{v},$$

where $\mathbb{S}^2 := \{\boldsymbol{v} := (v_1, v_2, v_3)^\top \mid v_1^2 + v_2^2 + v_3^2 = 1\}$ denotes a unit sphere, $\boldsymbol{v}$ is an orientation of a nerve fiber, $f(\boldsymbol{v})$ is the FOD to be estimated, and $R(\boldsymbol{g}, (0,0,1)^\top)$ is the common response function corresponding to DW-MRI signals of a typical fiber with the orientation $(0,0,1)^\top$. A detailed response function used in this paper will be shown in (2.15).

To ensure a clearly statistically meaning [11], the FOD should be nonnegative

$$(2.2) \qquad f(\boldsymbol{v}) \geq 0 \qquad \forall \, \boldsymbol{v} \in \mathbb{S}^2$$

and has a unit-mass

$$(2.3) \qquad \int_{\mathbb{S}^2} f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = 1.$$

Here, the inequality constraint (2.2) has infinitely many constraints since the orientation $\boldsymbol{v}$ takes all value on the unit sphere.

2.1. A basic convex semidefinite programming model.
We suppose the FOD function f is a homogeneous polynomial with order R, which is even for symmetry. To deal with the nonnegative constraint (2.2), our basic idea is to approximate FOD by an SOS polynomial, since an SOS polynomial is certainly nonnegative [43] and any nonnegative polynomial could be closely approximated by SOS polynomials [37]. For $\boldsymbol{v} = (v_1, v_2, v_3)^\top$, we denote a basis of $R/2$th order homogeneous polynomials as

$$(2.4) \qquad \boldsymbol{u} := (v_1^{R/2}, v_1^{R/2-1} v_2, \ldots, v_3^{R/2})^\top,$$

which contains $Q := \binom{R/2+2}{2}$ entries.[2] Suppose that there are several $R/2$th order homogeneous polynomials $\boldsymbol{c}^{(k)^\top} \boldsymbol{u}$ whose coefficient vectors are $\boldsymbol{c}^{(k)} \in \mathbb{R}^Q$ for $k = 1, \ldots, L$. We define a matrix

$$\mathbf{C} := \begin{pmatrix} \vdots \\ \boldsymbol{c}^{(k)^\top} \\ \vdots \end{pmatrix}.$$

Then, an SOS representation of FOD could be rewritten as

$$(2.5) \qquad f(\boldsymbol{v}) = \sum_{k=1}^{L} \left(\boldsymbol{c}^{(k)^\top} \boldsymbol{u} \right)^2 = (\mathbf{C}\boldsymbol{u})^\top (\mathbf{C}\boldsymbol{u}) = \boldsymbol{u}^\top (\mathbf{C}^\top \mathbf{C}) \boldsymbol{u} = \langle \boldsymbol{u}\boldsymbol{u}^\top, \mathbf{C}^\top \mathbf{C} \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the matrix inner-product such that $\langle \mathbf{A}, \mathbf{B} \rangle := \sum_{i,j} a_{ij} b_{ij}$. Note that for any given $L > 0$, the Q-by-Q matrix $\mathbf{X} := \mathbf{C}^\top \mathbf{C}$ is symmetric positive semidefinite (PSD), whose

---

[2]The binomial coefficient $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ is the number of ways of picking k unordered outcomes from n possibilities. Please see http://mathworld.wolfram.com/BinomialCoefficient.html.

entries are undetermined. Generally, a larger L can lead to a relatively more accurate approximation; however, there is no need to preset it in an elaborate way since the numerical results nearly do not depend on it. Conversely, for any $\mathbf{X}$ in the PSD matrix cone $\mathbb{S}_+^Q$, i.e., $\mathbf{X} \succeq 0$, the homogeneous polynomial $f(\boldsymbol{v}) = \langle \boldsymbol{u}\boldsymbol{u}^\top, \mathbf{X} \rangle$ is nonnegative.

On the other hand, we define

$$(2.6) \qquad \boldsymbol{\phi} := (v_1^R, v_1^{R-1} v_2, \dots, v_3^R)^\top$$

as a basis of Rth order homogeneous polynomials, which contains $P := \binom{R+2}{2}$ entries. Under this basis, FOD has a linear form

$$(2.7) \qquad f(\boldsymbol{v}) = \boldsymbol{\phi}^\top \boldsymbol{w} = \sum_{j=1}^P w_j \phi_j(\boldsymbol{v}),$$

where $\boldsymbol{w} \in \mathbb{R}^P$ is its coefficient vector. According to the basis Lemma 2.1 established in [12], there is a linear map

$$(2.8) \qquad \mathcal{A}(\mathbf{X}) := \begin{pmatrix} \vdots \\ \langle \mathbf{A}_j, \mathbf{X} \rangle \\ \vdots \end{pmatrix} : \mathbb{R}^{Q \times Q} \mapsto \mathbb{R}^P,$$

where $\mathbf{A}_j$ is a Q-by-Q matrix that indicates the location of the entry ϕ_j in the matrix $\boldsymbol{u}\boldsymbol{u}^\top$. Then, we have

$$\langle \boldsymbol{u}\boldsymbol{u}^\top, \mathbf{X} \rangle = \sum_{j=1}^P \phi_j \langle \mathbf{A}_j, \mathbf{X} \rangle = \boldsymbol{\phi}^\top \mathcal{A}(\mathbf{X}).$$

Recalling (2.7), we get the semidefinite programming (SDP) constraint

$$(2.9) \qquad \boldsymbol{w} = \mathcal{A}(\mathbf{X}) \qquad \text{and} \qquad \mathbf{X} \succeq 0.$$

The unit-mass constraint (2.3) can be reformulated as

$$\int_{\mathbb{S}^2} f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = \int_{\mathbb{S}^2} \sum_{j=1}^P w_j \phi_j(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = \sum_{j=1}^P w_j \int_{\mathbb{S}^2} \phi_j(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = 1.$$

We define

$$(2.10) \qquad s := \begin{pmatrix} \vdots \\ \int_{\mathbb{S}^2} \phi_j(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} \\ \vdots \end{pmatrix} \in \mathbb{R}^P$$

and rewrite the unit-mass constraint as

$$(2.11) \qquad \boldsymbol{w}^\top \boldsymbol{s} = 1.$$

The involved integrals in the vector $\boldsymbol{s}$ could be calculated exactly using the MATLAB symbolic integration.

The SD equation (2.1) could be processed similarly,

$$S(\boldsymbol{g})/S_0 = \int_{\mathbb{S}^2} \sum_{j=1}^{P} w_j \phi_j(\boldsymbol{v}) \, R(\boldsymbol{g}, \boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = \sum_{j=1}^{P} w_j \int_{\mathbb{S}^2} \phi_j(\boldsymbol{v}) R(\boldsymbol{g}, \boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}.$$

Suppose HARDI measures N gradient orientations $\boldsymbol{g}^{(i)}$ and generates corresponding signals $S^{(i)}$ for $i = 1, \ldots, N$. Then, we define the signal-attenuation vector

$$(2.12) \qquad \boldsymbol{f} := \begin{pmatrix} \vdots \\ S^{(i)}/S_0 \\ \vdots \end{pmatrix} \in \mathbb{R}^N$$

and the deconvolution matrix

$$(2.13) \qquad \boldsymbol{\Phi} := \begin{pmatrix} & \vdots & \\ \cdots & \int_{\mathbb{S}^2} \phi_j(\boldsymbol{v}) R(\boldsymbol{g}^{(i)}, \boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} & \cdots \\ & \vdots & \end{pmatrix} \in \mathbb{R}^{N \times P}.$$

Due to the elaborate selection of gradient orientations, the deconvolution matrix $\boldsymbol{\Phi}$ has a full column rank if $P \leq N$. Then the SD fitting could be rewritten as

$$(2.14) \qquad \boldsymbol{f} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon}$ is an unavoidable noise.

For the response function corresponding to a single fiber, we employ the bipolar Watson function [59]

$$(2.15) \qquad R(\boldsymbol{g}, \boldsymbol{v}) = \lim_{\delta \to +\infty} \exp[-\delta(\boldsymbol{g}^\top \boldsymbol{v})^2].$$

It is a refined Gaussian response corresponding to DW-MRI signal attenuation of a signal fiber, as long as δ is large enough. Here, we fix the parameter $\delta = 600$ for the purpose of obtaining a well-conditioned deconvolution matrix $\boldsymbol{\Phi}$. To compute the entries of $\boldsymbol{\Phi}$, we use the Lebedev quadrature [38]. The Lebedev quadrature on a unit sphere $\mathbb{S}^2$ is similar to the Gaussian quadrature in an interval $[-1, 1]$. They all compute a weighted sum of function values at specified points within the domain of integration. Hence, they all have a higher precision in practice.

Taking all the components (2.9)–(2.14) together, we obtain a basic convex SDP model

$$(2.16) \qquad \begin{cases} \min & \frac{1}{2}\|\boldsymbol{f} - \boldsymbol{\Phi}\boldsymbol{w}\|^2 + \mu \mathrm{Reg}(\mathbf{X}) \\ \text{s.t.} & \boldsymbol{w} = \mathcal{A}(\mathbf{X}), \quad \mathbf{X} \succeq 0, \\ & \boldsymbol{w}^\top \boldsymbol{s} = 1, \end{cases}$$

where $\mathrm{Reg}(\mathbf{X})$ is a convex regularization term and μ is its regular parameter.

2.2. A new heuristic regularization. Since the other terms are well understood, we now focus on the regularization term $\text{Reg}(\mathbf{X})$.

As an Rth order homogeneous polynomial, FOD corresponds to a unique Rth order symmetric tensor

$$\mathcal{T} = [t_{i_1 i_2 \cdots i_R}] \in \mathbb{R}^{3^R}, \tag{2.17}$$

where symmetry means that the value of $t_{i_1 i_2 \cdots i_R}$ is unchanged under any permutation of its indices. In fact, if we define $\pi(p, q)$ as the index set of all the permutations of $p \leftrightarrow 1$, $q \leftrightarrow 2$, and $(R-p-q) \leftrightarrow 3$, then FOD function $f(\boldsymbol{v})$ can be represented as the following homogeneous polynomial:

$$\begin{aligned}
(2.18) \qquad & \sum_{i_1=1}^{3} \sum_{i_2=1}^{3} \cdots \sum_{i_R=1}^{3} t_{i_1 i_2 \cdots i_R} v_{i_1} v_{i_2} \cdots v_{i_R} \\
&= \sum_{p=0}^{R} \sum_{q=0}^{R-p} \underbrace{\sum_{(i_1,i_2,\ldots,i_R)\in\pi(p,q)} t_{i_1 i_2 \cdots i_R}}_{w_j} \underbrace{v_1^p v_2^q v_3^{R-p-q}}_{\phi_j(\boldsymbol{v})} \\
&= \sum_{j=1}^{P} w_j \phi_j(\boldsymbol{v}) \\
&= f(\boldsymbol{v}).
\end{aligned}$$

Qi [46] called $(\lambda, \boldsymbol{v})$ a Z-eigenpair of a given symmetric tensor $\mathcal{T}$ if they satisfy

$$\begin{cases} \mathcal{T}\boldsymbol{v}^{R-1} = \lambda\boldsymbol{v}, \\ \boldsymbol{v}^\top \boldsymbol{v} = 1, \end{cases} \tag{2.19}$$

where the vector $\mathcal{T}\boldsymbol{v}^{R-1}$ is defined as

$$\left(\mathcal{T}\boldsymbol{v}^{R-1}\right)_i := \sum_{i_2=1}^{3} \cdots \sum_{i_R=1}^{3} t_{i i_2 \cdots i_R} v_{i_2} \cdots v_{i_R} \qquad \forall\, i = 1, 2, 3. \tag{2.20}$$

If $(\lambda, \boldsymbol{v})$ satisfies (2.19), then λ and $\boldsymbol{v}$ are called a Z-eigenvalue and its corresponding Z-eigenvector of the tensor $\mathcal{T}$, respectively. Recently, Cui, Dai, and Nie [15] proposed an interesting numerical approach which can compute all the real Z-eigenvalues of a given symmetric tensor.

In the context of a real-world DW-MRI signal analysis, Qi, Yu, and Xu [49] showed that the Z-eigenvectors corresponding to the principal Z-eigenvalues of the tensor $\mathcal{T}$ agree with the orientations of multiple fibers that cross or kiss in a voxel. In neuroanatomy, the number of nerve fibers within white matter voxels such as the brain and spinal cord is small. Therefore, an ideal regularizer is the number of the principal Z-eigenvalues of the tensor $\mathcal{T}$. However, how many Z-eigenvalues a supersymmetric tensor have is still an open problem. Qi [46] gave one of its upper bound $3(R-1)^2 - 1$.

Our idea comes from a particular function

$$(2.21) \qquad\qquad m(\lambda, \boldsymbol{u}) := \langle\, \mathbf{X} - \lambda\mathbf{E}, \boldsymbol{u}\boldsymbol{u}^\top \rangle + \lambda,$$

where $\mathbf{X}$ is the matrix estimated from the basic convex SDP model (2.16) and $\mathbf{E}$ is a Q-by-Q diagonal matrix such that

$$(2.22) \qquad\qquad \langle\, \mathbf{E}, \boldsymbol{u}(v)\boldsymbol{u}(v)^\top \rangle = (\boldsymbol{v}^\top\boldsymbol{v})^{R/2}.$$

In fact, the diagonal matrix $\mathbf{E}$ is positive definite since its diagonal entries are the corresponding coefficients of $\boldsymbol{u}_j^2(\boldsymbol{v})$ for $j = 1,\ldots,Q$ in the polynomial $(\boldsymbol{v}^\top\boldsymbol{v})^{R/2}$ and all the coefficients are positive by the multinomial theorem. We remark that the merit function (2.21) is similar to the one used in Qi [46] except that (2.21) has an additional λ.

Before introducing the novel regularization, we show a relationship between a stationary point of the merit function (2.21) and a Z-eigenpair of the interesting tensor $\mathcal{T}$.

