Basis Pursuit based algorithm for intra-voxel recovering information in DW-MRI.

Alonso Ramírez-Manzanares and Mariano Rivera

Centro de Investigación en Matemáticas A.C.

Apdo. Postal 402, Guanajuato, Gto.

C.P. 3600, México, (52) 473 73 271 55

{alram}{mrivera}@cimat.mx

Abstract

In this work we apply the Basis Pursuit (BP) methodology for recovering the intra-voxel information in Diffusion Weighted MR Images (DW-MRI). We compare the BP approach with a Diffusion Basis Function Estimation (DBFE) algorithm. DBFE approach was previously applied to recover intra-voxel diffusion information in brain DW-MRI. The intra-voxel information is recovered at voxels that contain axon fiber crosses or bifurcations by means of a linear combination of a known diffusion functions. We state the DBFE solution in the signal decomposition context, i.e., the measured DW-MRI signal is decomposed as a linear combination of signals that belongs to a Base of Diffusion Functions (BDF). In this BDF, each signal is a M-dimensional vector in which, each component indicates a coefficient of water diffusion in a known three-dimensional orientation. In this work, we analyze and compare the solution given by DBFE method with a BP methodology. The BP methodology is used in order to select the set of base signals (which are taken from a dictionary) that best explain a given arbitrary signal. Moreover, solution strategies used in the BT and DBFE algorithm are compared and discussed. Examples in synthetic and real images are shown in order to demonstrate the performance of the compared methods.

1. Introduction

A compact signal representation have been largely used in signal processing problems [1, 2, 3] . In this context, is desirable to represent a given signal by a set of coefficients associated with elements of a dictionary of functions. The elements of such dictionary are called *atoms*. Very often the used dictionary is redundant (or over-complete), in the sense that the signal atoms are not orthogonal. In the signal processing context it is common to select atoms as Wavelets, Gabor dictionaries, Cosine Packets, Chirplets and Warplets, among others. The idea is to select the atoms that best match the signal structures, using a criteria for choosing among equivalent decompositions.

A common using criteria is to accomplish with the basic feature of sparsity, i.e., we want to recover a representation with the fewest significant coefficients. Additionally, a desirable feature of the algorithm is to spend a reasonable time in achieve the decomposition.

The mathematical model that represents the decomposition of a signal $s \in \Re^M$ as a linear combination of atoms ϕ's, each one of them belonging to a dictionary Γ is

$$s = \sum_{i \in \Gamma} \alpha_i \phi_i + \eta = \Phi\alpha + \eta, \qquad (1)$$

where Φ is a $M \times N$ matrix composed by N atoms, each atom ϕ_i is a column of Φ. The $\alpha \in \Re^N$ vector contains the linear combination coefficients. The residual η grasp the signal energy that can not be explained (or fitted) by the dictionary. When the dictionary was selected such that it is capable to represent all possible signals in the work domain, we expect that η captures the energy of the present noise in the signal.

Commonly $N >> M$, that makes the problem ill-posed: we have only a few measures $s_i, i = 1, \ldots, M$ and we want to select between a (normally) huge dictionary, the atoms that best represent the signal as a linear combination. The solution of (1) for α, given by the inversion of the pseudo-inverse of Φ is prohibitive in such case, and it does not allow us to introduce prior information about α or lead the algorithm to a solution with some desired propieties.

