Comparison of Sampling Strategies and Sparsifying Transforms to Improve Compressed Sensing Diffusion Spectrum Imaging

Michael Paquette,† Sylvain Merlet,‡ Guillaume Gilbert,§ Rachid Deriche,∥ and Maxime Descoteaux⊥∗

INTRODUCTION

Diffusion Magnetic Resonance Imaging (dMRI) is a recent technique introduced by (1). Following the introduction of the pulsed gradient spin-echo sequence (2), the quantification of the water diffusion was possible by estimating the displacement of particles from the phase change that occurs during the acquisition process. Under the narrow pulse assumption (2) (gradient pulses are sufficiently short), the relationship between the diffusion signal attenuation, \( E(q) \), in q-space, and the diffusion propagator \( P(R) \), in real space, is given by a Fourier transform (FT) relationship (3) such that

\[
E(q) = \int_{\mathbb{R}^3} P(R) \exp(-2\pi i q \cdot R) dR,
\]

where \( E(q) = S(q)/S_0 \), where \( S(q) \) is the diffusion signal measured at position \( q \) in q-space, and \( S_0 \) is the baseline image acquired without any diffusion sensitization (\( q = 0 \)). We denote \( q = |q| \) and \( q = qu, R = hr \), where \( u \) and \( r \) are three-dimensional (3D) unit vectors. The wave vector \( q \) is \( q = \gamma G / 2\pi \), with \( \gamma \) the nuclear gyromagnetic ratio and \( G = gu \) the applied diffusion gradient vector. The norm of the wave vector, \( q \), is related to the diffusion weighting factor (the b-value), \( b = 4\pi^2 q^2 \tau \), where \( \gamma = \Delta - \delta/3 \) is the effective diffusion time with \( \delta \) the time of the applied diffusion sensitizing gradients and \( \Delta \) the time between the two pulses. We can measure the approximation of the average diffusion propagator by taking the ensemble average over the imaging voxel, hence the name Ensemble Average Propagator (EAP). The EAP is the full 3D displacement probability function of water molecules, which faithfully characterizes the water diffusion phenomenon. Note that the Fourier relationship between the EAP and the diffusion signal of Eq. 1 is strictly valid only if the narrow pulse condition is met, which is rarely the case for in vivo 3D q-space MRI. However, the violation of this condition induces a convolution over a range of diffusion times in the measurements, preserving the large-scale structure and orientation of the inferred propagator (4).

In Diffusion Spectrum Imaging (DSI) (5), we obtain the EAP \( P(R) \) by directly taking the inverse FT of \( E(q) \). However, DSI requires the acquisition of many diffusion weighted images sensitized to different \( q \) orientations and magnitude in the sampling space, to obtain a high-resolution EAP. In brief, while this technique has the advantage of giving a good approximation of the water diffusion phenomenon, it is limited by the long acquisition time due to the large number of required samples. Recent improvements have been shown to decrease the scan time by at least a factor of three using fast

†Department of Computer Science, Sherbrooke Connectivity Imaging Laboratory, Université de Sherbrooke, Sherbrooke, Quebec, Canada.
‡Athena Project-Team, INRIA Sophia Antipolis-Méditerranée, Sophia-Antipolis Cedex, France.
§Department of Computer Science, Sherbrooke Connectivity Imaging Laboratory, Université de Sherbrooke, Sherbrooke, Quebec, Canada.
∥MR Clinical Science, Philips Healthcare, Cleveland, Ohio, USA.
⊥⁎Correspondence to: Maxime Descoteaux, Ph.D., Department of Computer Science, Science Faculty, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, Quebec, Canada J1K2R1. E-mail: m.descoteaux@usherbrooke.ca

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

© 2014 Wiley Periodicals, Inc.
acquisition sequences and parallel imaging (6). Further improvements can be achieved using signal processing techniques in order to decrease the acquisition time. This is where the Compressive Sensing (CS) technique can be useful to accelerate DSI.

CS is a recent technique, which aims to accurately reconstruct sparse signals from undersampled measurements acquired below the Shannon–Nyquist rate.

Many contributions regarding the application of CS to dMRI have emerged in the last years (7–18). These techniques are all different and do not aim to recover the same features of the diffusion phenomenon. First, Landman et al. (13) applied the CS-based recovery in diffusion tensor imaging (19,20) to reconstruct features such as the fractional anisotropy and mean diffusivity from clinical diffusion tensor imaging-like acquisitions. The dMRI signal is characterized in (13) as a sparse combination of diffusion tensors and obtains a diffusion tensor-based propagator. Similarly, (7) defines a parametric representation of the intravoxel fiber geometry using a discrete mixture of Gaussians. Second, other techniques such as (9,11) have proposed CS-based techniques for estimation of the angular part of the EAP, the orientation distribution function (ODF) (5,21,22). In (9,11), the dMRI signal acquired at a unique b-value and this single-shell dMRI signal is modeled as a sparse and linear combination of spherical ridgelets from an overcomplete and redundant dictionary. They also provide analytical formulae to estimate the ODF. Third, continuous formulations of the CS problem have emerged in order to analytically estimate the 3D dMRI signals (18). It allows one to interpolate and extrapolate the observed data, to analytically estimate the EAP and several other diffusion features such as the return-to-origin probability and the mean squared displacement. In (18), compressed sensing in conjunction with parametric continuous function is used in the signal model. Lastly, dictionary learning has recently become popular in dMRI. In particular, several techniques have been introduced in the context of DSI (12,14,23). For instance, (12,14,23) learn dictionaries from DSI-like acquisitions and use it to either denoise full DSI data or to perform undersampled DSI acquisitions and reconstructions. In particular, Gramfort et al. (14) exploit the symmetry and positivity of the signal to assess free parameters of the dictionary learning problem. Reference (23) applies a $l_2$ reconstruction using pseudoinverse and Tikhonov regularization with respect to a learned dictionary. Other dictionary learning approaches are parametric and provide continuous representations of the atom, such as (17), where the dictionary is formed by a weighted combination of third-order B-splines. The work of (17) appears promising in reconstructing the diffusion signals, and further enhancement could be done regarding the development of analytical formulae to estimate other diffusion features. More recently, Merlet et al. proposed in (16) to learn a dictionary where each atom is constrained to be a parametric function and propose a computational framework to analytically recover the EAP and the ODF.