Lemma 2.1. *Suppose the supersymmetric tensor $\mathcal{T}$ is corresponding to the homogeneous polynomial $f(\boldsymbol{v}) = \langle\, \boldsymbol{u}(\boldsymbol{v})\boldsymbol{u}(\boldsymbol{v})^\top, \mathbf{X} \rangle$. Then, the tuple $(\lambda, \boldsymbol{v})$ is a stationary point of the merit function $m(\lambda, \boldsymbol{u}(\boldsymbol{v}))$ in variables λ and $\boldsymbol{v}$ if and only if $(\lambda, \boldsymbol{v})$ is a Z-eigenpair of the tensor $\mathcal{T}$.*

Proof. The tuple $(\lambda, \boldsymbol{v})$ is a stationary point of $m(\lambda, \boldsymbol{u}(\boldsymbol{v}))$ in variables λ and $\boldsymbol{v}$ if it satisfies the following optimality condition:

$$(2.23) \qquad\qquad \begin{cases} \nabla_\lambda m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = 0, \\ \nabla_{\boldsymbol{v}} m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = 0. \end{cases}$$

Hence, it suffices to prove the equivalence between systems (2.23) and (2.19).

Obviously, we have

$$\nabla_\lambda m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = 1 - \langle\, \mathbf{E}, \boldsymbol{u}(\boldsymbol{v})\boldsymbol{u}(\boldsymbol{v})^\top \rangle = 1 - (\boldsymbol{v}^\top\boldsymbol{v})^{R/2}.$$

Hence, $\nabla_\lambda m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = 0$ is equivalent to the requirement $\boldsymbol{v}^\top\boldsymbol{v} = 1$.

According to (2.18) and (2.20), we obtain

$$f(\boldsymbol{v}) = \sum_{i_1=1}^{3} v_{i_1}\left(\sum_{i_2=1}^{3}\cdots\sum_{i_R=1}^{3} t_{i_1 i_2 \cdots i_R}v_{i_2}\cdots v_{i_R}\right) = \sum_{i_1=1}^{3} v_{i_1}(\mathcal{T}\boldsymbol{v}^{R-1})_{i_1} = \boldsymbol{v}^\top(\mathcal{T}\boldsymbol{v}^{R-1}).$$

Hence, $\nabla f(\boldsymbol{v}) = R\mathcal{T}\boldsymbol{v}^{R-1}$. Then, using this equality, we get

$$\begin{aligned} \nabla_{\boldsymbol{v}} m(\lambda, \boldsymbol{u}(\boldsymbol{v})) &= \nabla_{\boldsymbol{v}}\left(f(\boldsymbol{v}) - \lambda(\boldsymbol{v}^\top\boldsymbol{v})^{R/2}\right) \\ &= R\mathcal{T}\boldsymbol{v}^{R-1} - \lambda R(\boldsymbol{v}^\top\boldsymbol{v})^{(R-2)/2}\boldsymbol{v} \\ &= R(\mathcal{T}\boldsymbol{v}^{R-1} - \lambda\boldsymbol{v}). \end{aligned}$$

Therefore, $\nabla_{\boldsymbol{v}} m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = 0$ equals $\mathcal{T}\boldsymbol{v}^{R-1} = \lambda\boldsymbol{v}$. $\blacksquare$

Though there are several theoretical results on the Z-eigenvalues of a tensor and the little numerical methods for computing them, it is still hard to apply it in our study. We hence turn

to the stationary points of the merit function (2.21) in variables λ and $\boldsymbol{u}$, which is related to the following generalized eigenproblem:

$$(2.24) \qquad\qquad \mathbf{X}\boldsymbol{u} = \lambda \mathbf{E}\boldsymbol{u}, \qquad \boldsymbol{u} \neq 0.$$

If this equation holds, we call λ and $\boldsymbol{u}$ an eigenvalue and its corresponding eigenvector of the generalized eigenproblem (2.24), respectively.

    Lemma 2.2. The tuple $(\lambda, \boldsymbol{u})$ with $\boldsymbol{u} \neq 0$ is a stationary point of the merit function $m(\lambda, \boldsymbol{u})$ in variables λ and $\boldsymbol{u}$ if and only if λ is the eigenvalue of the generalized eigenproblem (2.24) with eigenvectors $\bar{\boldsymbol{u}}$ satisfying

$$(2.25) \qquad\qquad \boldsymbol{u} = \frac{\bar{\boldsymbol{u}}}{\sqrt{\bar{\boldsymbol{u}}^\top \mathbf{E}\bar{\boldsymbol{u}}}}.$$

    Proof. Suppose $(\lambda, \boldsymbol{u})$ is a stationary point of $m(\lambda, \boldsymbol{u})$ in variables λ and $\boldsymbol{u}$. Then, we have

$$(2.26) \qquad\qquad \nabla_\lambda m(\lambda, \boldsymbol{u}) = 1 - \langle \mathbf{E}, \boldsymbol{u}\boldsymbol{u}^\top \rangle = 1 - \boldsymbol{u}^\top \mathbf{E}\boldsymbol{u} = 0.$$

Hence, $\boldsymbol{u} \neq 0$ since $\mathbf{E}$ is positive definite. Moreover, we get

$$(2.27) \qquad\qquad \nabla_{\boldsymbol{u}} m(\lambda, \boldsymbol{u}) = \mathbf{X}\boldsymbol{u} - \lambda \mathbf{E}\boldsymbol{u} = 0.$$

Consequently, λ is an eigenvalue of the generalized eigenproblem (2.24) and any nonzero vector paralleling to $\boldsymbol{u}$ is its corresponding eigenvector.

    On the other hand, we assume that λ and $\bar{\boldsymbol{u}}$ satisfy (2.24). Since $\mathbf{E}$ is positive definite and $\bar{\boldsymbol{u}} \neq 0$, we can compute $\boldsymbol{u}$ as in (2.25). Then, equalities (2.26) and (2.27) are obviously valid. ∎

    A natural problem is whether an eigenvalue of the generalized eigenproblem (2.24) equals to a Z-eigenvalue of the supersymmetric tensor $\mathcal{T}$ in some case. The following lemma gives a positive answer.

    Lemma 2.3. Suppose λ and $\bar{\boldsymbol{u}}$ are an eigenvalue and its corresponding eigenvector of the generalized eigenproblem (2.24) and there is a vector $\boldsymbol{v} \in \mathbb{S}^2$ such that $\boldsymbol{u}(\boldsymbol{v})$ is parallel to $\bar{\boldsymbol{u}}$. Then, $(\lambda, \boldsymbol{v})$ is a Z-eigenpair of the supersymmetric tensor $\mathcal{T}$.

    Proof. From Lemmas 2.1 and 2.2, it is enough to show that

$$(2.28) \qquad\qquad \nabla_{\boldsymbol{v}} m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = 0.$$

    By the optimality condition, we have

$$\nabla_{\boldsymbol{v}} m(\lambda, \boldsymbol{u}(\boldsymbol{v})) = \mathbf{J}_{\boldsymbol{u}}^\top (\mathbf{X}\boldsymbol{u}(\boldsymbol{v}) - \lambda \mathbf{E}\boldsymbol{u}(\boldsymbol{v})),$$

where $\mathbf{J}_{\boldsymbol{u}}$ is the Jacobian of $\boldsymbol{u}(\boldsymbol{v})$. Because of the assumptions of this lemma, we have

$$\mathbf{X}\boldsymbol{u}(\boldsymbol{v}) - \lambda \mathbf{E}\boldsymbol{u}(\boldsymbol{v}) = 0.$$

Hence, the equality (2.28) holds. ∎

    Now, we devote ourselves to constructing the novel heuristic regularization using the eigenvalues of the generalized eigenproblem (2.24). From the theory of compressive sensing

[21, 9, 10], minimizing the ℓ_1 norm of all the eigenvalues of the generalized eigenproblem (2.24) is satisfactory. Since the matrix $\mathbf{E}$ is positive definite, we have

$$\mathbf{X}\boldsymbol{u} = \lambda \mathbf{E}\boldsymbol{u} \qquad \Longleftrightarrow \qquad \mathbf{E}^{-1}\mathbf{X}\boldsymbol{u} = \lambda\boldsymbol{u}.$$

Moreover, $\mathbf{E}^{-1}\mathbf{X}$ is PSD since $\mathbf{X} \succeq 0$. Therefore, the ℓ_1 norm of all the eigenvalues of $\mathbf{E}^{-1}\mathbf{X}$ is its trace

$$(2.29) \qquad \qquad \text{trace}(\mathbf{E}^{-1}\mathbf{X}) = \langle \mathbf{E}^{-1}, \mathbf{X} \rangle.$$

We remark here that this heuristic regularization differs from the nuclear norm used in [12].

Based on the basic model (2.16), we give the novel convex SDP model formally:

$$(2.30) \qquad \begin{cases} \min & \frac{1}{2}\|\boldsymbol{f} - \boldsymbol{\Phi}\boldsymbol{w}\|^2 + \mu\langle \mathbf{E}^{-1}, \mathbf{X}\rangle \\ \text{s.t.} & \boldsymbol{w} = \mathcal{A}(\mathbf{X}), \quad \mathbf{X} \succeq 0, \\ & \boldsymbol{w}^\top \boldsymbol{s} = 1. \end{cases}$$

The global solution of this model exists.

3. A new Peaceman–Rachford splitting method.
In this section, we devote ourselves to a new PRSM which could solve the novel convex SDP efficiently. The new algorithm comprises two steps in each iteration. In the first prediction step, we employ the PRSM to generate a search direction. Here, an overdetermined relaxation factor is admitted in the second updating of the multiplier. The second step is a correction step. We compute a suitable step size to promote the convergence of the new algorithm.

For convenience of our analysis, we consider the following general separable convex optimization problem:

$$(3.1) \qquad \begin{cases} \min & \theta_1(\boldsymbol{x}) + \theta_2(\boldsymbol{y}) \\ \text{s.t.} & \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c} = 0, \\ & \boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{y} \in \mathcal{Y}, \end{cases}$$

where $\mathcal{X} \subseteq \mathbb{R}^m$ and $\mathcal{Y} \subseteq \mathbb{R}^n$ are nonempty closed convex sets, $\mathbf{A} \in \mathbb{R}^{\ell \times m}$, $\mathbf{B} \in \mathbb{R}^{\ell \times n}$, $\boldsymbol{c} \in \mathbb{R}^{\ell}$, and $\theta_1(\boldsymbol{x})$ and $\theta_2(\boldsymbol{y})$ are convex and continuously differentiable functions.[3] Hence, the set of optimal solutions of (3.1), which we assumed is nonempty, is convex and closed.

The convex optimization (3.1) has an equivalent variational inequality (VI) reformulation [29]: find $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^{\ell}$ such that

$$\begin{cases} (\boldsymbol{x} - \boldsymbol{x}^*)^\top [\nabla\theta_1(\boldsymbol{x}^*) - \mathbf{A}^\top\boldsymbol{\lambda}^*] \geq 0 & \forall\, \boldsymbol{x} \in \mathcal{X}, \\ (\boldsymbol{y} - \boldsymbol{y}^*)^\top [\nabla\theta_2(\boldsymbol{y}^*) - \mathbf{B}^\top\boldsymbol{\lambda}^*] \geq 0 & \forall\, \boldsymbol{y} \in \mathcal{Y}, \\ (\boldsymbol{\lambda} - \boldsymbol{\lambda}^*)^\top [\mathbf{A}\boldsymbol{x}^* + \mathbf{B}\boldsymbol{y}^* - \boldsymbol{c}] \geq 0 & \forall\, \boldsymbol{\lambda} \in \mathbb{R}^{\ell}. \end{cases}$$

---

[3]The functions $\theta_1(\boldsymbol{x})$ and $\theta_2(\boldsymbol{y})$ could be proper convex and lower semicontinuous if we slightly modify our proofs. However, it is more readable in the current setting.

Define $\boldsymbol{w} := (\boldsymbol{x}^\top, \boldsymbol{y}^\top, \boldsymbol{\lambda}^\top)^\top$, $\Omega := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^\ell$, and the map

$$(3.2) \qquad F(\boldsymbol{w}) := \begin{pmatrix} \nabla\theta_1(\boldsymbol{x}) - \mathbf{A}^\top\boldsymbol{\lambda} \\ \nabla\theta_2(\boldsymbol{y}) - \mathbf{B}^\top\boldsymbol{\lambda} \\ \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c} \end{pmatrix},$$

the above VI can be formulated in the following compact form: find $\boldsymbol{w}^* \in \Omega$, such that

$$(3.3) \qquad (\boldsymbol{w} - \boldsymbol{w}^*)^\top F(\boldsymbol{w}^*) \geq 0, \quad \forall \boldsymbol{w} \in \Omega.$$

Since θ_1 and θ_2 are convex, the map $F(\cdot)$ is monotone. We remark that VI (3.3) is critical for our following analysis.

3.1. The predication step. The augment Lagrangian function of the separable convex optimization (3.1) is

$$(3.4) \qquad \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = \theta_1(\boldsymbol{x}) + \theta_2(\boldsymbol{y}) - \boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c}) + \frac{\beta}{2}\|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c}\|^2,$$
$$(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^\ell,$$

where β is a positive penalty parameter.

The PRSM is expressed as follows.

Prediction step. In iteration k, we set $\alpha \in (0, 1)$ and $\gamma \in [1, \infty)$. And $(\boldsymbol{y}^k, \boldsymbol{\lambda}^k)$ is the current iterate. We perform the following four steps successively:

$$(3.5) \qquad \widetilde{\boldsymbol{x}}^k = \operatorname{argmin}\left\{\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\lambda}^k) \ : \ \boldsymbol{x} \in \mathcal{X}\right\},$$

$$(3.6) \qquad \boldsymbol{\lambda}^{k+\frac{1}{2}} = \boldsymbol{\lambda}^k - \alpha\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c}),$$

$$(3.7) \qquad \widetilde{\boldsymbol{y}}^k = \operatorname{argmin}\{\mathcal{L}(\widetilde{\boldsymbol{x}}^k, \boldsymbol{y}, \boldsymbol{\lambda}^{k+\frac{1}{2}}) \ : \ \boldsymbol{y} \in \mathcal{Y}\},$$

$$(3.8) \qquad \widetilde{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda}^{k+\frac{1}{2}} - \gamma\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c}).$$

We comment that the original PRSM sets $\alpha = \gamma = 1$. If $\alpha = 0$ and $\gamma = 1$, it is the Douglas–Rachford splitting method. Recently, He et al. [28] introduced an underdetermined relaxation factor $\alpha = \gamma \in (0, 1)$ to construct a strictly contractive algorithm. Remarkably, we admit an overdetermined relaxation factor $\gamma \in [1, \infty)$, which could accelerate the convergence of the new algorithm.