One important approach for solving this kind of problem is the Matching Pursuit (MP) method [3]. This approach, differently to a orthogonal expansion is a non-linear process as is explained in the following. The method initializes a residual $R^{(0)} = s$ and iterates by looking the atoms that matches better with the signal and minimizes the resid-

ual $R^{(k)}$. At iteration k, the method identifies the atom ϕ_j that best correlates with the actual residual. The correlation measure was proposed as the inner point between the normalized atoms and the actual residual, and denotes the value of the coefficient of the linear combination at the k-*th* step, i.e. $\alpha_k = <|\phi_j|, R^{(k)}>$. The new residual is then, the previous residual less the contribution of the selected atom: $R^{(k+1)} = R^{(k)} - \alpha_k \phi_j$. After m steps we obtain the following representation:

$$s = \sum_{n=0}^{m-1} <|\phi_n|, R^{(n)}>|\phi_n| + R^{(m)} \tag{2}$$

Note that in this representation, the same atom could be present several times in the reconstruction, this feature makes that the interpretation of the decomposition could be not clear in some applications where we want to decompose s in different atoms. Since the MP is an algorithm that optimizes in each step the amount of signal energy that it grasp, very often the first selected atom globally fits several signal structures, but is not best adapted to the signal local structures. The resultant is an algorithm that might choose first wrongly and spend most of its iterations correcting the differences with the original signal generated by the first few terms, see [2].

More recently, a interesting approach have been proposed, the method called Basis Pursuit (BP) [1] deal with a very similar problem formulated as:

$$\min \|\alpha\|_1$$
$$\text{subject to} \quad \Phi\alpha = s \tag{3}$$

where the constraint over the norm of the α vector converts the problem in to one with unique solution. In this way, BP has a interesting relationship to areas with ill-posed problems. The problem above, can be transformed in a linear programming one (see [1]), and solved by the simplex method. Although, in the presence of noise and depending to the chosen dictionary, the constraint in (3) could not to be accomplished given a over-constrained linear programming problem. In those cases, we need to use a more adequate minimization procedure, like a interior-point method which tries to minimize the residual vector $r_b = \Phi\alpha - s$ (see [5]). This method have shown a better performance with respect to MP, with the drawback of the requirement of a more sophisticated minimization tool.

1.1. The Diffusion Weighted Intra–voxel Problem

Amongst the most challenging goals in neuroimaging is the estimation of connectivity patterns in the brain in vivo. For this purpose, a special magnetic resonance imaging (MRI) technique named Diffusion Weighted Magnetic
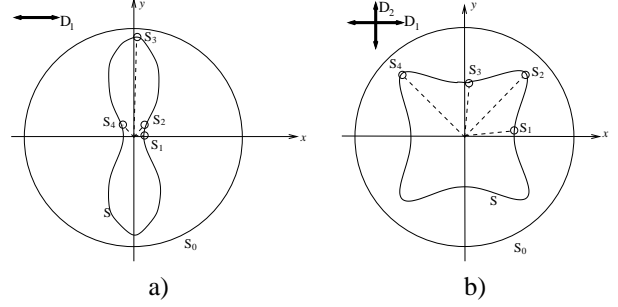

Figure 1. Schema of the measured s signals. a) The signal s (continuous peanut shape) when there is only one water diffusion orientation D_1 in a voxel. b) The shape of s (continuous concave shape) when there are 2 almost orthogonal diffusion directions D_1 and D_2. The small circles indicates the distance to the center, or the magnitude of s measured at some 2D angles.

Resonance Imaging (DW-MRI) is used, allowing one to obtain an estimation of the orientation of water diffusion in the brain. Such diffusion is constrained by the direction of axon nerve bundles. Such an information is very useful in neuroscience research, due to the relationship of brain connectivity with several diseases and, in general, with brain development [10].

The physical relationship of the water diffusion process for each image position, in a given orientation, was established by the Stejskal–Tanner [13]:

$$s_i = S_0 \exp(-b\mathbf{g}_i^T \mathbf{D}\mathbf{g}_i) + \eta_i, \tag{4}$$

where S_0 is the measured signal magnitude without diffusion gradient, s is the attenuated signal in the tissue and b is a constant directly proportional to the applied time and magnitude of the directional gradients. The unitary vector $\mathbf{g}_i = [g_x, g_y, g_z]^T$ indicates the *i–th* direction in which a directional independent magnetic gradient is applied, η_i represents a residual (produced by an inadequate model as well as for the present noise in the signal), and the diffusion coefficients in all directions are summarized by the positive definite symmetric 3×3 tensor $\mathbf{D}$. This model was used in the Diffusion Tensor MRI methodology (DT-MRI) (see [13] for more details).