The context of this article is the acceleration of DSI acquisitions via the CS technique (CS-DSI) using existing fast sparsifying transforms. This problem has already been addressed by several teams (6,10,12,15) and consists in reconstructing a discrete EAP, which is sparse with respect to a prespecified sparse transform that is not based on dictionary learning. For instance, Menzel et al. (10) solve the CS-DSI problem by considering the gradient operation as sparse transform, also known as total variation (TV) regularization. Bilgic et al. (12) combine a TV regularization and a sparse constraint using the Haar-based discrete wavelet transform (DWT). In (8), Merlet and Deriche take the identity operator as sparse transform. However, these works were conducted independently and there has not been an analysis and comparison of the differences between them, which makes it difficult to choose an optimal sparse transform and decide which evaluation metrics to be used. Moreover, the sampling protocol is also rarely studied and often overlooked. We argue that it is crucial to jointly optimize sampling scheme (SC) and sparsifying transform to improve CS-DSI results.

We, therefore, propose an extensive experimental comparison of three sampling strategies and six sparsifying transforms to accurately reconstruct the EAP and its many features, and favorably compare it to two other existing CS SCs. At first, we present and then compare the Haar DWT, the Cohen–Daubechies–Feauveau (CDF) 9/7 DWT, the stationary wavelet transform (SWT), the dual-tree wavelet transform (DTWT), the gradient of the EAP (also known as TV regularization), and the identity (I), i.e., the canonical basis. Then, we propose a new uniform angular and radial SC. Finally, the full 3D EAP is reconstructed and diffusion features such as the ODF and its discrete number of compartments (DNC) and angular error (AE) are obtained. We also compute metrics that capture both the radial and angular part of the EAP such as the kurtosis (24), the normalized mean squared error (NMSE), and Pearson’s correlation coefficient (PCcoeff) computed from the full EAP. These metrics are confronted and compared for all SCs and sparsifying transforms throughout the synthetic and real data experiments.

Hence, the contributions of this work are 3-fold. (i) We propose a new SC that is uniform in angular coverage but random in radial sampling to accurately reconstruct the EAP and its many features, when compared to two other existing CS SCs. (ii) We show that the DWT with CDF 9/7 wavelets is quantitatively better on all metrics than the other five sparse transforms tested for CS-DSI. (iii) Extensive quantitative results from experiments on synthetic and real human brain data demonstrate that an undersampling factor of approximately four is sufficient to recover robust EAP, kurtosis, and ODFs.

**METHODS**

**Compressed Sensing**

The EAP $P(R)$ and the normalized diffusion signal $E(q)$ are related by a 3D FT (see Eq. 1). Using this relation, we seek a robust compressed sensing (25,26) reconstruction based on a well-chosen SC and sparse representation to recover $P(R)$ from an undersampled number of measurements.

Mathematically speaking, we consider the problem of recovering a vector $x \in \mathbb{R}^D$ from an observation vector,
\[ y = Ax + z, \quad \text{where } z \in R^n \text{ is the acquisition noise, } A \in R^{m \times n} \text{ is the, so called, sensing matrix. For a particular application, the sensing matrix } A \text{ can be decomposed as a product of a sparse system } \Psi \text{ and an orthogonal measurement system } \Phi, \text{ i.e., } A = \Phi \Psi. \] In this case, the signal of interest, \( f = \Psi x \), and \( x \) is a vector of coefficients, \( x_i = (f, \psi_i) \), with \( \psi_i \) a column of \( \Psi \).

For example, in this work, \( \Phi \) is characterized by a FT, \( y \) is the diffusion signal, and \( f \) is the EAP. The aim of CS is to infer the coefficient vector \( x \) (and the underlying signal of interest \( f \)) from an observation vector \( y \) of size \( m \ll n \). This is done by solving the following convex problem for \( x \):

\[
\min_{x \in \mathbb{R}^n} \| y - Ax \|_{\ell_1} + \lambda \| x \|_{\ell_1}, \tag{2}
\]

where \( \| y - Ax \|_{\ell_1} \) is the data consistency constraint, \( \| x \|_{\ell_1} \) is the sparsity constraint, \( \ell_1 \) and \( \ell_2 \), respectively, indicate the \( \ell_1 \) and \( \ell_2 \) vector norm, and \( \lambda \) is a constant that controls the degree of sparsity of \( x \). Note that \( \lambda \) depends on many factors such as the signal sparsity and the level of noise. In this work, we use a cross validation procedure (27) to find \( \lambda \) and we solve Eq. 2 with a FISTA algorithm (28).