We now show some useful properties on the prediction step. To this purpose, we first define two important matrices:

$$(3.9) \qquad \mathbf{M} := \begin{pmatrix} \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta\mathbf{B}^\top\mathbf{B} & -\frac{\alpha}{\alpha+\gamma}\mathbf{B}^\top \\ -\frac{\alpha}{\alpha+\gamma}\mathbf{B} & \frac{1}{\alpha+\gamma}\beta^{-1}\mathbf{I}_\ell \end{pmatrix}$$

and

$$(3.10) \qquad \mathbf{Q} := \begin{pmatrix} \frac{(\alpha+\gamma)^2-\alpha\gamma(\alpha+\gamma+1)}{(\alpha+\gamma)^2}\beta\mathbf{B}^\top\mathbf{B} & \frac{-\alpha(\alpha+\gamma+1)+\gamma}{2(\alpha+\gamma)^2}\mathbf{B}^\top \\ \frac{-\alpha(\alpha+\gamma+1)+\gamma}{2(\alpha+\gamma)^2}\mathbf{B} & \frac{1}{(\alpha+\gamma)^2}\beta^{-1}\mathbf{I}_\ell \end{pmatrix}.$$

In the following lemma, we show that $\mathbf{M}$ and $\mathbf{Q}$ are PSD.

Lemma 3.1. *Suppose $\alpha \in [0, 1)$, $\gamma \in [1, \infty)$. Then the matrix $\mathbf{M}$ is PSD. Additionally, $\mathbf{M}$ is positive definite if the matrix $\mathbf{B}$ has a full column rank.*

Proof. By some calculations, we have

$$\mathbf{M} = \begin{pmatrix} \sqrt{\beta}\mathbf{B}^\top & 0 \\ 0 & \frac{1}{\sqrt{\beta}}\mathbf{I}_\ell \end{pmatrix} \begin{pmatrix} \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\mathbf{I}_\ell & -\frac{\alpha}{\alpha+\gamma}\mathbf{I}_\ell \\ -\frac{\alpha}{\alpha+\gamma}\mathbf{I}_\ell & \frac{1}{\alpha+\gamma}\mathbf{I}_\ell \end{pmatrix} \begin{pmatrix} \sqrt{\beta}\mathbf{B} & 0 \\ 0 & \frac{1}{\sqrt{\beta}}\mathbf{I}_\ell \end{pmatrix}$$

and

$$(3.11) \qquad \begin{pmatrix} \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\mathbf{I}_\ell & -\frac{\alpha}{\alpha+\gamma}\mathbf{I}_\ell \\ -\frac{\alpha}{\alpha+\gamma}\mathbf{I}_\ell & \frac{1}{\alpha+\gamma}\mathbf{I}_\ell \end{pmatrix} = \frac{1}{\alpha+\gamma} \begin{pmatrix} \alpha+\gamma-\alpha\gamma & -\alpha \\ -\alpha & 1 \end{pmatrix} \otimes \mathbf{I}_\ell,$$

where $\otimes$ denotes the matrix Kronecker product [30]. The matrix Kronecker product enjoys a useful property: for any matrices $\mathbf{X}$ and $\mathbf{Y}$, the product of their eigenvalues $\lambda(\mathbf{X})\lambda(\mathbf{Y})$ is the eigenvalue of $\mathbf{X} \otimes \mathbf{Y}$. Hence, to show the positive definiteness of the matrix (3.11), we need only show that the 2-by-2 matrix

$$\begin{pmatrix} \alpha+\gamma-\alpha\gamma & -\alpha \\ -\alpha & 1 \end{pmatrix}$$

is positive definite. Obviously,

$$\alpha + \gamma - \alpha\gamma - \alpha^2 = (1-\alpha)(\alpha+\gamma) > 0.$$

Therefore, the matrix (3.11) is positive definite and $\mathbf{M}$ is PSD.

The second statement is straightforward. ∎

Lemma 3.2. *Suppose $\alpha \in [0, 1)$ and $\gamma \in [1, \infty)$. Then the matrix*

$$2(\alpha+\gamma)\mathbf{Q} - (\alpha+1)\mathbf{M}$$

is PSD.

Moreover, the matrix $\mathbf{Q}$ is PSD.

Proof. By some calculations, we have

$$2(\alpha+\gamma)\mathbf{Q} - (\alpha+1)\mathbf{M} = \frac{1}{\alpha+\gamma} \begin{pmatrix} \Xi\beta\mathbf{B}^\top\mathbf{B} & \gamma(1-\alpha)\mathbf{B}^\top \\ \gamma(1-\alpha)\mathbf{B} & (1-\alpha)\beta^{-1}\mathbf{I}_\ell \end{pmatrix},$$

where $\Xi := (\alpha+\gamma-\alpha\gamma)(2\gamma+\alpha-1) - 2\alpha\gamma$. Similar to the proof in Lemma 3.1, we show that

$$\begin{aligned}
&[(\alpha+\gamma-\alpha\gamma)(2\gamma+\alpha-1) - 2\alpha\gamma](1-\alpha) - [\gamma(1-\alpha)]^2 \\
&= (1-\alpha)[2(1-\alpha)\gamma^2 - (1-\alpha)^2\gamma + 2\alpha\gamma + \alpha(\alpha-1) - 2\alpha\gamma - (1-\alpha)\gamma^2] \\
&= (1-\alpha)[(1-\alpha)\gamma^2 - (1-\alpha)^2\gamma + \alpha(\alpha-1)] \\
&= (1-\alpha)^2[\gamma^2 - (1-\alpha)\gamma - \alpha] \\
&= (1-\alpha)^2(\gamma-1)(\gamma+\alpha).
\end{aligned}$$

Since $\gamma \geq 1$, this lemma is valid. ■

For the convenience of our analysis, we introduce a new variable

$$(3.12) \qquad \boldsymbol{z} := (\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top)^\top$$

and define

$$(3.13) \qquad \|\boldsymbol{z}\|_{\mathbf{W}} := \sqrt{\boldsymbol{z}^\top \mathbf{W} \boldsymbol{z}}$$

if the involved matrix $\mathbf{W}$ is PSD.

Theorem 3.3. *Suppose that $\widetilde{\boldsymbol{z}}^k$ is generated by the predication step from $\boldsymbol{z}^k$ and the matrix $\mathbf{M}$ is defined as* (3.9). *Then, if $\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}} = 0$, $\widetilde{\boldsymbol{w}}^k$ is a solution of the VI* (3.3).

Proof. See the appendix for the proof. ■

For the prediction step, we have the following main result.

Theorem 3.4. *Suppose $\boldsymbol{w}^*$ is a solution of VI* (3.3) *and parameters $\alpha \in [0,1)$ and $\gamma \in [1,\infty)$. Then, we have*

$$(3.14) \qquad (\boldsymbol{z}^* - \boldsymbol{z}^k)^\top \mathbf{M}(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k) \geq \|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{Q}}^2 \geq 0.$$

Proof. See the appendix for the proof. ■

This theorem shows that $\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k$ is a decreasing direction for the merit function $\|\cdot - \boldsymbol{z}^*\|_{\mathbf{M}}^2$. This phenomenon is critical for the following correction step.

3.2. Correction step. We define the correction step as the following form [61]:

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \rho(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k),$$

where ρ is a step size that will be determined soon. By some calculations, we have

$$
\begin{aligned}
\|\boldsymbol{z}^k - \boldsymbol{z}^*\|_{\mathbf{M}}^2 - \|\boldsymbol{z}^{k+1} - \boldsymbol{z}^*\|_{\mathbf{M}}^2 &= \|\boldsymbol{z}^k - \boldsymbol{z}^*\|_{\mathbf{M}}^2 - \|\boldsymbol{z}^k - \boldsymbol{z}^* + \rho(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k)\|_{\mathbf{M}}^2 \\
&= 2\rho(\boldsymbol{z}^* - \boldsymbol{z}^k)^\top \mathbf{M}(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k) - \rho^2 \|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2 \\
&\geq 2\rho\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{Q}}^2 - \rho^2 \|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2,
\end{aligned}
$$

$$(3.15)$$

where the last inequality holds for Theorem 3.4. To obtain a maximal decrease in the merit function $\|\cdot - \boldsymbol{z}^*\|_{\mathbf{M}}^2$, we compute the maximizer of its lower approximation (3.15)

$$(3.16) \qquad \rho_k := \frac{\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{Q}}^2}{\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2}.$$

In practice, to accelerate the convergence of iterates, we use an additional extrapolation in the correction step.

Correction step. Set $\varsigma \in [1,2)$. We compute a step size ρ_k by (3.16) and generate the new iterate:

$$(3.17) \qquad \boldsymbol{z}^{k+1} := \boldsymbol{z}^k + \varsigma\rho_k(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k).$$

According to Lemma 3.2, we get the lower bound of the step size ρ_k.

Lemma 3.5. *Suppose $\alpha \in (0,1)$ and $\gamma \in [1,\infty)$. Then*

$$(3.18) \qquad \rho_k \geq \frac{\alpha+1}{2(\alpha+\gamma)}.$$

Then, we can establish an interesting decreasing theorem.

Theorem 3.6. *Suppose $\boldsymbol{w}^*$ is a solution of VI (3.3) and parameters $\alpha \in (0,1)$, $\gamma \in [1,\infty)$, and $\varsigma \in [1,2)$. Then, we get*

$$(3.19) \qquad \|\boldsymbol{z}^k - \boldsymbol{z}^*\|_{\mathbf{M}}^2 - \|\boldsymbol{z}^{k+1} - \boldsymbol{z}^*\|_{\mathbf{M}}^2 \geq \frac{\varsigma(2-\varsigma)(\alpha+1)^2}{4(\alpha+\gamma)^2}\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2.$$

Proof. From (3.15), (3.16), and Lemma 3.5, we have

$$\begin{aligned}
\|\boldsymbol{z}^k - \boldsymbol{z}^*\|_{\mathbf{M}}^2 - \|\boldsymbol{z}^{k+1} - \boldsymbol{z}^*\|_{\mathbf{M}}^2 &\geq 2\varsigma\rho_k\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{Q}}^2 - \varsigma^2\rho_k^2\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2 \\
&= \varsigma(2-\varsigma)\rho_k^2\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2 \\
&\geq \varsigma(2-\varsigma)\frac{(\alpha+1)^2}{4(\alpha+\gamma)^2}\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}}^2.
\end{aligned}$$

This theorem is proved. ∎

Finally, we show the main convergence theorem.

Theorem 3.7. *Suppose $\boldsymbol{w}^*$ is a solution of VI (3.3) and parameters $\alpha \in (0,1)$, $\gamma \in [1,\infty)$, and $\varsigma \in [1,2)$. Then, we have the following assertions:*

(i) *The sequence $\{\|\boldsymbol{z}^k - \boldsymbol{z}^*\|_{\mathbf{M}}\}$ is monotonically nonincreasing.*

(ii) $\lim_{k\to\infty} \|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}} = 0$.

(iii) *If matrices $\mathbf{A}$ and $\mathbf{B}$ have full column ranks, the complete sequence $\{\widetilde{\boldsymbol{w}}^k\}$ converges to a solution of VI (3.3).*

Proof. From Theorem 3.6, statements (i) and (ii) are straightforward.

If $\mathbf{B}$ has a full column rank, the matrix $\mathbf{M}$ is positive definite by Lemma 3.1. Then, from statement (i), we have $\{\boldsymbol{z}^k\}$ is bounded. The sequence $\{\widetilde{\boldsymbol{z}}^k\}$ is also bounded for statement (ii). From (3.6) and (3.8), we have

$$(3.20) \qquad (\alpha+\gamma)\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c}) = \alpha\beta\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) - (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k).$$

Since $\mathbf{A}$ has a full column rank, (3.20) and the boundedness of $\{\boldsymbol{z}^k\}$ and $\{\widetilde{\boldsymbol{z}}^k\}$ indicate that $\{\widetilde{\boldsymbol{x}}^k\}$ is bounded. Hence, $\{(\widetilde{\boldsymbol{x}}^k, \widetilde{\boldsymbol{z}}^k)\}$ has one accumulation point $(\boldsymbol{x}^{\infty\top}, \boldsymbol{z}^{\infty\top})^\top$ at least. Suppose $\{(\widetilde{\boldsymbol{x}}^{k_i}, \widetilde{\boldsymbol{z}}^{k_i})\}$ is a subsequence that converges to $(\boldsymbol{x}^{\infty\top}, \boldsymbol{z}^{\infty\top})^\top$. Then, $\{\boldsymbol{z}^{k_i}\}$ also converges to $\boldsymbol{z}^\infty$ due to condition (ii). Taking limit along this subsequence, we have from (3.20) that

$$\mathbf{A}\boldsymbol{x}^\infty + \mathbf{B}\boldsymbol{y}^\infty = \boldsymbol{c},$$

and from the optimality conditions of (3.5) and (3.7) that

$$(\boldsymbol{x} - \boldsymbol{x}^\infty)^\top[\nabla\theta_1(\boldsymbol{x}^\infty) - \mathbf{A}^\top\boldsymbol{\lambda}^\infty] \geq 0 \qquad \forall\, \boldsymbol{x} \in \mathcal{X}$$

and

$$(\boldsymbol{y} - \boldsymbol{y}^\infty)^\top[\nabla\theta_2(\boldsymbol{y}^\infty) - \mathbf{B}^\top\boldsymbol{\lambda}^\infty] \geq 0 \qquad \forall\, \boldsymbol{y} \in \mathcal{Y},$$

respectively, indicating that $(\boldsymbol{x}^{\infty\top}, \boldsymbol{z}^{\infty\top})^\top$ is a solution of VI (3.3).