A standard acquisition protocol for a single orientation $\mathbf{g}_i$, gives a 3D image, where in each voxel, the signal intensity indicates the grade of attenuation due to the diffusion, a larger attenuation (S_i small) indicates significant water diffusion in the configured orientation. A 2D schematization of the measured signals are shown in Figure 1.

In each brain position could be several diffusion orien-

2

tations, for example in the case of axon fiber crossing or bifurcations, and then, a more adequate model have been proposed by Tuch et al. [12, 11] called High Angular Resolution Diffusion Imaging (HARDI) method, based on an observation model in which the signal s is built as a finite mixture of tensors D_j:

$$s_i = S_0 \sum\nolimits_{j=1}^{\kappa} \beta_j \exp(-b g_i^T D_j g_i), \qquad (5)$$

where κ denotes the number of significant fibers orientations within the voxel, the tensor D_j explains each one of these diffusion orientations, and the scalar β_j indicates its amount of presence.

If we want to recover the unknown tensors D_j (i.e. the orientation and magnitude of the present diffusion) using only the measured signal s, all unknowns must be computed (the number κ of orientations, plus six unknowns per tensor, plus the amount β_j of each one), independently for each voxel from a large set of acquired images $\{s\}$. This Diffusion Multi–Tensor Magnetic Resonance Imaging (DMT–MRI) technique allows one to recover the intra–voxel information that is not observed in the standard DT–MRI that uses the model in equation (4) [11]. The drawbacks of the DMT–MRI method are: the large number of additional diffusion images $\{s\}$ required (for instance, 126 diffusion 3D–images are used in [11], 54 in [6]), resulting in a large acquisition time, and algorithmic problems related to Equation (5); which is highly nonlinear. So, multiple restarts of the optimization method are required to prevent the algorithm from settling in a local minima. Furthermore, such an algorithm is not stable for more than two fiber bundles, i.e. for $\kappa > 2$ (see discussion on Ref. [11], Chap. 7). There are another kind of works in this area [7, 8], that have recovered the DT–MRI intra–voxel information. Those works have been used a spatial and intra-model regularization, in order to decompose the DT in a multi-tensor one, though, those works highly depend on the existing information in the voxel neighborhood.

1.2. Diffusion Basis Function Estimation

Recently, a solution method for equation (5) based on a dictionary of functions have been proposed in [9]. In this approach the atoms are called Diffusion Basis Functions (DFB) and are defined as

$$\Phi_{ij} = \exp(-b g_i^T \bar{\mathbf{T}}_j g_i), \qquad (6)$$

where each Φ_{ij} should be understood as the decay factor due to a fixed tensor $\bar{\mathbf{T}}_j$ in the i–th direction g_i. As in the MP and BP methods the set of Φ_{ij} values define a matrix such that each vector column $\Phi_{.j}$ is an atom, generated with the signal measured in all the orientations g_i for a fixed tensor $\bar{\mathbf{T}}_j$. By using the dictionary Φ (in [9]), with cardinality N, the composition of the signal s was established as:

$$s_i = S_0 \sum\nolimits_{j=1}^{N} \alpha_j \Phi_{ij} + \eta_i \qquad (7)$$

were $\alpha = [\alpha_1, \alpha_2, ..., \alpha_j, ..., \alpha_N]^T$ is a vector, such that the scalar $\alpha_j \geq 0$ denotes the contribution of the j-th atom Φ_j to all the S_i measures in different gradient directions, i.e. the contribution to linear mixture at a particular voxel. Note that the dictionary is not complete, due to the fact that the contained orientations $j = 1, \ldots, N$ in the basis are a discretization of the 3D space, leading to a remanent scalar for each measurement denoted by η_i. By increasing the cardinality, we can diminish this remainder, so that, it becomes insignificant and therefore could be attributed to the noise in the signal.