One important ingredient of a CS recovery is the sparsity (25,26). We need that the signal of interest \( f \) admits a sparse representation \( x \) with respect to the sparse system \( \Psi \), i.e., \( x \) contains a small number of nonzero coefficients. Considering \( x \) is a sparse vector, we can impose a sparsity constraint on \( x \) by minimizing its \( \ell_1 \) norm as done in the problem presented in Eq. 2 (25,26). The two following sections present two CS frameworks. Respectively, the case where the sparse system \( \Psi \) is characterized by an orthonormal and the case where \( \Psi \) is characterized by an overcomplete and redundant dictionary.

### CS with Respect to Orthonormal Bases

When the sparse system is characterized by an orthonormal basis, (25) describes two other important conditions to obtain accurate and robust CS recovery.

First, (25,29) imposes a certain amount of incoherence between the measurement system \( \Phi \) and the sparse system \( \Psi \). It can be interpreted as a rough characterization of the degree of similarity between \( \Phi \) and \( \Psi \). The higher is this incoherence, the fewer measurements are potentially needed. This finding holds when the sparse system \( \Psi \) is orthonormal.

Second, another key notion is the restricted isometry property (RIP). Reference (30) established some results about the accuracy of the reconstruction of a sparse signal \( x \) as long as the sensing matrix \( A \) obeys the RIP.

Reference (25) proved that the RIP can also hold for sensing matrices, where \( \Psi \) is an arbitrary orthonormal basis and \( \Phi \) is an \( m \times n \) measurements matrix drawn randomly from a suitable distribution. One can refer to (31) for examples of distributions for which this property has been proven to hold true. In dMRI, we have the possibility to provide the imaging system with a list of gradient directions and \( b \)-values in a gradient table, which allows one to acquire q-space samples in a random fashion. However, for an arbitrary chosen distribution, it is often computationally unfeasible to compute guarantees that the RIP holds. Even for a uniform random sampling, we cannot ensure that the RIP is respected. As an example, in (32), the authors observe that although the entries of the sensing matrix are chosen independently from a Gaussian distribution [which is a strong theoretical result ensuring that the sensing matrix respect the RIP (25)], the RIP does not hold in practice. Despite the fact that the RIP may not hold in practice, good results can nevertheless be obtained.

### CS with Respect to Redundant and Overcomplete Dictionaries

A recent work (33) was published to generalize the CS theory described in (25,26,29,30). Reference (33) shows the possibility to accurately recover \( f \) which is sparse in a redundant and overcomplete dictionary, when the sensing matrix \( A \) satisfies the RIP adapted to this dictionary. Moreover, (33) suggests that the incoherence property is not necessary but, instead, proves that the solution of Eq. 2 is very accurate provided that \( x \) is sparse enough.

The CS theory is well established considering discrete signals (25,26,29,30,33) and is adapted to the reconstruction of a discrete EAP. In this work, we consider two important parameters to obtain accurate and robust CS reconstructions: the system \( \Psi \), which has to be sparse and incoherent with the sensing system, and the sampling protocol itself. In the following section, we review six potential choices for \( \Psi \). The choice of the SC is discussed in the results section.

### Sparse Representation of the Discrete EAP

In this section, we present six candidates of EAP representation to be used in the CS context. Three of them are part of the orthonormal basis framework, i.e., the canonical basis and the two DWTs based on the Haar and CDF 9/7 wavelets. Two others are overcomplete and redundant representations, namely the discrete SWT and the real dual-tree discrete wavelet transform (DTWT). Then, we propose to study the gradient operation as a sparse transform, commonly known as a TV constraint when combined with the \( \ell_1 \) regularization.

**The Canonical Basis**

Because of the quasi-Gaussian nature of the EAP attenuation, Merlet and Deriche (8) assumed that the EAP is sparse in its original domain and do not use any sparse transform, i.e., \( \Psi = I \), where \( I \) represent the canonical basis. This assumption is rarely true but Merlet and Deriche (8) have nonetheless found that solving Eq. 2 in this setting leads to good qualitative EAP reconstructions in practice. This can be explained by the fact that the pair of canonical and Fourier bases provide a maximally incoherent system (29). Then, the \( \ell_1 \) term in Eq. 2 plays a role of denoiser. Hence, in this work, we quantitatively compare its performance against better sparsifying bases.

**The Discrete Wavelet Transform**

The DWT is an extremely well established tool in the image processing community and is used, for instance,
in image compression because it provides highly sparse representation of natural images.

Moreover, the system built from the pair of wavelet and Fourier bases has already been proved efficient in CS applications (34). Several public and open-source libraries such as WaveLab (http://www-stat.stanford.edu/~wavelet/Wavelab_850/index_wavelab850.html) in MATLAB and PyWavelets (http://www.pybytes.com/pjwavelets/) in Python propose efficient DWT implementations. Based on the work of (15), we consider a DWT based on the biorthogonal and symmetric CDF wavelet. In particular, we use the CDF 9/7, which has four vanishing moments. As a comparison, we also consider the Haar-based DWT as also successfully used by Bilgic et al. (12) to solve the CS-DSI problem.

The Discrete SWT

The classical DWT is not a shift-invariant transform, which is often a useful property to avoid visual artifacts around the discontinuities of a signal (35). The SWT overcomes this limitation at the expense of providing more coefficients than the size of the signal itself, leading to an overcomplete and redundant transform. The mother wavelet used for the SWT is the CDF 9/7 wavelet.