Note that

$$\lim_{i\to\infty} \|\boldsymbol{z}^{k_i} - \boldsymbol{z}^\infty\|_{\mathbf{M}} = 0,$$

and by statement (i), we have

$$\|\boldsymbol{z}^p - \boldsymbol{z}^\infty\|_{\mathbf{M}} \le \|\boldsymbol{z}^{k_i} - \boldsymbol{z}^\infty\|_{\mathbf{M}}$$

for all integers $p > k_i$. Therefore, we obtain

$$\lim_{k\to\infty} \|\boldsymbol{z}^k - \boldsymbol{z}^\infty\|_{\mathbf{M}} = 0,$$

meaning $\{\boldsymbol{z}^k\}$ converges; so does $\{\widetilde{\boldsymbol{z}}^k\}$. Finally, by (3.20) $\{\widetilde{\boldsymbol{x}}^k\}$ converges to $\boldsymbol{x}^\infty$. ∎

4. Implementing issues. To solve the novel convex SDP model (2.30) efficiently, we turn to solve its dual problem

$$(4.1) \qquad \begin{cases} \min & \dfrac{1}{2}(\boldsymbol{\xi} + \sigma\boldsymbol{s} + \boldsymbol{\Phi}^\top\boldsymbol{f})^\top(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}(\boldsymbol{\xi} + \sigma\boldsymbol{s} + \boldsymbol{\Phi}^\top\boldsymbol{f}) - \sigma - \dfrac{1}{2}\|\boldsymbol{f}\|^2 \\ \text{s.t.} & \mathcal{A}^*\boldsymbol{\xi} - \mathbf{Y} + \mu\mathbf{E}^{-1} = 0, \quad \mathbf{Y} \succeq 0, \end{cases}$$

where $\boldsymbol{\xi} \in \mathbb{R}^P$, $\sigma \in \mathbb{R}$, and $\mathbf{Y} \in \mathbb{S}_+^Q$ are multipliers corresponding to constraints of (2.30) $\boldsymbol{w} = \mathcal{A}(\mathbf{X})$, $\boldsymbol{w}^\top\boldsymbol{s} = 1$, and $\mathbf{X} \succeq 0$, respectively. Since there exists a positive definite matrix $\mathbf{X} = \frac{1}{4\pi}\mathbf{E}$ satisfying all the constraint of (2.30), i.e., the Slater's condition holds, the strong duality is valid [7]. Hence, the primal-dual solution of (4.1) is also a solution of its primal problem (2.30).

From Lemma 2.1 of [12], $\mathcal{A}\mathcal{A}^*$ is a positive definite diagonal matrix. Hence, $\mathcal{A}^*$ has a full column rank. Furthermore, the dual problem (4.1) satisfies all the assumptions in Theorem 3.7. Therefore, the sequence generated by the new Peaceman–Rachford splitting algorithm converges to a global solution of problems (4.1) and (2.30).

Now, we implement the new PRSM to solve the dual problem (4.1). The augmented Lagrangian function of (4.1) is

$$(4.2) \qquad \begin{aligned} \mathcal{L}(\boldsymbol{\xi}, \sigma, \mathbf{Y}, \mathbf{X}) &:= \frac{1}{2}(\boldsymbol{\xi} + \sigma\boldsymbol{s} + \boldsymbol{\Phi}^\top\boldsymbol{f})^\top(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}(\boldsymbol{\xi} + \sigma\boldsymbol{s} + \boldsymbol{\Phi}^\top\boldsymbol{f}) - \sigma - \frac{1}{2}\|\boldsymbol{f}\|^2 \\ &\quad - \langle \mathbf{X}, \mathcal{A}^*\boldsymbol{\xi} - \mathbf{Y} + \mu\mathbf{E}^{-1}\rangle + \frac{\beta}{2}\|\mathcal{A}^*\boldsymbol{\xi} - \mathbf{Y} + \mu\mathbf{E}^{-1}\|_F^2, \end{aligned}$$

where the variable $\mathbf{Y} \in \mathbb{S}_+^Q$, β is a positive penalty parameter, and $\mathbf{X}$ is the multiplier of the equality constraint of the dual problem (4.1) and it is also a variable of the primal problem (2.30).

Set parameters $\alpha \in [0, 1)$, $\gamma \in [1, \infty)$, $\varsigma \in [1, 2)$, and $k \leftarrow 1$. Given an initial point $(\mathbf{Y}^0, \mathbf{X}^0) \in \mathbb{S}_+^Q \times \mathbb{S}^Q$, we alternately perform the following steps.

(1) Update $\boldsymbol{\xi}$ and σ. We minimize

$$\begin{aligned} (\widetilde{\boldsymbol{\xi}}^k, \widetilde{\sigma}^k) &= \operatorname*{argmin}\left\{\mathcal{L}(\boldsymbol{\xi}, \sigma, \mathbf{Y}^k, \mathbf{X}^k)\right\} \\ &= \operatorname*{argmin}\left\{\frac{1}{2}(\boldsymbol{\xi} + \sigma\boldsymbol{s} + \boldsymbol{\Phi}^\top\boldsymbol{f})^\top(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}(\boldsymbol{\xi} + \sigma\boldsymbol{s} + \boldsymbol{\Phi}^\top\boldsymbol{f}) - \sigma \right. \\ &\qquad\quad \left. + \frac{\beta}{2}\|\mathcal{A}^*\boldsymbol{\xi} - \mathbf{Y}^k + \mu\mathbf{E}^{-1} - \beta^{-1}\mathbf{X}^k\|_F^2\right\}. \end{aligned}$$

Its optimality condition is

$$(4.3) \qquad (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1}(\widetilde{\boldsymbol{\xi}}^k + \widetilde{\sigma}^k \boldsymbol{s} + \boldsymbol{\Phi}^\top \boldsymbol{f}) + \beta \mathcal{A}(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k - \mathbf{Y}^k + \mu \mathbf{E}^{-1} - \beta^{-1} \mathbf{X}^k) = 0,$$

$$(4.4) \qquad \boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1}(\widetilde{\boldsymbol{\xi}}^k + \widetilde{\sigma}^k \boldsymbol{s} + \boldsymbol{\Phi}^\top \boldsymbol{f}) - 1 = 0.$$

By (4.4), we have

$$(4.5) \qquad \widetilde{\sigma}^k = \frac{1 - \boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1}(\widetilde{\boldsymbol{\xi}}^k + \boldsymbol{\Phi}^\top \boldsymbol{f})}{\boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{s}}.$$

Then, we substitute (4.5) into (4.3) and get

$$
\begin{aligned}
(4.6) \qquad \widetilde{\boldsymbol{\xi}}^k = &- \left[(\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \left(\mathbf{I}_P - \frac{\boldsymbol{s}\boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1}}{\boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{s}} \right) + \beta \mathcal{A}\mathcal{A}^* \right]^{-1} \\
&\cdot \left[(\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \left(\boldsymbol{\Phi}^\top \boldsymbol{f} + \frac{1 - \boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{f}}{\boldsymbol{s}^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{s}} \boldsymbol{s} \right) - \beta \mathcal{A}(\mathbf{Y}^k - \mu \mathbf{E}^{-1} + \beta^{-1} \mathbf{X}^k) \right].
\end{aligned}
$$

We note that $\widetilde{\sigma}^k$ is unused in the following steps. So we do not calculate it.

(2) The first update of $\mathbf{X}$:

$$(4.7) \qquad \mathbf{X}^{k+\frac{1}{2}} = \mathbf{X}^k - \alpha\beta(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k - \mathbf{Y}^k + \mu \mathbf{E}^{-1}).$$

(3) Update $\mathbf{Y}$. We compute

$$
\begin{aligned}
\widetilde{\mathbf{Y}}^k &= \operatorname{argmin} \left\{ \mathcal{L}(\widetilde{\boldsymbol{\xi}}^k, \widetilde{\sigma}^k, \mathbf{Y}, \mathbf{X}^{k+\frac{1}{2}}) \ : \ \mathbf{Y} \in \mathbb{S}_+^Q \right\} \\
&= \operatorname{argmin} \left\{ \frac{\beta}{2} \| \mathbf{Y} - \mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k - \mu \mathbf{E}^{-1} + \beta^{-1} \mathbf{X}^{k+\frac{1}{2}} \|_F^2 \ : \ \mathbf{Y} \in \mathbb{S}_+^Q \right\} \\
(4.8) \qquad &= \mathcal{P}(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k + \mu \mathbf{E}^{-1} - \beta^{-1} \mathbf{X}^{k+\frac{1}{2}}),
\end{aligned}
$$

where $\mathcal{P}(\cdot)$ is the projection onto the PSD matrix cone $\mathbb{S}_+^Q$.

(4) The second update of $\mathbf{X}$:

$$(4.9) \qquad \widetilde{\mathbf{X}}^k = \mathbf{X}^{k+\frac{1}{2}} - \gamma\beta(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k - \widetilde{\mathbf{Y}}^k + \mu \mathbf{E}^{-1}).$$

(5) Compute $\varphi = \beta \|\widetilde{\mathbf{Y}}^k - \mathbf{Y}^k\|_F^2$, $\chi = -\langle \widetilde{\mathbf{Y}}^k - \mathbf{Y}^k, \widetilde{\mathbf{X}}^k - \mathbf{X}^k \rangle$, and $\psi = \beta^{-1} \|\widetilde{\mathbf{X}}^k - \mathbf{X}^k\|_F^2$. Then, the step size is

$$(4.10) \qquad \rho_k = \frac{[(\alpha + \gamma)^2 - \alpha\gamma(\alpha + \gamma + 1)]\varphi - [\alpha(\alpha + \gamma + 1) - \gamma]\chi + \psi}{(\alpha + \gamma)[(\alpha + \gamma - \alpha\gamma)\varphi - 2\alpha\chi + \psi]}.$$

(6) Generate the new iterates:

$$(4.11) \qquad \mathbf{Y}^{k+1} = \mathbf{Y}^k + \varsigma\rho_k(\widetilde{\mathbf{Y}}^k - \mathbf{Y}^k),$$
$$(4.12) \qquad \mathbf{X}^{k+1} = \mathbf{X}^k + \varsigma\rho_k(\widetilde{\mathbf{X}}^k - \mathbf{X}^k).$$

(7) Set $k \leftarrow k + 1$.

According to Theorem 3.3, we terminate the algorithm if $\|\widetilde{\mathbf{Y}}^k - \mathbf{Y}^k\|_F$ and $\|\widetilde{\mathbf{X}}^k - \mathbf{X}^k\|_F$ are sufficiently small. The resulting coefficients of FOD are $\boldsymbol{w} = \mathcal{A}(\mathcal{P}(\widetilde{\mathbf{X}}^k))$.

How to determine the regular parameter μ? This is a challenging problem for the regularization model (2.30). In the tissue environment of human white matter, the number of nerve fibers is small. Heuristically, the PSD matrix $\mathbf{X}$, which is the coefficient matrix of FOD $f(\boldsymbol{v}) = \boldsymbol{u}^\top \mathbf{X} \boldsymbol{u}$, should be of low rank. Based on this viewpoint, we choose the regular parameter μ iteratively such that the generated matrix $\mathbf{X}$ has rank three at most. In the context of the classical augmented Lagrangian method [41], when variables $\widetilde{\boldsymbol{\xi}}^k$ and $\widetilde{\mathbf{Y}}^k$ are updated, the new multiplier $\mathbf{X}$ should be

$$
\begin{aligned}
&\mathbf{X}^{k+\frac{1}{2}} - \beta(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k - \widetilde{\mathbf{Y}}^k + \mu \mathbf{E}^{-1}) \\
&= \beta\big[\mathcal{P}(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k + \mu \mathbf{E}^{-1} - \beta^{-1} \mathbf{X}^{k+\frac{1}{2}}) - (\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k + \mu \mathbf{E}^{-1} - \beta^{-1} \mathbf{X}^{k+\frac{1}{2}})\big] \\
&= \beta \mathcal{P}\left(-(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k + \mu \mathbf{E}^{-1} - \beta^{-1} \mathbf{X}^{k+\frac{1}{2}})\right).
\end{aligned}
$$

Since $\mathbf{E}^{-1}$ is a positive definite diagonal matrix, we have

$$
-(\mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k + \mu \mathbf{E}^{-1} - \beta^{-1} \mathbf{X}^{k+\frac{1}{2}}) = \mathbf{E}^{-\frac{1}{2}}(\mathbf{E}^{\frac{1}{2}}(\beta^{-1} \mathbf{X}^{k+\frac{1}{2}} - \mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k)\mathbf{E}^{\frac{1}{2}} - \mu \mathbf{I}_Q)\mathbf{E}^{-\frac{1}{2}}.
$$

Suppose the matrix $\mathbf{E}^{\frac{1}{2}}(\beta^{-1} \mathbf{X}^{k+\frac{1}{2}} - \mathcal{A}^* \widetilde{\boldsymbol{\xi}}^k)\mathbf{E}^{\frac{1}{2}}$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_Q$. Then, we take

(4.13)
$$
\mu^{k+1} := \max(\lambda_4, 0)
$$

for the $(k+1)$th iteration.

5. Numerical experiments. To illustrate the power of the novel convex SDP model and the efficiency of the new PRSM for estimating a statistically meaningful FOD, we perform the following experiments.