The use of the model in (7), with a large dictionary, allow us to have a high angular resolution, only by computing the α vector in each brain position. The solution proposed in [9] was derived by the minimization of the following error function

$$E(\alpha) = \|\Phi\alpha - s\|_2^2$$
$$\textbf{s.t. } \alpha_j \geq 0 \,, \forall j \qquad (8)$$

The minimization was achieved in [9] by the derivation of (8) with respect to a each α_k, and by equalize to zero each derivative. The solution is obtained by application of a Gauss-Seidel approach with the projection of the negative α_j values to zero in each iteration. This minimization approach has the drawback that it requires several iterations to achieve the real α value.

2. Basis Pursuit Adaptation

In this paper we propose to solve the problem stated by (7), by the modification of the the minimization problem in (8) in a BP methodology.

We propose to compute the α-coefficients in (7) by solving the BP problem:

$$\begin{aligned} \min \quad & \|\alpha\|_1 \\ \text{subject to} \quad & \Phi\alpha = s \\ & \alpha_j > 0 \,, \forall j \end{aligned} \qquad (9)$$

Since we constrained already the sign of the α components, (9) can be formulated in the following linear program

$$\begin{aligned} \min \quad & \sum\nolimits_j \alpha_j = \hat{e}^T \alpha \\ \text{subject to} \quad & \Phi\alpha = s \\ & \alpha_j > 0 \,, \forall j \end{aligned} \qquad (10)$$

where $\hat{e}$ is a vector with all its components equal to one. The above linear programming carry us to solve the following problem stated in a matrix form

$$
\begin{array}{ll}
\min & \hat{e}^T \alpha \\
\text{subject to} & \left[\begin{array}{cc} \Phi & \mathbf{0} \\ I & -I \end{array} \right] \left[\begin{array}{c} \alpha \\ \gamma \end{array} \right] = \left[\begin{array}{c} s \\ \mathbf{0} \end{array} \right]
\end{array}
\tag{11}
$$

where I is a $N \times N$ identity matrix, $\gamma \in \Re^N$ is a slack variable vector (see the simplex method in [5]) and $\mathbf{0}$'s denotes a matrix or a vector with all the components equal to zero, and with dimensions corresponding to their positions in (11). This linear program, could be solved, in theory, with the simplex method, but, the signal s contains noise that converts the linear problem in a over–constrained one, that means that we expect to have the remanent η_i different from zero. For this reason we solve it with a powerful interior-point minimization method. In the experiments we used the predictor-corrector Merhotra's algorithm (see [4, 5]) that gives results in a fraction of the computational effort that requires the method described in Section 1.2.

3. Results

This section shows experimental results and a performance comparison of the DBFE and the BP methods.

3.1. Synthetic Experiments

In order to test the performance of the proposed method for recovering the diffusion signals ϕ_j that best composed the observed s, we generate several synthetic examples. In the first example we generated a signal s which is exactly composed by a linear combination of two atom signals, $s = \frac{1}{2}\phi_i + \frac{1}{2}\phi_j$. Then we used a different number of measures M and we compare the norm of the error betwen the obtained solution and the known real one. The dictionary was composed by $N = 36$ atoms and we used $b = 1000$ in (6). The comparison between the method DBFE reported in [9] and the one that uses BP proposed in Section 2 is shown in Figure 2. Note that the error diminishes in both methods by increasing the number of measures s_i. However note that the error for the method BP is smaller and requires less measures (only 4) for obtaining the exact result.