The Real 3D Dual-Tree Discrete Wavelet Transform

The DTWT transform has the benefit of being oriented, which can be useful to sparsely represent 3D signals with directional properties such as the EAP. Here, the DTWT is implemented using four separable 3D DWTs in parallel. Then, the subbands of the four DWTs are combined appropriately (36). However, because the DTWT is four times more expensive compared to the DWT, it comes under the framework of redundant and overcomplete transforms.

The Gradient Transform

Setting \( \Psi \) as a gradient transform comes to replace the sparsity constraint by a TV regularization (37). Strictly speaking, this method is known as a TV reconstruction. The TV reconstruction involves nonlinear optimization and the use of the conjugate gradient algorithm to find the corresponding solution. Menzel et al. (10) were the first to use the TV regularization as a diffusion-domain constraint in the CS-DSI problem. Note that the TV regularization was also applied in the spatial domain in combination with a sparse constraint in the diffusion domain by (9) for ODF reconstruction in high angular resolution diffusion imaging (HARDI).

Evaluation Metrics and Features Characterizing the Diffusion Phenomenon

There is no consensus in the existing CS-DSI literature on what diffusion features and what metrics to be reported for evaluation of EAP reconstruction. In Menzel et al. (10), radial and axial kurtosis measures as well as voxelwise EAP correlations are reported between full DSI and their CS-DSI technique. Qualitative ODF reconstructions are also shown. Conversely, in Bilgic et al. (12, 23), root mean squared normalized EAP error are reported, qualitative ODFs are shown and the impact of CS-DSI is computed from averaged fractional anisotropy from 18 white matter bundles produced using ODF tractography. Finally, in Gramfort et al. (14), metrics from AE and DMC error from the ODF as well as NMSE on the EAP are quantified to evaluate the effect of undersampling the DSI acquisition.

Hence, to cover existing evaluation techniques, we validate the EAP reconstruction by computing the NMSE and the Pearson’s correlation coefficient, which are computed from the discrete EAP from full DSI and the one estimated via CS. The NMSE between a signal \( x \) and its estimation \( \hat{x} \) is given by

\[
\text{NMSE} = \frac{||x - \hat{x}||_2^2}{||x||_2^2}
\]

and the Pearson correlation coefficient (PCcoeff) is given by

\[
\text{PCcoeff} = \frac{1}{n - 1} \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (z_i - \bar{z})^2}},
\]

where \( n \) is the number of discrete points in the propagator, \( Z_k \) is the vector of standard score of each sample in \( x \), and \( \text{std} \) is the standard deviation.

In addition to the full EAP, we can compute the ODF, which gives the probability that a water molecule diffuses in a given orientation.

We express the ODF, which only captures angular information of the diffusion process \( Y(r) \), as the integration of the EAP \( P(\mathbf{r} \mathbf{r}) \) over a solid angle (5), i.e.,

\[
Y(r) = \int_0^\pi P(\mathbf{r} \mathbf{r}) R^2 d\theta \approx \int_\alpha^\beta P(\mathbf{r} \mathbf{r}) R^2 d\theta,
\]

where \( r \) is a 3D unit vector and \( (\alpha, \beta) \) are the lower and upper bounds that can be set for the numerical ODF approximation of the true theoretical ODF. From each estimated ODF, we assess the quality of the angular information contained in the EAP by computing the difference in the number of fiber compartments (DNC) and the AE with respect to the known ground truth. To compute the DNC and the AE, we extract the maxima on a discrete grid of 4000 points of the estimated ODFs and compare them to the ground truth maxima. Then, the DNC becomes the mean difference between the number of maxima extracted on the estimated ODFs and the true number of maxima, and the AE is computed between the maxima extracted on the estimated ODFs and the respective maxima within the ground truth.

Finally, another computed diffusion feature is the kurtosis, which accounts for both the angular and radial information of the diffusion phenomenon. The kurtosis aims to quantify the non-Gaussianity of a probability density function. The diffusion kurtosis \( K \) in the direction \( n \) is defined by

\[
K(n) = \frac{\langle (R \cdot n)^4 \rangle}{\langle (R \cdot n)^2 \rangle^2} - 3,
\]

with \( R \) is a 3D unit vector, \( R = r \mathbf{r} \) and \( \langle (R \cdot n)^m \rangle = \int P(R) (R \cdot n)^m d^3R \) the \( m \)th-order moment of \( P(\mathbf{r}) \) about its
mean value. In our experiments, we use the method described in (24) to estimate \( K(n) \), we consider only the \( b \)-values below 3000 s/mm\(^2\) for the kurtosis computation. This upper bound for the \( b \)-value ensures the convergence of the Taylor expansion that is derived form the dMRI signal to find the kurtosis (24,38).

Synthetic Data Simulation

We evaluate the reconstruction on synthetic data using the dataset provided by the ISBI 2012 HARDI contest (http://hardi.epfl.ch/). The contest was organized to provide a way for different groups to fairly compare their methods against a common set of ground-truth data. In this contest, the normalized diffusion signal \( E(q) \) is generated from the multitensor model for \( F \) fibers, \( E(q) = \sum_{f=1}^{F} p_f \exp(-4\pi^2\tau q^2 u^T T_f u) \), where a fiber \( f \) is defined by a tensor matrix \( T_f \) and weight \( p_f \), such that \( \sum p_f = 1 \). \( q \) denotes the norm of the effective gradient and \( u \) is a unitary vector in Cartesian coordinate. From this synthetic model, ground-truth diffusion features can be derived to estimate the ODF (5,21,22) and the kurtosis (24). Note that, for the synthetic experiments, we consider the dMRI signals contained in the file Training IV, with two fibers crossing at equally represented angles from 30 to 90° and different volume fractions, yielding 610 synthetic signals. Some experiments are performed in a noisy scenario where Rician noise is added. For these experiments, the noisy signal is computed as

\[
E_{\text{noisy}} = \sqrt{(E + \epsilon_1)^2 + \epsilon_2^2},
\]

where \( \epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma) \) with \( \sigma = 1/\text{SNR} \).