5.1. Synthetic simulations. A multitensor model [18] is employed here to generate the synthetic data. In one voxel, we assume there are K crossed fibers with volume factions $\nu_k > 0$ and $\sum_{k=1}^{K} \nu_k = 1$. Each fiber corresponds to a diffusion tensor $\mathbf{D}_k$, whose principle eigenvector is the orientation of a predesigned fiber. All the diffusion tensors share common eigenvalues $(1.7, 0.2, 0.2) \times 10^{-3} mm^2/s$. Then, the truth signals of this voxel are the weighted sum of signals generated by these diffusion tensors

(5.1)
$$
S^{(i)} = S_0 \sum_{k=1}^{K} \nu_k \exp(-b \boldsymbol{g}^{(i)^\top} \mathbf{D}_k \boldsymbol{g}^{(i)}),
$$

where $S_0 = 1$ is a zero-gradient signal, $b = 3000 s/mm^2$ is a diffusion-weighted b-value, and 81 gradient orientations $\boldsymbol{g}^{(i)}$, which are generated by a second order tessellation of the icosahedron, are located on a unit semisphere.

In the context of DW-MRI, measured signals usually have complex-valued noise. Hence, we corrupt the truth signals of one voxel by the Rician noise [18]:
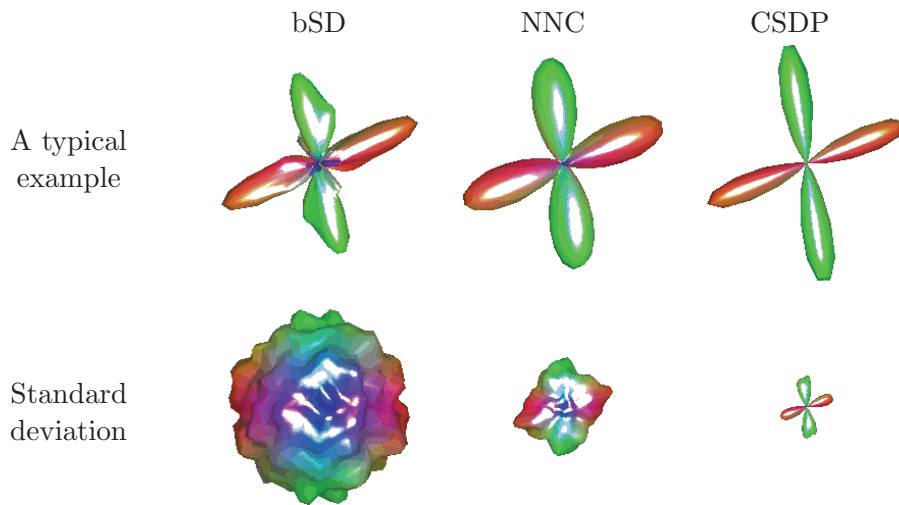
$$
S_{noisy}^{(i)} = \sqrt{(S^{(i)} + n_r)^2 + n_i^2},
$$

Figure 1. *We illustrate the FODs of a typical example estimated by three different models: bSD, NNC, and novel CSDP. The standard deviations of 100 tests are also shown.*

where $n_r \sim \mathcal{N}(0, \sigma^2)$ and $n_i \sim \mathcal{N}(0, \sigma^2)$ are independent Gaussian white noise that simulate the real part and the imaginary part of the Rician noise, respectively. We define the signal-to-noise ratio $\mathrm{SNR} := S_0/\sigma$ to describe the level of Rician noise.

Power of constraints. In the convex SDP model (2.30), there are two constraints: nonnegativity and unit-mass. We now give a typical example to illustrate the necessity of these two constraints. For this purpose, we examine the following three models.

- (bSD) The basic SD model without any constraints:

$$\min \quad \tfrac{1}{2}\|\boldsymbol{f} - \boldsymbol{\Phi w}\|^2.$$

- (NNC) Based on the bSD, we consider an additional nonnegative constraint:

$$\begin{cases} \min & \tfrac{1}{2}\|\boldsymbol{f} - \boldsymbol{\Phi w}\|^2 + \mu\langle \mathbf{E}^{-1}, \mathbf{X}\rangle \\ \text{s.t.} & \boldsymbol{w} = \mathcal{A}(\mathbf{X}), \quad \mathbf{X} \succeq 0. \end{cases}$$

- (CSDP) The new CSDP model (2.30).

The first row of Figure 1 shows the detailed improvement of contour profiles of estimated FODs when nonnegativity and unit-mass constraints are imposed. A typical example uses two fibers with a crossed angle $80°$ and the signal-to-noise ratio $\mathrm{SNR} = 20$. For the convenience of comparisons, we normalize each FOD by its maximum. A spot of negative values are detected when we examine the FOD estimated by bSD. Hence, we should impose the nonnegative constraint. Moreover, when we enforce the unit-mass constraint, the contour profile of FOD is slender. This feature is favorable for us to abstract the detailed fiber orientations.

The last row of Figure 1 shows the standard deviations of estimated FODs. The mean values of estimated FODs are omitted since they are almost the same. Here, we perform 100 random tests and keep $\mathrm{SNR} = 20$. Compared with bSD, nonnegativity and unit-mass

constraints could efficiently reduce the magnitude of error of estimated FODs. Furthermore, since the standard deviation corresponding to the novel model has almost the same contour profile to the estimated FOD, we observe a directional error which surprises us.
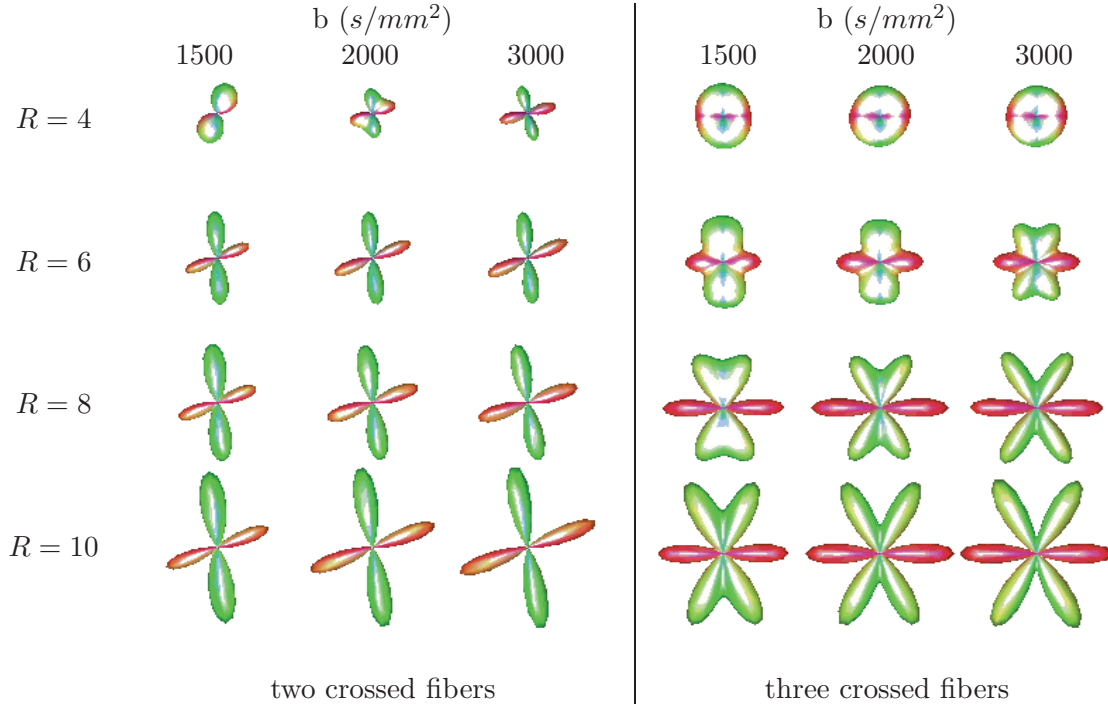

Figure 2. *An overview of contour profiles of FODs estimated by CSDP with different b-values, various orders, and multiple crossed fibers.*

Model parameters. We show the contour profiles of mean FODs estimated by CSDP in various settings in Figure 2, where 100 tests are performed with SNR = 20 in each case. For two and three fiber-crossing, we consider different diffusion-weighted b-values $b = 1500, 2000, 3000 s/mm^2$ and various orders of the homogeneous polynomial $R = 4, 6, 8, 10$. The fibers with crossed angles 80° and 60° in the left and right subfigures of Figure 2, respectively.

When we see the first row of the left subfigure that illustrate a crossing of two fibers, the estimated FODs become better as the equipped diffusion-weighted b-value increases. The second row of the right subfigure shows a similar phenomenon for a crossing with three fibers. Hence, a higher b-value is recommended when we gather DW-MRI signals. However, the signals equipped with a higher b-value always contains higher noise. A feasible approach is to employ a higher order homogeneous polynomial to approximate FOD. When the order R is greater than or equal to 8, the corresponding homogeneous polynomial could capture the profile of the crossing with two and three fibers. The best quality estimation is observed when we set $R = 10$. Hence, to deal with a complicated multifiber crossing problem, we should explore higher order homogeneous polynomials. Hence, we set $b = 3000 s/mm^2$ and $R = 10$ in the following experiments.

(a)
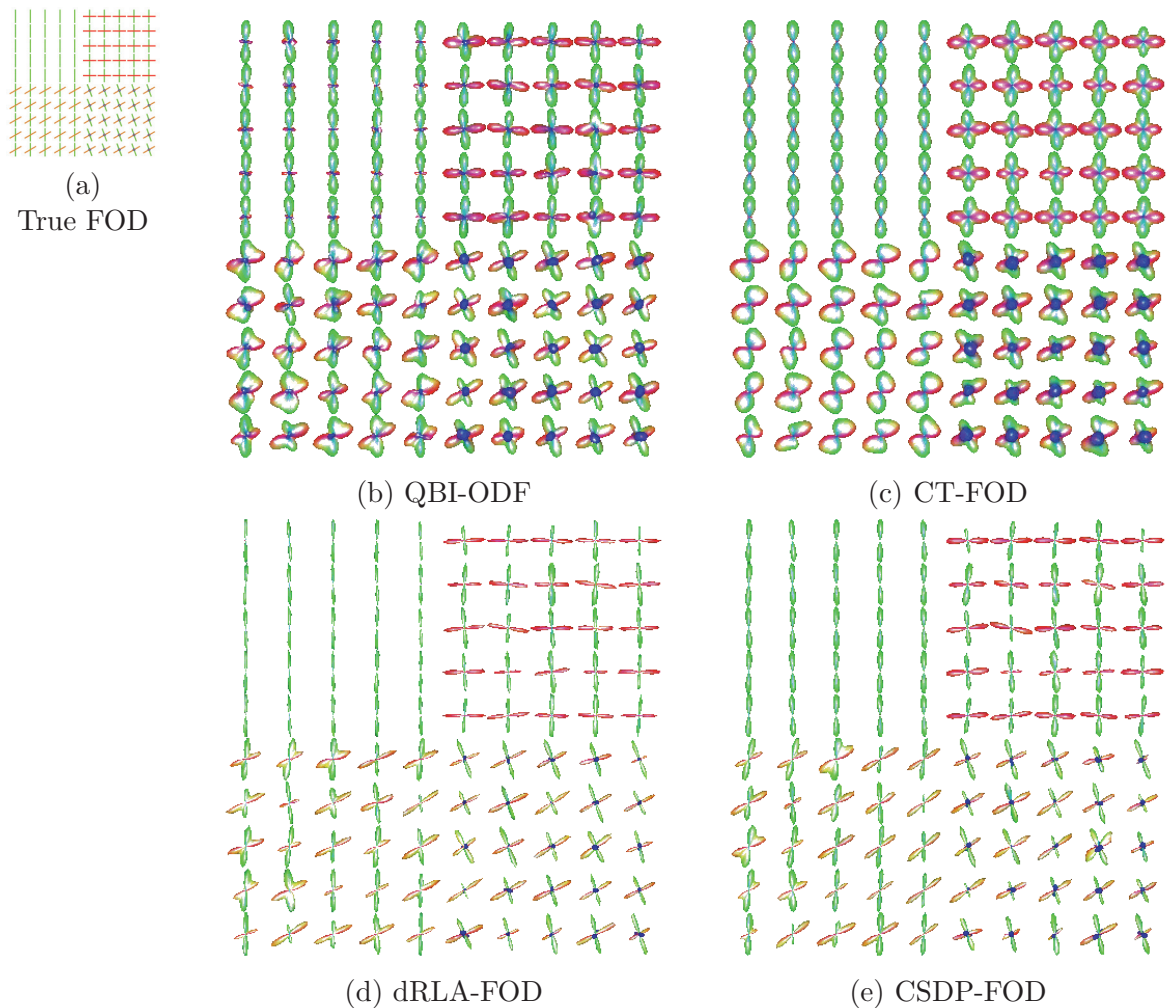True FOD

(b) QBI-ODF   (c) CT-FOD

(d) dRLA-FOD   (e) CSDP-FOD

Figure 3. *Reconstructions of a synthetic data set* (a) *where fibers in each voxel are known. We compare results obtained from four sorts of methods:* (b) *QBI-ODF from* [1, 8, 55], (c) *CT-FOD from* [59], *and* (d) *dRLA-FOD from* [17], (e) *our method CSDP-FOD.*

Synthetic data. Now, we compare the novel CSDP method with three existing approaches for synthetic data sets. As shown in Figure 3(a), the first synthetic data set contains four kinds of voxels: a single fiber in the northwest area, a horizontal fiber meeting a vertical fiber in the northeast area, two fibers crossing with 60° in the southwest area, and three fibers crossing (two of them has a crossing angle 80° and they are perpendicular to the third one) in the southeast area. In each voxel, we generate synthetic signals in 81 gradients with SNR = 20.

The counter profiles of these fibers are reconstructed by four sorts of approaches.

- QBI-ODF is produced by an improved Q-ball imaging [1, 8, 55], which generates a sharper diffusion-based orientation distribution function (ODF) than the original Q-ball imaging. The MATLAB code is downloaded from http://neuroimagen.es/webs/hardi_tools/.

- CT-FOD is estimated by a basic SD approach [59] which generates FOD in a finite set of unit vectors. Its MATLAB code is downloaded from http://cn.mathworks.com/matlabcentral/fileexchange/26997-fandtasia-toolbox.
- dRLA-FOD is produced by an improved SD approach [17] which reduces isotropic background effects and estimates FOD in a finite set of orientations. The MATLAB code is also downloaded from http://neuroimagen.es/webs/hardi_tools/.
- CSDP-FOD is estimated by the novel CSDP model established in section 2 and generates a continuous FOD function.