In fact, Figure 3 give us insights about the reason of the large error of the DBFE method. In this experiment we generated a synthetic signal s that is composed by 2 diffusion orientations in a 2D plane, 63° and 157° respectively. The dictionary is composed by diffusion signals that splits the 2D orientation in intervals of 10°, i.e., the set of orientation was $[0^o, 10^o, 20^o, \dots, 170^o]$. This, dictionary can not exactly represents the signal s. The result obtained with the
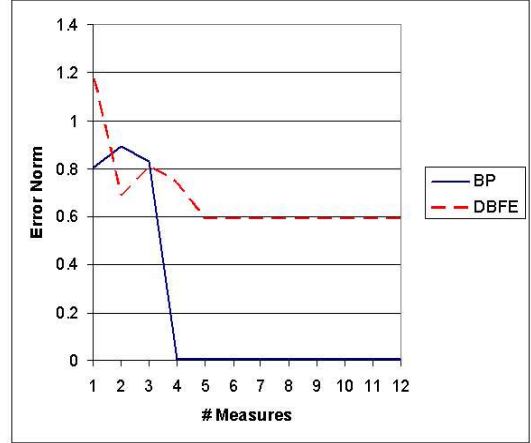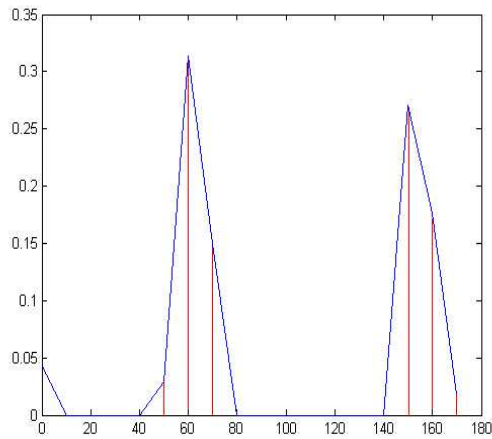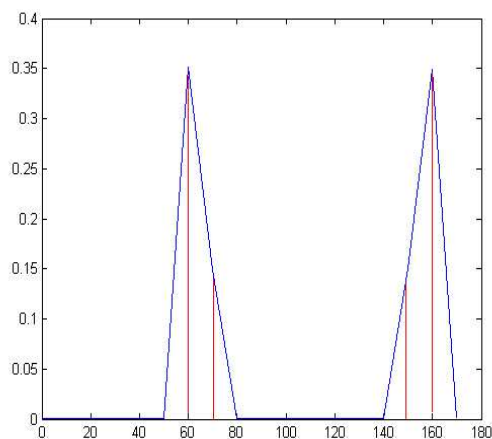

Figure 2. Comparison for both methods, of the error generated by the variation of the number of measures M, see text.

DBFE algorithm is shown in panel 3a, and the result for the BP method is shown in panel 3b. In both figures the x axis corresponds to the orientation of the $N = 18$ atoms (or diffusion signals in the dictionary) and the y axis indicates the value of the associated α_j value. As can be seen, the result obtained by the DBFE involves more coefficients than the needed, and the result given by the BP method uses only the necessary number of coefficients, i.e., in the case of BP, the signal ϕ_7 associated with 60° have the biggest α coefficient, and join with the coefficient associated to 70° (a small one) fits good enough the diffusion measured in the 63° orientation. In the same way, the second diffusion is fitted only by the combination of the atom associated to 160° and 150° in this order, and all the others coefficients are zero. The BP method required 12 iterations of the Mehrotra minimization method. On the other hand, the method DBFE was iterated 1300 times, and the obtained result contains more coefficients that the needed for representing s. Note that the solution given by DBFE is almost right, i.e., indicates with the biggest α coefficient the diffusion function that best explains the signal, altough, the solution given by BP is sharper, thus, better defined, indicating better the magnitude of the coefficients of the linear combination, and for the medical purposes its solution is better.