Real Data Acquisition

A standard DSI acquisition mimicking the original DSI protocol (5) was done on a 3 T system (Philips Achieva X, Best, The Netherlands), equipped with a whole body gradient (40 mT/m and 200 T/m/s) and a 8-channel head coil. Single-shot spin-echo echo-planar imaging measurements with isotropic 2-mm spatial resolution and 515 DW measurements were acquired comprising q-space points of a cubic lattice within the sphere of five lattice units in radius. TE/TR = 116 ms/14.9 s (including time for dynamic B\(_0\) stabilization), bandwidth in echo-planar imaging direction = 1101 Hz, 128 × 128 matrix, 60 axial slices with a parallel imaging (SENSE) factor of 2, delta and Delta were 45.4 and 57.7 ms and maximal \( b \)-value of 6000 s/mm\(^2\). We compute the signal-to-noise ratio (SNR) of the data as done in (39) and find a value of 38 at a \( b \)-value \( b = 0 \) s/mm\(^2\) and 6.5 at \( b = 6000 \) s/mm\(^2\). The SNR remains higher than 4–5, which is known to be the limit under which the noise profile becomes Rician.

RESULTS

In this section, we first present and study the efficiency and the robustness of the sampling protocol on synthetic data. Then, we validate the reconstruction using various metrics while considering both synthetic and real data. Before going further, we introduce the notation for the CS reconstruction using the six sparse representations presented in the Methods section:

- A CS reconstruction without applying any sparse transform on the EAP, i.e., \( \Psi = I \) (CS-I).
- A CS reconstruction while applying a DWT on the EAP (CS-DWT). In particular, we use the term CS-DWTHaar when the DWT is based on the Haar wavelet and CS-DWTCD9/7 when the DWT is based on the CDF 9/7.
- A CS reconstruction while applying a SWT on the EAP (CS-SWT).
- A CS reconstruction while applying a DTWT on the EAP (CS-DTWT).
- A TV reconstruction (CS-TV).

Choice of SCs

As seen in the Methods section, random sampling should, in theory, allow one to accurately recover the EAP from noisy measurements. Moreover, (10) shows the importance of the sampling pattern to reconstruct oriented structures. In particular, (10) observes that a SC performs well with samples randomly distributed according to a Gaussian distribution with its mean being at the center of the grid. This can be explained by the fact that most of the signal energy is focused around its center. However, when considering 3D signals of size 11 × 11 × 11 (common size in DSI), such random under SCs do not always ensure a uniform angular distribution of the samples. This is especially important when dealing with directional signals such as dMRI signals, where an angular feature such as an accurate ODF is critical.

For this purpose, we propose a novel SC that assures uniform angular sampling and random radial q-space sampling. We generate \( N \) samples uniformly distributed on the sphere using the static repulsion algorithm (40). Then, for each sample, we draw a random radius between the origin and the maximum radius of a sphere comprised in our acquisition grid. Finally, we match each of those samples in our Cartesian grid in a way to obtain \( N \) samples in the Cartesian DSI grid. This SC both ensures a uniform angular covering of the q-space and has a random radial sampling.

Figure 1 shows a two-dimensional (2D) illustration of our novel SC strategy. In this section, we compare three SCs and show the importance of uniform angular covering of the q-space:

- A uniform random SC (RU-SC), i.e., the samples are randomly distributed according to a uniform distribution, as prescribed by SC theory.
- A Gaussian random SC (RG-SC), i.e., the samples are randomly distributed according to a Gaussian distribution with its mean corresponding to the center of the grid (as used in (10)).
- A random SC ensuring a uniform angular covering of the q-space (HA-SC) as described earlier.

To have an overview of these three SCs, we generated one instance of each SC in a 3D grid of size 11 × 11 × 11 for \( N = 64 \) measurements and summed all the 2D slices along the x-axis, which provides sampling densities of the RU-SC, RG-SC, and HA-SC. The first row of Figure 2 shows these sampling densities. We also projected the
FIG. 1. SC construction in 2D. (left) $N$ samples uniformly distributed on the sphere using the static repulsion algorithm (40). (middle) Considering one sample, we draw at a random distance, $d$, from the origin and the maximum radius of a sphere comprised in the DSI grid. (right) This is repeated until all the samples are visited. Note that, in the 3D case, we generate the SC on one hemisphere and then symmetrize it. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

FIG. 2. Sampling densities corresponding to a uniform random SC (RU-SC), a Gaussian random SC (RG-SC), a random SC ensuring a uniform angular covering of the q-space (HA-SC). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
Sampling Strategies and Sparsifying Transforms for CS-DSI

over, the values of vary much for a given sparse transform and SNR. More-

radial integration of the ODF are presented in Table 1.

optimal values of sen bounds for the radial integration lead to better

values of tion, amplifying these errors even more). These optimal

(respective to the range for the

factor in the ODF computa-

were used for future ODF estimation.