Figure 3(b)–(e) illustrate estimated QBI-ODF, CT-FOD, dRLA-FOD, and CSDP-FOD respectively. Since QBI does not enforce a nonnegative constraint, the resulting QBI-ODFs suffer from negative values in some orientations; see Figure 3(b). Three kinds of SD methods impose nonnegative constraints explicitly. Hence no negative values are detected. Let us examine the cases of two fibers crossing with a 60° angle in the southwest area of Figure 3(c). The estimated CT-FODs seem a little obtuse because only nonnegative constraints are considered. When the isotropic background are eliminated, dRLA-FOD looks rather sharp, as shown in Figure 3(d). However, it is a challenging problem to determine the magnitude of isotropic background. Hence, we turn to explore a unit-mass constraint in CSDP. Owing to the continuity of the estimated CSDP-FOD, it seems smoother than dRLA-FOD and still keeps slender.

Second, we consider a phantom data from the HARDI reconstruction challenge,[4] where DW-MRI signals are collected in 64 gradient orientations with the diffusion-weighted b-value $b = 3000s/mm^2$ and the signal-to-noise ratio SNR $= 20$. In CSDP-FOD, we set the order $R = 8$, since the case of $R = 10$ requires $66(> 64)$ gradient orientations. As an example, we test an interesting region marked by a yellow box in the 30th transverse slice of the phantom data, whose FA map is shown in Figure 4(a). We compare the above four sorts of methods and illustrate resulting contour profiles of FODs in the interesting region in Figure 4(b)-(e). Compared with results illustrated in Figure 3, QBI-ODF, CT-FOD, and dRLA-FOD perform better this time. This may be due to the well condition of the phantom data. The performance of CSDP-FOD seems to be in the medium place of CT-FOD and dRLA-FOD. So, to represent a sharp and continuous FOD using homogeneous polynomials, the order is an important parameter.

5.2. Human brain study. We turn to study a real-world human brain HARDI data. Each voxel is of size $1.875 \times 1.875 \times 2mm^3$, the diffusion-weighted b-value is $b = 3000s/mm^2$, and 200 gradient orientations is equipped.

We first select a coronal slice of a healthy human brain, whose fractional anisotropy (FA) map is reported in Figure 5(a). The interesting region is marked by a yellow box. We illustrate detailed contour profiles of estimated QBI-ODF, CT-FOD, dRLA-FOD, and CSDP-FOD for this area in Figure 5(b)–(e), respectively. In the east, there is an important nerve fiber bundle named the corpus callosum (CC), which connects the left and right cerebral hemispheres and facilitates interhemispheric communication. Owing to the impact of noise, there are some noise taken place in the center of each voxel of QBI-ODF and CT-FOD. dRLA-FOD is good and CSDP-FOD seems better. Next, we examine a voxel with three fibers crossing, which is

---

[4]See http://hardi.epfl.ch/static/events/2013_ISBI/.

(a) FA map

(b) QBI-ODF
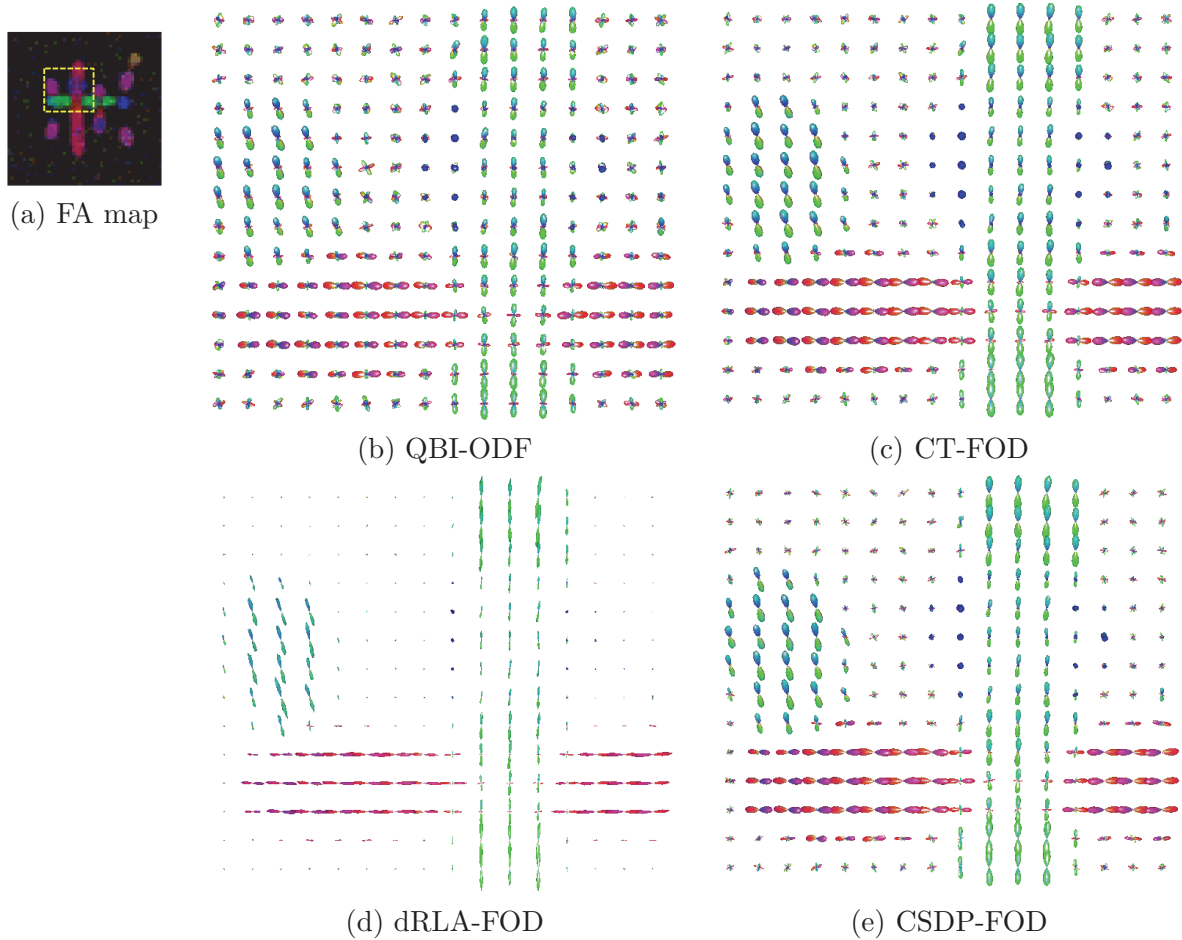
(c) CT-FOD

(d) dRLA-FOD

(e) CSDP-FOD

Figure 4. *Reconstructions of a phantom data from the HARDI reconstruction challenge. An interesting region is marked by the yellow box shown in FA map* (a). *We compare results obtained from four sorts of methods:* (b) *QBI-ODF from* [1, 8, 55], (c) *CT-FOD from* [59], *and* (d) *dRLA-FOD from* [17], (e) *our method CSDP-FOD.*

marked by a red circle. The enlarged counter of QBI-ODF, CT-FOD, dRLA-FOD, and CSDP-FOD in this voxel are illustrated in the southeast of Figure 5(b)–(e), respectively. Compared with QBI-ODF, CT-FOD, and dRLA-FOD, CSDP-FOD reports a sharper profile.

Second, we consider a transverse slice of a healthy human brain, whose FA map is shown in Figure 6(a). For the interesting region marked by the yellow box, we illustrate counter profiles of QBI-ODF, CT-FOD, dRLA-FOD, and CSDP-FOD in Figure 6(b)–(e), respectively. Obviously, we obtain similar observation. CSDP-FOD seems more clear and satisfactory. These results indicate that, compared with some existing methods for reconstructing in vivo tissue fibers, the new method CSDP is competitive.

5.3. Numerical performance of the new PRSM algorithm. To show the efficiency of the new Peaceman–Rachford splitting method (newPRSM), we compare it with its relatives: the Douglas–Rachford splitting method, which has a more famous name, the alternating direction
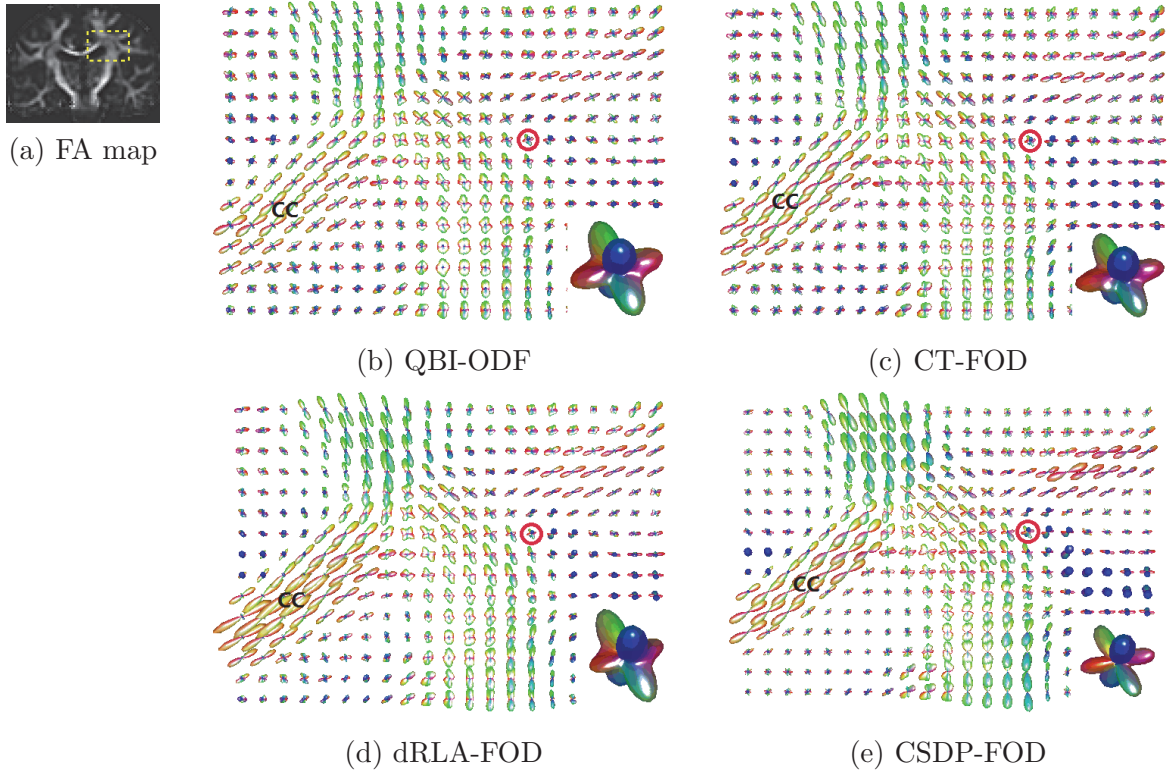
(a) FA map

(b) QBI-ODF

(c) CT-FOD

(d) dRLA-FOD

(e) CSDP-FOD

Figure 5. *The reconstruction of CC crossing corticospinal tracts for the interesting region shown in FA map (a). We compare results obtained from four sorts of methods: (b) CSA-QBI from [1, 8, 55], (c) CT-FOD from [59], (d) dRLA-FOD from [17], and (e) our method CSDP-FOD.*

Table 1
The mean iterative numbers and the mean CPU times of algorithms ADMM, scPRSM, and newPRSM for one voxel.

	ADMM	scPRSM	newPRSM
Iterative	1879.8	1502.8	983.4
numbers	100%	77%	51%
CPU time	.3573	.2976	.2123
(second)	100%	81%	62%

method with multiplier (ADMM), and the strictly contractive Peaceman–Rachford splitting method (scPRSM) [28] which was proposed recently. We remark that all of the above three algorithms have almost the same computational cost in each iteration.

For the synthetic data used in Figure 1, ADMM, scPRSM, and newPRSM start from the same initial point and we find that they return the same solution in each case. Now, we report the mean iterative numbers and the mean CPU times for each algorithms in Table 1. We say that these two modified Peaceman–Rachford splitting algorithms are faster than the Douglas-Rachford splitting method. When compared with ADMM, newPRSM saves about fifty percent iterative numbers and about forty percent CPU times, which obviously outperforms scPRSM.
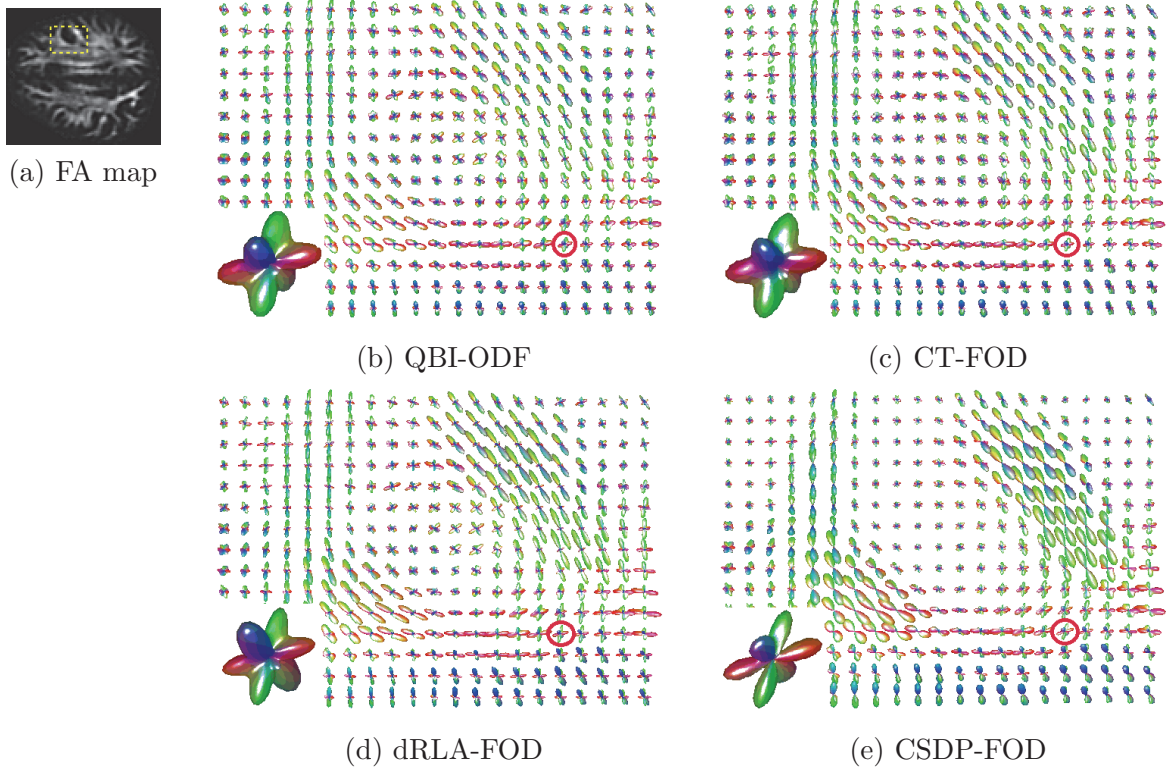
Figure 6. *The reconstruction of the centrum semiovale for the interesting region shown in FA map* (a). *We compare results obtained from four methods:* (b) *QBI-ODF from* [1, 8, 55], (c) *CT-FOD from* [59], (d) *dRLA-FOD from* [17], *and* (e) *our method CSDP-FOD.*

The reason for this efficiency is that newPRSM has a large overdetermined relaxation factor and an additional correction step.