Additionally, we tested the robustness of both methods to the noise in the measured signal. Figure 4 shows the obtained results. We generate a synthetic signal and then we added independent Gaussian noise with zero mean and different standard deviations. The x axis indicates the SNR ratio and the y axis shows the norm of the difference between the obtained solution vector α and the real one. As can be seen, the error of the method BP is the smallest one.

a)

b)

Figure 3. α **coefficients obtained in the case where the diffusion orientation was not in the dictionary. a) For the DBFE method, b) for the BP method, see text.**
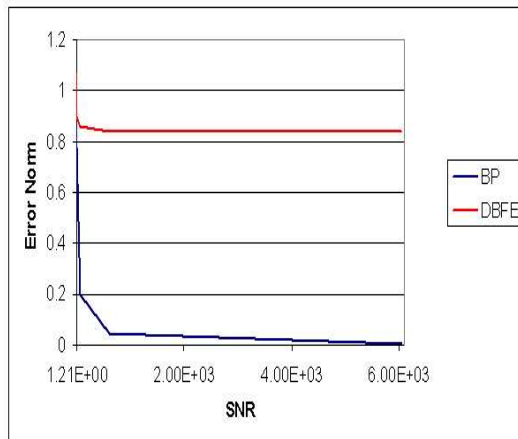

Figure 4. Variation for both methods of the error in the results by the SNR variation in the synthetic signal, see text.

3.2. Experiments with real DW-MRI data

The results obtained with brain DW–MRI are shown in figure 5. In this case, we used the DW images that are included in the BioTensor software, provided by the web site of the Scientific Computing and Imaging Institute of the University of Utah (`http://software.sci.utah.edu/archive/archive_main.html`). Which are constructed to $2.0 \times 2.0 \times 2.0 mm^3$, and have volume dimensions of $75 \times 109 \times 29$. The DW data have $M = 12$ measurements per voxel in independent directions g_i, $b = 1000$. The dictionary of diffusion functions was composed by $N = 33$ equidistant orientations. Panel 5a depicts the brain map (a sagital slice) with the region of interest marked with a white square. Panel 5b shows the classical orientations given by the Diffusion Tensor (see [13] for details) that give us a reference for the orientations. We show in panel 5c the computed orientations, weighted by its associated α coefficient, recovered by the DBFE method. Finally, the computed orientations, weighted by its associated α coefficient, with the BP method are shown in Panel 5d. It is important to note that this results are congruent with the a priori anatomical knowledge for this region which corresponds to the brain *corona radiata*. Note that in the center of the slice there is a fiber cross (depicted in the DT–MRI representation in panel 5b with no predominant orientation tensors,i.e. balls) and the orientations recovered shown in panel 5c and 5d indicates the crossing by means of almost orthogonal diffusion orientations. Note that, in some regions, the recovered orientations by the BP method are more defined and contrasted, i.e. there are some brain positions where only one diffusion direction was detected instead of two, like in

the DBFE result.

We measured the computational time for both methods, they were implemented in the same language and were executed in the same computer. BP approach is about 40 times faster than the DBFE approach.

4. Conclusions

In this paper, we stated the problem of intra-voxel diffusion estimation in DW–MRI in the context of signal decomposition, such decomposition is composed by atoms that belongs to a dictionary. The BP methodology was appplied, demonstrating that the obtained results are better compared with the DBFE methodology, that represents the state of the art in this approach. The obtained results are better in the sense that the uncertainty in the diffusion orientation was diminished and the required computational effort was lesser. Such a differences can be explained, in part, because the BP-like approach uses a objective function based in the L-1 norm (that is known to be a robust potential) and the DBFE uses a non-robust L-2 based cost norm.

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[2] R. Gribonval, Ph. Depalle, X. Rodet, E. Bacry, and S. Mallat. Sound signal decomposition using a high resolution matching pursuit. In *ICMC'96*.

[3] S. M. and Zhifeng Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions in Signal Processing*, 41:3397–3415, 1993.

[4] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal of Optimization*, 2:575–601, 1992.

[5] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operation research, 2000.