Choice of the Sparse Representation

In this section, we compare the six approaches presented

at the beginning of the section (CS-I, CS-DWTHaar, CS-

DWTcdf9/7, CS-SWT, CS-DTWT, CS-TV) on the recon-

struction of synthetic data provided by the ISBI 2012 HARDI contest. We also compare the CS approaches to low-pass DSI approximation. These low-pass DSI are obtained by taking only the central portion of the full DSI grid up to a certain radius. An example of low-pass DSI is presented in (41), called DSI102, where samples from the full DSI (DSI257) are taken up to a radius of 3.61 instead of 5. This gives a maximal b-value of approximately 4500 s/mm². We add Rician noise with SNR = 20 and SNR = 10 to the data. From the reconstructed EAPs, we estimate both the ODFs and kurtosis (Eq. 4) for optimal (α, β) bounds mentioned in Table 1. Hundred simulations are performed for each scenario (SNR/sparse transform/number of samples), for a total of 6000 simulations. Regarding the CS methods,

Table 1

Optimal Lower and Upper (α and β) Bounds Found for Best DNC and AE in the ODF Integration of Eq. 3

<table>
<thead>
<tr>
<th>SNR</th>
<th>CS-I</th>
<th>CS-DWTHaar</th>
<th>CS-DWTcdf9/7</th>
<th>CS-SWT</th>
<th>CS-DTWT</th>
<th>TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 samples</td>
<td>0.2/0.6</td>
<td>0.2/0.6</td>
<td>0.2/0.7</td>
<td>0.3/0.8</td>
<td>0.3/0.8</td>
<td>0.2/0.8</td>
</tr>
<tr>
<td>64 samples</td>
<td>0.2/0.6</td>
<td>0.2/0.6</td>
<td>0.2/0.7</td>
<td>0.3/0.8</td>
<td>0.3/0.8</td>
<td>0.3/0.8</td>
</tr>
<tr>
<td>96 samples</td>
<td>0.2/0.6</td>
<td>0.2/0.6</td>
<td>0.3/0.7</td>
<td>0.3/0.7</td>
<td>0.3/0.6</td>
<td>0.2/0.6</td>
</tr>
<tr>
<td>128 samples</td>
<td>0.2/0.6</td>
<td>0.3/0.6</td>
<td>0.2/0.7</td>
<td>0.3/0.7</td>
<td>0.3/0.6</td>
<td>0.2/0.6</td>
</tr>
<tr>
<td>160 samples</td>
<td>0.2/0.6</td>
<td>0.3/0.6</td>
<td>0.3/0.7</td>
<td>0.3/0.7</td>
<td>0.3/0.6</td>
<td>0.2/0.6</td>
</tr>
<tr>
<td>192 samples</td>
<td>0.2/0.6</td>
<td>0.3/0.6</td>
<td>0.3/0.7</td>
<td>0.3/0.7</td>
<td>0.3/0.6</td>
<td>0.2/0.6</td>
</tr>
</tbody>
</table>

These bounds are computed for a number of samples and a CS reconstruction methods. The signals are contaminated by Rician noise at SNR = 10 and SNR = 20. For DSI, α = 0.2 and β = 0.7 was optimal for SNR = 10 and α = 0.3 and β = 0.8 for SNR = 20.

Moreover, we compute the mean values of the DNC and AE over the 100 experiments and repeat the process with a number of samples N = 32, 64, 96, 128, 160. The results, shown in Figures 3 and 4, give an overview of the angular information obtained.

Figures 3 and 4 show that the HA-SC leads to better accuracy than the two other schemes in terms of DNC and AE. These results show that a higher degree of angular information is contained in the HA-SC. Figures 3 and 4 also confirm the finding of (10) regarding that RG-SC is a better choice than RU-SC. Hence, even if the CS theory is usually based on uniform random sampling, these experiments and previous work by (10) show that random sampling protocols are more appropriate when we control the distribution of points. In particular, the HA-SC appears efficient and robust to recover directional information, which is an important aspect in ODF fiber tractography applications. For the rest of this work, the focus will be on HA-SC but we will nonetheless compare all three SCs on real data for the other EAP metrics.

We first observe that the bounds in Table 1 do not vary much for a given sparse transform and SNR. Moreover, the values of and are lower at SNR = 10, as the signal is likely to be more noisy when going away from the origin and, thus, provide erroneous information (especially considering the $r^2$ factor in the ODF computation, amplifying these errors even more). These optimal values of and will be used for future ODF estimation.
we generate a new HA-SC for each repetition. The mean DNC, AE, and kurtosis NMSE are shown in Figure 5. Note that the kurtosis computed for the full DSI and DSI102 are equal, as we consider only the $b$-values below 3000 s/mm$^2$ for the kurtosis computation, as mentioned before.
In Figure 5, we first see that the CS-DWT-CDF9/7 method always provides more accurate angular information (DNC, AE) and kurtosis than the CS-DWT-Haar method. This difference, even if small, is due to the use of an appropriate wavelet basis, i.e., the CDF 9/7 instead of the Haar wavelet basis. Indeed, Haar wavelet is a step...

FIG. 4. Mean of the AE (right) and the difference in the number of fiber compartments (DNC) (left) when using a uniform random SC (RU-SC), a Gaussian random SC (RG-SC), a random SC ensuring a uniform angular covering of the q-space (HA-SC). We compare the CS-SWT, CS-I, CS-DTWT reconstruction. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
function designed to capture piece-wise signals, whereas the CDF 9/7 wavelet adequately describes the quasi-Gaussian attenuation of the EAP (15).