We close this section with Figure 7 which shows the optimal measurement $\|\widetilde{\mathbf{X}}^k - \mathbf{X}^k\|_F + \|\widetilde{\mathbf{Y}}^k - \mathbf{Y}^k\|_F$ versus the iterative number of the above three algorithms applied in a typical example. Obviously, the newPRSM is the best algorithm for the statistically meaningful fiber orientation distribution estimation.

6. Conclusion. In this paper, using the methods of the SOS polynomial approximation and the convex SDP, we established a novel SD method. The estimated FOD has a clear statistical meaning. Preliminary numerical results on synthetic and real-world DW-MRI data showed that the new method is competitive since it gives clear reconstructions of nerve fibers in the human brain white matter.

For the separable convex optimization, we proposed a new Peaceman–Rachford splitting method in a prediction-correction framework. Preliminary numerical experiments showed that the new algorithm is efficient for estimating statistically meaningful fiber orientation distributions.

From our experiments we can see that although the corresponding homogeneous polynomial could capture the profile of the crossing with two and three fibers when the order R is
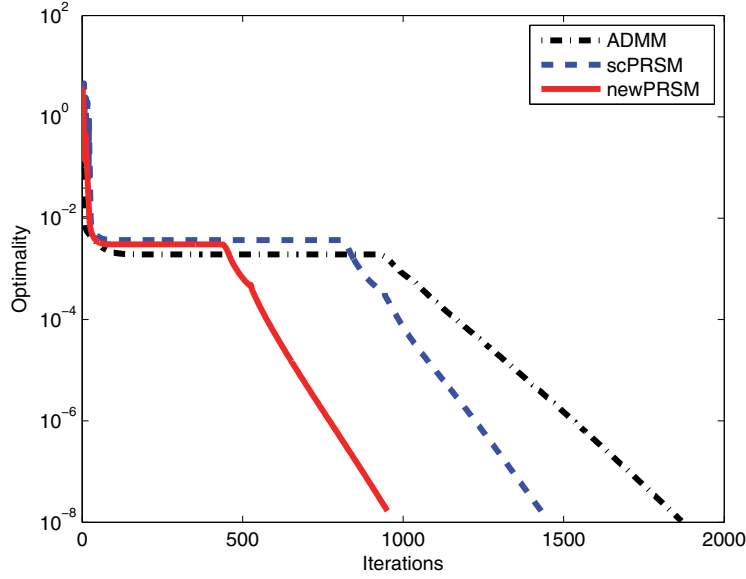
Figure 7. *A typical example illustrate the optimal measurement versus the iterative number.*

greater than or equal to 8, the larger order $R = 10$ gets better results. However, larger order means greater number of gradients used, and in applications longer imaging time. For certain patients (e.g., dementia), requiring a longer imaging time is just not possible. Hence, it is one of our future research topics to determine how to eliminate the dependence of the order of polynomial from our model and algorithm.

Appendix A. Some proofs of Section 3.1.

Proof for Theorem 3.3. We analyze optimality conditions of subproblems in the predication step.

▷   For the $\boldsymbol{x}$-subproblem, we have

$$\widetilde{\boldsymbol{x}}^k = \operatorname{argmin}\left\{ \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\lambda}^k) \; : \; \boldsymbol{x} \in \mathcal{X} \right\}$$

$$= \operatorname{argmin}\left\{ \theta_1(\boldsymbol{x}) + \frac{\beta}{2}\|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c} - \beta^{-1}\boldsymbol{\lambda}^k\|^2 \; : \; \boldsymbol{x} \in \mathcal{X} \right\}.$$

So $\widetilde{\boldsymbol{x}}^k$ satisfies the following VI:

(A.1) $(\boldsymbol{x}' - \widetilde{\boldsymbol{x}}^k)^\top \left[\nabla\theta_1(\widetilde{\boldsymbol{x}}^k) + \beta\mathbf{A}^\top(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c} - \beta^{-1}\boldsymbol{\lambda}^k) \right] \geq 0 \quad \forall\, \boldsymbol{x}' \in \mathcal{X}.$

Then, we add (3.6) and (3.8) and get

$$-(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) = \alpha\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c}) + \gamma\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c})$$

$$= (\alpha + \gamma)\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c}) + \gamma\beta\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k).$$

Hence,

$$\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c}) + (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) = -\frac{\gamma}{\alpha + \gamma}\beta\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) + \frac{\alpha + \gamma - 1}{\alpha + \gamma}(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k).$$

Substituting this equality into (A.1), we obtain

$$(\boldsymbol{x}' - \widetilde{\boldsymbol{x}}^k)^\top \left[\nabla\theta_1(\widetilde{\boldsymbol{x}}^k) - \mathbf{A}^\top \widetilde{\boldsymbol{\lambda}}^k - \frac{\gamma}{\alpha+\gamma}\beta \mathbf{A}^\top \mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) + \frac{\alpha+\gamma-1}{\alpha+\gamma}\mathbf{A}^\top(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \right] \geq 0,$$

(A.2)
$$\forall\, \boldsymbol{x}' \in \mathcal{X}.$$

▷   Similarly, for the $\boldsymbol{y}$-subproblem, we have

$$\widetilde{\boldsymbol{y}}^k = \operatorname{argmin}\{\mathcal{L}(\widetilde{\boldsymbol{x}}^k, \boldsymbol{y}, \boldsymbol{\lambda}^{k+\frac{1}{2}}) \ : \ \boldsymbol{y} \in \mathcal{Y}\}$$
$$= \operatorname{argmin}\left\{ \theta_2(\boldsymbol{y}) + \frac{\beta}{2}\|\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y} - \boldsymbol{c} - \beta^{-1}\boldsymbol{\lambda}^{k+\frac{1}{2}}\|^2 \ : \ \boldsymbol{y} \in \mathcal{Y} \right\}.$$

Hence, $\widetilde{\boldsymbol{y}}^k$ satisfies

$$(\boldsymbol{y}' - \widetilde{\boldsymbol{y}}^k)^\top \left[\nabla\theta_2(\widetilde{\boldsymbol{y}}^k) + \beta\mathbf{B}^\top(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c} - \beta^{-1}\boldsymbol{\lambda}^{k+\frac{1}{2}}) \right] \geq 0 \quad \forall\, \boldsymbol{y}' \in \mathcal{Y}.$$

Using (3.6) and (3.8) again

$$-(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) = \alpha\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{c}) + \gamma\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c})$$
$$= (\alpha+\gamma)\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c}) - \alpha\beta\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k).$$

We have

(A.3)
$$\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c}) = \frac{\alpha}{\alpha+\gamma}\beta\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) - \frac{1}{\alpha+\gamma}(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k).$$

Then, we recall (3.8) and get

$$\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c}) + (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^{k+\frac{1}{2}}) = (1-\gamma)\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c})$$
$$= \frac{(1-\gamma)\alpha}{\alpha+\gamma}\beta\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) - \frac{1-\gamma}{\alpha+\gamma}(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k).$$

We obtain the following VI when we substitute the above equality into (A.3):

$$(\boldsymbol{y}' - \widetilde{\boldsymbol{y}}^k)^\top \left[\nabla\theta_2(\widetilde{\boldsymbol{y}}^k) - \mathbf{B}^\top\widetilde{\boldsymbol{\lambda}}^k + \frac{(1-\gamma)\alpha}{\alpha+\gamma}\beta\mathbf{B}^\top\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) - \frac{1-\gamma}{\alpha+\gamma}\mathbf{B}^\top(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \right] \geq 0,$$

(A.4)
$$\forall\, \boldsymbol{y}' \in \mathcal{Y}.$$

▷   Finally, from (A.3), we have

$$(\boldsymbol{\lambda}' - \widetilde{\boldsymbol{\lambda}}^k)^\top \left[(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c}) - \frac{\alpha}{\alpha+\gamma}\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) + \frac{1}{\alpha+\gamma}\beta^{-1}(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \right] \geq 0,$$

(A.5)
$$\forall\, \boldsymbol{\lambda}' \in \mathbb{R}^\ell.$$

Combining (A.2), (A.4), (A.5), and using the compact notation, we obtain

(A.6)
$$(\boldsymbol{w}' - \widetilde{\boldsymbol{w}}^k)^\top [F(\widetilde{\boldsymbol{w}}^k) + \mathbf{M}_0(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k)] \geq 0 \qquad \forall\, \boldsymbol{w}' \in \Omega,$$

where

$$\mathbf{M}_0 := \begin{pmatrix} 0 & -\frac{\gamma}{\alpha+\gamma}\beta\mathbf{A}^\top\mathbf{B} & \frac{\alpha+\gamma-1}{\alpha+\gamma}\mathbf{A}^\top \\ 0 & \frac{(1-\gamma)\alpha}{\alpha+\gamma}\beta\mathbf{B}^\top\mathbf{B} & -\frac{1-\gamma}{\alpha+\gamma}\mathbf{B}^\top \\ 0 & -\frac{\alpha}{\alpha+\gamma}\mathbf{B} & \frac{1}{\alpha+\gamma}\beta^{-1}\mathbf{I}_\ell \end{pmatrix}.$$

According to the proof of Lemma 3.1, we have

$$\|\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)\| = 0 \quad \text{and} \quad \|\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k\| = 0$$

because of the assumption $\|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_{\mathbf{M}} = 0$. Recalling (A.6), we get $\mathbf{M}_0(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) = 0$. That is to say, $\widetilde{\boldsymbol{w}}^k$ solves VI (3.3). ∎

Proof for Theorem 3.4. Since $\boldsymbol{w}^* \in \Omega$ and $\widetilde{\boldsymbol{w}}^k$ satisfies (A.6), we have

$$(\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top[F(\widetilde{\boldsymbol{w}}^k) + \mathbf{M}_0(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k)] \geq 0.$$

It could be rewritten as

$$(\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top\mathbf{M}_0(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) \geq (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^\top F(\widetilde{\boldsymbol{w}}^k).$$

Because the map $F(\boldsymbol{w})$ is monotone and $\boldsymbol{w}^*$ is a solution of VI (3.3), we get

$$(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^\top F(\widetilde{\boldsymbol{w}}^k) \geq (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^\top F(\boldsymbol{w}^*) \geq 0.$$

Combining the above two inequalities yield that

$$(A.7) \qquad\qquad (\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top\mathbf{M}_0(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) \geq 0.$$

Next, we decompose $\mathbf{M}_0$ as $\mathbf{M}_0 = \mathbf{N}_1 + \mathbf{N}_2 + \widehat{\mathbf{M}}$, where

$$\mathbf{N}_1 := \begin{pmatrix} 0 & -\frac{\gamma}{\alpha+\gamma}\beta\mathbf{A}^\top\mathbf{B} & 0 \\ 0 & -\frac{\gamma}{\alpha+\gamma}\beta\mathbf{B}^\top\mathbf{B} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{N}_2 := \begin{pmatrix} 0 & 0 & \frac{\alpha+\gamma-1}{\alpha+\gamma}\mathbf{A}^\top \\ 0 & 0 & \frac{\alpha+\gamma-1}{\alpha+\gamma}\mathbf{B}^\top \\ 0 & 0 & 0 \end{pmatrix},$$

and

$$\widehat{\mathbf{M}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta\mathbf{B}^\top\mathbf{B} & -\frac{\alpha}{\alpha+\gamma}\mathbf{B}^\top \\ 0 & -\frac{\alpha}{\alpha+\gamma}\mathbf{B} & \frac{1}{\alpha+\gamma}\beta^{-1}\mathbf{I}_\ell \end{pmatrix}.$$

Then, from (A.7), we have

$$(A.8) \quad (\boldsymbol{w}^* - \boldsymbol{w}^k)^\top\widehat{\mathbf{M}}(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) \geq (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k)^\top\widehat{\mathbf{M}}(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) - (\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top(\mathbf{N}_1 + \mathbf{N}_2)(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k).$$