[6] G.J.M. Parker and D.C. Alexander. Probabilistic monte carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In *IPMI'03*.

[7] A. Ramírez-Manzanares and M. Rivera. Brain nerve bundles estimation by restoring and filtering intra–voxel information in diffusion tensor mri. In *Procc. 2 IEEE Workshop on VLSM'03, Oct., Nice, Fr.*, 2003.

[8] A. Ramirez-Manzanares and M. Rivera. Basis tensor decomposition for restoring intra–voxel structure and
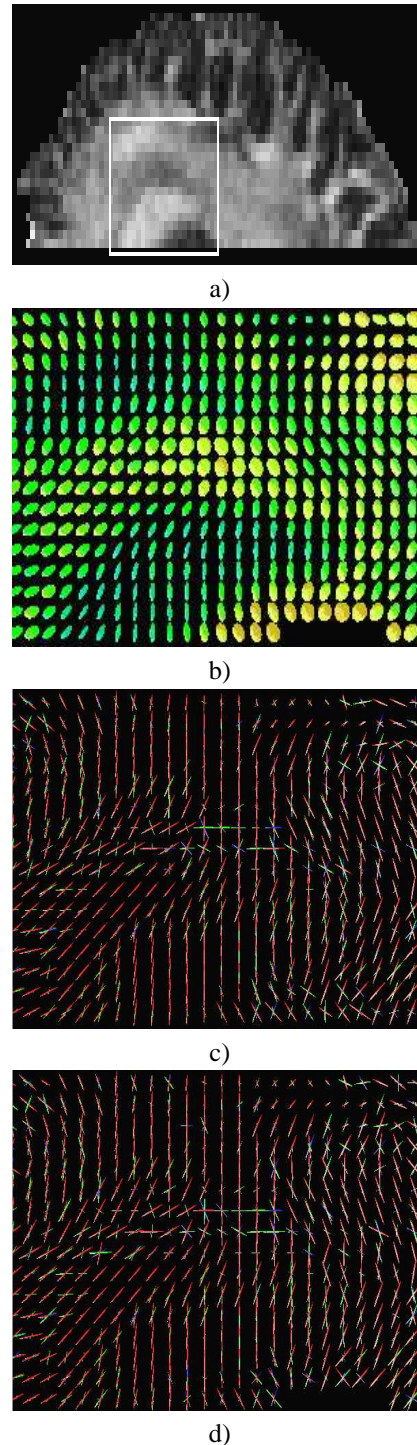
a)

b)

c)

d)

Figure 5. Results obtained in real DW–MR images, by the application of the BP method. a) Brain map of the region of interest in a sagital slice, showing a part of the *corona radiata***. b) DT-MRI shown as reference for the orientations c) Diffusion orientations weighted by the α coefficients obtained by the DFBE method. d) Diffusion orientations weighted by the α coefficients obtained by the BP method. See text.**

stochastic walks for inferring brain connectivity dt–mri. *to appear in IJCV*, 2005.

[9] A. Ramírez-Manzanares, M. Rivera, B. C. Vemuri, and T. C. Mareci. Basis functions for estimating intra–voxel structure in dw–mri. In *Medical Imaging Conference, 2004*, 2004.

[10] B. Richard. *Introduction to Functional Magnetic Resonance Imaging Principles and Techniques*. Cambridge: Cambridge University Press, 2002.

[11] D.S. Tuch. *Diffusion MRI ofcomplex tissue structure*. PhD thesis, Harvard-MIT, Cambridge MA, 2002.

[12] D.S. Tuch, R.M. Weisskoff, J.W.Belliveau, and V.J. Wedeen. High angular resolution diffusion imaging of the human brain. In *7th Annual Meeting of ISMRM, Philadelphia*.

[13] C.F. Westin, S.E. Maier, H.Mamata, F.A. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor mri. *Medical Image Analysis*, 6(2):93–108, 2002.