Although the Haar wavelet is less appropriate in sparsely representing the EAP (15), we nonetheless note that the CS-DWTHaar method gives acceptable

---

**FIG. 5.** Evolution of the DNC, AE, and kurtosis NMSE in function of the number of samples for the six reconstruction methods, i.e., CS-I, CS-DWT with CDF 9/7 and Haar, CS-SWT, CS-DTWT, TV. We also show the results regarding the full DSI reconstruction, i.e., when 257 samples are used, and the DSI102, i.e., when 102 samples are used according to (41) Rician noise is added with SNR = 10 (top) and SNR = 20 (bottom). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
results in terms of angular information (DNC, AE) and kurtosis.

Figure 5 also shows that the CS-I method gives satisfactory results even if no sparse transform is applied on the EAP. This can be explained by the high incoherence between the real and Fourier space. Note that in our work, the CS-I method is not better than the CS-DWT methods, whereas the author in (12) observed that the application of a DWT led to worse results compared to the case where no transform is applied.

We also compare the CS-DWT and the CS-TV methods used in (10,12). In Figure 5, we globally observe that the CS-DWTCDF9/7 method performs better than the CS-TV method in terms of angular information and kurtosis, especially at SNR = 10, where the CS-TV method is not able to provide accurate directional information. Indeed, the AE is underestimated in voxels where the number of detected compartments is higher than the true number of compartments. This implies that we cannot rely on the AE as the number of maxima is not well approximated (high DNC values). Hence, the CS-TV method is sensitive to noise and should only be used on high SNR datasets.

Next, we note in Figure 5 that the two overcomplete and redundant wavelet transforms do not improve the results of the DWT, even if these two transforms benefit from a shift-invariance property (the SWT) or an orientation property (the DTWT). We can thus observe that orthonormality of the transforms improves reconstructions in this CS-DSI application.

Figure 5 also compares the CS-DWT method against the full DSI reconstruction and the DSI102 reconstruction presented in (41). Overall, it is not surprising to see that the DNC and the AE obtained with the CS-DWT method come closer to the respective values obtained with the full DSI when the number of measurement increases. In particular, at SNR = 20, from N = 160 and above, which is 97 less measurements than the full DSI acquisition, the CS-DWT method leads to a better approximation in terms of AE and number of compartments and kurtosis NMSE, compared to the full DSI method. As for the DSI102, we observe that the CS-DWT method performs better with respect to the same number of measurements (i.e., 102 measurements) both at a SNR equal to 10 and 20.

We thus conclude from these synthetic data experiments that the DWT based on the CDF 9/7 mother wavelet remains the most appropriate choice to reconstruct a discrete EAP using the CS technique amongst the various transforms investigated. In terms of angular information (AE and DNC) and kurtosis NMSE, the CS-DWT globally leads to lower values than the other methods considering both SNR = 20 and SNR = 10. The CS-DWT method remains robust to noise (even for a low SNR) and is efficient. Finally, we note that for most curves, the elbow of the curve occurs at approximately N = 64 measurements. This observation is true for all sparse transforms and measures, except the overcomplete techniques (SWT and DTWT), where the elbow of the curve occurs closer to 100 measurements. For the rest of this article, we therefore perform the CS reconstruction on real data using the DWTCDF9/7 as main sparse transform. DWTHaar and TV are also included in many of the next experiments to strengthen the conclusion of synthetic experiment that DWTCDF9/7 is most appropriate.

Real Data Experiments

We estimate the EAPs from the human brain data described in the Real Data Acquisition section using the DSI method with the full set of measurements (i.e., N = 257 measurements), the low-pass DSI with N = 129, 62, 29 (as described in the Choice of the Sparse Representation section) and the CS-DWTCDF9/7 method with N = 128, 64, 32 undersampled measurements. Note that an additional DW image without any diffusion sensitization is also acquired and used by all these method at the center of q-space. Then, for the ODF estimation (Fig. 6), we choose the integration parameters as α = 0.4 and β = 0.8, inspired by Table 1. Note that β could be higher due to the low level of noise.

The ODFs computed from the CS reconstructed EAPs with N = 128 and N = 64 measurements describe well the underlying fiber structure shown by the ODFs estimated with the full DSI technique (i.e., with N = 257 measurements). The DSI low-pass perform similarly to their CS-DSI counterpart in term of AE but the DNC values are significantly better for CS-DSI with N = 64 and N = 32 with DNC = 0.152, 0.183, 0.183, 0.274 for CS-DSI64, CS-DSI32, DSi62, and DSi29, respectively. With CS-DSI32, only single fiber structures and a small number of the large crossings are correctly estimated via the ODFs (Fig. 6). We observe that the ODF estimated from low-pass DSI in Figure 6 are smooth and that the CS-DWT method provides sharper ODF when we consider an equivalent number of samples. This observation is reflected by the DNC and AE values.

In addition, the whole EAP is recovered whereas single-shell HARDI techniques only estimate the ODF, which is a small part of the information provided by the EAP.

Next, to quantitatively compare sparsifying transforms, we also confront the CS-DWTCDF9/7, CS-DWTHaar, and CS-TV methods on the EAPs estimated with the classical DSI and the low-pass DSI on the real human brain data. The CS-DWT- and CS-TV-based EAPs are computed for a large number of undersampled measurements of N = 32, 48, 64, 80, 96, 112, 128, 144, 160, and the low-pass DSI are also computed for a large number of low-pass filtering N = 29, 41, 47, 62, 74, 86, 90, 102, 126, 129, 153, 171. Figure 7 shows the values resulting of the NMSE computation between these DSI approximation and the full DSI (N = 257).