By some calculations and recalling (A.3), we have

$$-(\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top\mathbf{N}_1(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k)$$
$$= \frac{\gamma}{\alpha+\gamma}\beta(\mathbf{A}\boldsymbol{x}^* + \mathbf{B}\boldsymbol{y}^* - \mathbf{A}\widetilde{\boldsymbol{x}}^k - \mathbf{B}\widetilde{\boldsymbol{y}}^k)^\top\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)$$
$$= -\frac{\gamma}{\alpha+\gamma}\beta(\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c})^\top\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)$$
$$= -\frac{\alpha\gamma}{(\alpha+\gamma)^2}\beta(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)^\top\mathbf{B}^\top\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) + \frac{\gamma}{(\alpha+\gamma)^2}(\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k)^\top\mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)$$

and

$$
\begin{aligned}
&- (\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top \mathbf{N}_2 (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) \\
&= \frac{\alpha + \gamma - 1}{\alpha + \gamma} (\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \mathbf{A}\boldsymbol{x}^* - \mathbf{B}\boldsymbol{y}^*)^\top (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \\
&= \frac{\alpha + \gamma - 1}{\alpha + \gamma} (\mathbf{A}\widetilde{\boldsymbol{x}}^k + \mathbf{B}\widetilde{\boldsymbol{y}}^k - \boldsymbol{c})^\top (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \\
&= \frac{\alpha(\alpha + \gamma - 1)}{(\alpha + \gamma)^2} (\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)^\top \mathbf{B}^\top (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) - \frac{\alpha + \gamma - 1}{(\alpha + \gamma)^2} \beta^{-1} (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k)^\top (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&- (\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^k)^\top (\mathbf{N}_1 + \mathbf{N}_2)(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) \\
&= -\frac{\alpha\gamma}{(\alpha+\gamma)^2} \beta (\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)^\top \mathbf{B}^\top \mathbf{B}(\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k) - \frac{\alpha + \gamma - 1}{(\alpha + \gamma)^2} \beta^{-1} (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k)^\top (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \\
&\quad + \frac{\gamma + \alpha(\alpha + \gamma - 1)}{(\alpha + \gamma)^2} (\widetilde{\boldsymbol{y}}^k - \boldsymbol{y}^k)^\top \mathbf{B}^\top (\widetilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^k) \\
&= (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k)^\top \widehat{\mathbf{N}} (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k),
\end{aligned}
$$

where

$$
\widehat{\mathbf{N}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\frac{\alpha\gamma}{(\alpha+\gamma)^2}\beta \mathbf{B}^\top \mathbf{B} & \frac{\gamma+\alpha(\alpha+\gamma-1)}{2(\alpha+\gamma)^2}\mathbf{B}^\top \\ 0 & \frac{\gamma+\alpha(\alpha+\gamma-1)}{2(\alpha+\gamma)^2}\mathbf{B} & -\frac{\alpha+\gamma-1}{(\alpha+\gamma)^2}\beta^{-1}\mathbf{I}_\ell \end{pmatrix}.
$$

Therefore, the inequality (A.8) is equivalent to

$$
\text{(A.9)} \qquad (\boldsymbol{w}^* - \boldsymbol{w}^k)^\top \widehat{\mathbf{M}} (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k) \geq (\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k)^\top (\widehat{\mathbf{M}} + \widehat{\mathbf{N}})(\widetilde{\boldsymbol{w}}^k - \boldsymbol{w}^k).
$$

Recalling the definitions (3.9) and (3.10), we could rewrite the inequality (A.9) as

$$
\text{(A.10)} \qquad (\boldsymbol{z}^* - \boldsymbol{z}^k)^\top \mathbf{M} (\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k) \geq (\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k)^\top \mathbf{Q} (\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k).
$$

Since the matrix $\mathbf{Q}$ is PSD by Lemma 3.2, we immediately have $(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k)^\top \mathbf{Q}(\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k) = \|\widetilde{\boldsymbol{z}}^k - \boldsymbol{z}^k\|_\mathbf{Q}^2 \geq 0$. The proof is complete. ∎

Acknowledgments. The authors are grateful to associate editor René Vidal and the three anonymous referees for helping us improve the original manuscript.

REFERENCES

[1] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, *Reconstruction of the orientation distribution function in single- and multiple-shell Q-ball imaging within constant solid angle*, Magn. Reson. Med., 64 (2010), pp. 554–566.

[2] D. C. Alexander, *Maximum entropy spherical deconvolution for diffusion MRI*, Inf. Process. Med. Imaging, 3565 (2005), pp. 76–87.

[3] D. C. Alexander, G. J. Barker, and S. R. Arridge, *Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data*, Magn. Reson. Med., 48 (2002), pp. 331–340.

[4] A. Barmpoutis, J. Ho, and B. C. Vemuri, *Approximating symmetric positive semidefinite tensors of even order*, SIAM J. Imaging Sci., 5 (2012), pp. 434–464.

[5] P. J. Basser, J. Mattiello, and D. LeBihan, *Estimation of the effective self-diffusion tensor from the NMR spin echo*, J. Magn. Reson., 103 (1994), pp. 247–254.

[6] T. E. J. Behrens, H. J. Berg, S. Jbabdi, M. F. S. Rushworth, and M. W. Woolrich, *Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?*, Neuroimage, 34 (2007), pp. 144–155.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[8] E. J. Canales-Rodríguez, C. Lin, Y. Iturria-Medina, C. Yeh, K. Cho, and L. Melie-García, *Diffusion orientation transform revisited*, Neuroimage, 49 (2010), pp. 1326–1339.

[9] E. J. Candés, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[10] E. J. Candés, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[11] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed., Duxbury, Pacific Grove, CA, 2002.

[12] Y. Chen, Y. Dai, D. Han, and W. Sun, *Positive semidefinite generalized diffusion tensor imaging via quadratic semidefinite programming*, SIAM J. Imaging Sci., 6 (2013), pp. 1531–1552.

[13] J. Cheng, R. Deriche, T. Jiang, D. Shen, and P. T. Yap, *Non-negative spherical deconvolution (NNSD) for estimation of fiber orientation distribution function in single-/multi-shell diffusion MRI*, Neuroimage, 101 (2014), pp. 750–764.

[14] O. Ciccarelli, M. Catani, H. Johansen-Berg, C. Clark, and A. Thompson, *Diffusion-based tractography in neurological disorders: Concepts, applications, and future developments*, LANCET Neurol., 7 (2008), pp. 715–727.

[15] C. Cui, Y. Dai, and J. Nie, *All real eigenvalues of symmetric tensors*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1582–1601.

[16] F. Dell'Acqua, G. Rizzo, P. Scifo, R. A. Clarke, G. Scotti, and F. Fazio, *A model-based deconvolution approach to solve fiber crossing in diffusion-weighted MR imaging*, IEEE Trans. Biomed. Eng., 54 (2007), pp. 462–472.

[17] F. Dell'Acqua, P. Scifo, G. Rizzo, M. Catani, A. Simmons, G. Scotti, and F. Fazio, *A modified damped Richardson-Lucy algorithm to reduce isotropic background effects in spherical deconvolution*, Neuroimage, 49 (2010), pp. 1446–1458.

[18] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, *Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications*, Magn. Reson. Med., 56 (2006), pp. 395–410.

[19] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, *Regularized, fast, and robust analytical Q-ball imaging*, Magn. Reson. Med., 58 (2007), pp. 497–510.

[20] M. Descoteaux, R. Deriche, T. R. Knösche, and A. Anwander, *Deterministic and probabilistic tractography based on complex fibre orientation distributions*, IEEE Trans. Med. Imaging, 28 (2009), pp. 269–286.

[21] D. L. Donoho, *For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution*, Comm. Pure Appl. Math., 59 (2006), pp. 797–829.

[22] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[23] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrange Methods: Applications to the Solution of Boundary-Valued Problems, M. Fortin and R. Glowinski, eds., North–Holland, Amsterdam, 1983, pp. 59–82.

[24] R. Glowinski, *On alternating direction methods of multipliers: A historical perspective*, in Modeling, Simulation and Optimization for Science and Technology, William Fitzgibbon, Yuri A. Kuznetsov, Pekka Neittaanmäki, and Olivier Pironneau, eds., Comput. Methods Appl. Sci. 34, Springer, Dordrecht, Netherlands, 2014, pp. 59–82.

[25] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problémes de Dirichlet non linéaires*, Revue Française d'Automatique, Informatique, Recherche Opérationnelle, Analyse numérique, 9 (1975), pp. 41–76.

[26] A. Goh, C. Lenglet, P. M. Thompson, and R. Vidal, *Estimating orientation distribution functions with probability density constraints and spatial regularity*, in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009, Lecture Notes in Comput. Sci. 5761, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 877–885.

[27] D. Han, L. Qi, and Ed X. Wu, *Extreme diffusion values for non-Gaussian diffusions*, Optim. Methods Softw., 23 (2008), pp. 703–716.

[28] B. He, H. Liu, Z. Wang, and X. Yuan, *A strictly contractive Peaceman-Rachford splitting method for convex programming*, SIAM J. Optim., 24 (2014), pp. 1011–1040.

[29] B. He, M. Tao, and X. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim., 22 (2012), pp. 313–340.

[30] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.

[31] S. Hu, Z. Huang, H. Ni, and L Qi, *Positive definiteness of diffusion kurtosis imaging*, Inverse Probl. Imaging, 6 (2012), pp. 57–75.

[32] J. H. Jensen, J. A. Helpern, A. Ramani, H. Lu, and K. Kaczynski, *Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging*, Magn. Reson. Med., 53 (2005), pp. 1432–1440.

[33] B. Jian and B. C. Vemuri, *A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI*, IEEE Trans. Med. Imaging, 26 (2007), pp. 1464–1471.

[34] B. Jian, B. C. Vemuri, E. Özarslan, P. R. Carney, and T. H. Mareci, *A novel tensor distribution model for the diffusion-weighted MR signal*, NeuroIimage, 37 (2007), pp. 164–176.

[35] H. Johansen-Berg and M. F. S. Rushworth, *Using diffusion imaging to study human connectional anatomy*, Annu. Rev. Neurosci., 32 (2009), pp. 75–94.

[36] E. Kaden, T. R. Knösche, and A. Anwander, *Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging*, Neuroimage, 37 (2007), pp. 474–488.

[37] J. B. Lasserre, *A sum of squares approximation of nonnegative polynomials*, SIAM J. Optim., 16 (2006), pp. 751–765.

[38] V. I. Lebedev and D. N. Laikov, *A quadrature formula for the sphere of the 131st algebraic order of accuracy*, Dokl. Math., 59 (1999), pp. 477–481.

[39] C. Lin, W. I. Tseng, H. Cheng, and J. Chen, *Validation of diffusion tensor magnetic resonance axonal fiber imaging with registered manganese-enhanced optic tracts*, Neuroimage, 14 (2001), pp. 1035–1047.

[40] R. Neji, N. Azzabou, N. Paragios, and G. Fleury, *A convex semi-definite positive framework for DTI estimation and regularization*, in Advances in Visual Computing, Lecture Notes in Comput. Sci. 4841, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 220–229.

[41] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., Springer Ser. Oper. Res. Financ. Eng. Springer, New York, 2006.

[42] Y. Ouyang, Y. Chen, Y. Wu, and H. Zhou, *Total variation and wavelet regularization of orientation distribution functions in diffusion MRI*, Inverse Probl. Imaging, 7 (2013), pp. 565–583.

[43] P. A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Math. Program., 96 (2003), pp. 293–320.

[44] V. Patel, Y. Shi, P. M. Thompson, and A. W. Toga, *Mesh-based spherical deconvolution: A flexible approach to reconstruction of non-negative fiber orientation distributions*, Neuroimage, 51 (2010), pp. 1071–1081.

[45] D. W. Peaceman and H. H. Rachford, Jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Ind. Appl. Math., 3 (1955), pp. 28–41.

[46] L. Qi, *Eigenvalues of a real supersymmetric tensor*, J. Symbolic Comput., 40 (2005), pp. 1302–1324.

[47] L. Qi, D. Han, and Ed X. Wu, *Principal invariants and inherent parameters of diffusion kurtosis tensors*, J. Math. Anal. Appl., 349 (2009), pp. 165–180.

[48] L. Qi, Y. Wang, and Ed X. Wu, *D-eigenvalues of diffusion kurtosis tensors*, J. Comput. Appl. Math., 221 (2008), pp. 150–157.

[49] L. Qi, G. Yu, and Y. Xu, *Nonnegative diffusion orientation distribution function*, J. Math. Imaging Vis., 45 (2013), pp. 103–113.

[50] A. Ramirez-Manzanares, M. Rivera, B. C. Vemuri, P. Carney, and T. Mareci, *Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry*, IEEE Trans. Med.

Imaging, 26 (2007), pp. 1091–1102.

[51] E. Schwab, B. Afsari, and R. Vidal, *Estimation of non-negative ODFs using the eigenvalue distribution of spherical functions*, in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 322–330.

[52] J. D. Tournier, F. Calamante, and A. Connelly, *Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution*, Neuroimage, 35 (2007), pp. 1459–1472.

[53] J. D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly, *Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution*, Neuroimage, 23 (2004), pp. 1176–1185.

[54] A. Tristán-Vega, C. F. Westin, and S. Aja-Fernández, *Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging*, Neuroimage, 47 (2009), pp. 638–650.

[55] A. Tristán-Vega, C. F. Westin, and S. Aja-Fernández, *A new methodology for the estimation of fiber populations in the white matter of the brain with the Funk-Radon transform*, Neuroimage, 49 (2010), pp. 1301–1315.

[56] D. S. Tuch, *Q-ball imaging*, Magn. Reson. Med., 52 (2004), pp. 1358–1372.

[57] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen, *High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity*, Magn. Reson. Med., 48 (2002), pp. 577–582.

[58] V. J. Wedeen, P. Hagmann, W. I. Tseng, T. G. Reese, and R. M. Weisskoff, *Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging*, Magn. Reson. Med., 54 (2005), pp. 1377–1386.

[59] Y. T. Weldeselassie, A. Barmpoutis, and M. S. Atkins, *Symmetric positive semi-definite cartesian tensor fiber orientation distributions (CT-FOD)*, Med. Image Anal., 16 (2012), pp. 1121–1129.

[60] S. Wolfers, E. Schwab, and R. Vidal, *Nonnegative ODF estimation via optimal constraint selection*, in 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE, Piscataway, NJ, 2014, pp. 734–737.

[61] C. Ye and X. Yuan, *A descent method for structured monotone variational inequalities*, Optim. Methods Softw., 22 (2007), pp. 329–338.