Based on this figure, we see that both of the CS-DWT method and the CS-TV method obtain better EAP in terms of NMSE with the fully sampled DSI compared to their low-pass equivalent (similar values of N). The CS-DWT curves show a shift in the slope at approximately N = 64. This observation is coherent with what was found in the synthetic experiments. As with the other metrics shown, the CS-DWT methods outperform CS-TV and CS-DWTCDF9/7 obtains lower EAP NMSE than CS-DWTHaar, which again confirm the synthetic experiments.

Figure 8 shows the NMSE between the kurtosis of the full DSI EAPs and the CS-DWTCDF9/7, CS-DWTHaar,
and CS-TV based EAPs, again with $N = 32, 48, 64, 80, 96, 112, 128, 144, 160$. The NMSE kurtosis values exhibit a similar behavior as observed on the synthetic data experiments in Figures 3 and 4.

Figure 9 shows the Pearson correlation coefficient between EAPs reconstructed with CS-DWTHaar, CS-DWTcdf9/7, and CS-TV at $N = 64$ with RU-SC, RG-SC, and HA-SC. All the reconstructions using HA-SC perform globally better than their RG-SC and RU-SC counterpart, most likely due to the better angular coverage of HA-SC. The reconstruction quality of the wavelet-based scheme is appreciably better than the TV reconstruction. Moreover, CS-DWTcdf9/7 outperforms CS-DWTHaar and shows more uniform correlation maps. The values are bounded at a minimum of 0.8 to enhance visibility. Note also that NMSE were produced as similarly done in Bilgic et al. (12) using the root mean squared error. The NMSE remains small and stable with undersampling, going from around 0.22 for $N = 64$ to 0.16 for $N = 128$ for CS-DWTcdf9/7 using HA-SC. Pearson’s correlation coefficient maps were also produced as similarly done in Menzel et al. Menzel et al. (10) showed best-case correlation around 97% whereas, here, we show that the optimal HA-SC sampling combined with CDF9-7 transform leads to more than 96% over all voxels of the white matter reliably over many different reconstruction experiments. Hence, the results shown in Figure 9 further strengthen our findings on synthetic data that

FIG. 6. ODFs examples on DNC map estimated from a human brain data, using the full DSI method, the low-pass DSI method with $N = 129, 62, 29$ and the CS-DWTcdf9/7 method with a number of samples $N = 128, 64, 32$ using HA-SC. We compute the pair (DNC,AE) between the reconstructions and the full DSI estimation of the ODFs: CS-DSI128 (DNC = 0.116, AE = 5.951), CS-DSI64 (DNC = 0.152, AE = 7.972), CS-DSI32 (DNC = 0.183, AE = 13.295), DSI129 (DNC = 0.104, AE = 4.682), DSI62 (DNC = 0.183, AE = 8.632), DSI29 (DNC = 0.274, AE = 16.133). The reader is encouraged to zoom-in to appreciate the ODF glyphs. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
DWTCDF9/7 is a well suited sparse transform for DSI reconstruction and that using HA-SC increases the reconstructed EAPs quality.

Finally, Figure 10 shows the Pearson correlation coefficient between EAPs reconstructed with CS-DWTCDF9/7 with HA-SC and DSI257 on an axial slice of a human brain for \( N = 32, 64, 128 \). The EAPs reconstructed at \( N = 128 \) demonstrate a very high correlation with the EAPs of DSI257 and the EAPs in the white matter portions of the brain with \( N = 64 \) still retain a high correlation. At \( N = 32 \), we see larger degradation of the EAPs quality but the correlation coefficient remains high in the large white matter zones. The values are bounded at a minimum of 0.85 to enhance visibility. These results again strengthen the observation that \( N = 64 \) is a good trade-off between undersampling ratio and reconstruction quality as observed in our synthetic experiments and in Figure 9.

**DISCUSSION**

As recently demonstrated by several recent articles (10,12,15), CS offers an efficient way to accelerate the DSI acquisition. These works were conducted independently and rarely explored all potential steps that can be optimized in the CS reconstruction, from SC, to sparse transforms and evaluation on metrics capturing both the angular and radial part of the EAP.

In this work, we first proposed a novel SC that assures uniform angular and random radial q-space samples. Our desire to find this new SC was motivated by Menzel et al. (10), where it was showed that a SC performs well with samples randomly distributed according to a Gaussian distribution with its mean being at the center of the grid. From this observation, we decided to investigate the choice of a SC that is better tailored to reconstruct both the angular and radial features of the EAP. Also in this article, six EAP sparse representations were considered in the CS reconstruction, i.e., the canonical basis (CS-I method), the DWT (CS-DWT method), the discrete SWT based on the CDF 9/7 wavelet (CS-DWTCDF9/7 method) and on the Haar wavelet (CS-DWTHaar method), the real dual-tree DWT (CS-DTWT method), the gradient transform (CS-TV method). Furthermore, we studied and compared three sampling protocols, i.e., a uniform random SC (RU-SC), a Gaussian random SC (RG-SC), and a random SC, which ensures a uniform angular covering of the q-space (HA-SC). We also compared the CS approaches with the full DSI reconstruction with 257 samples, and with several low-pass undersampling of the half sphere from DSI. All the EAP representations and sampling protocols were studied and compared through extensive synthetic and human brain data experiments.

We also compared the kurtosis and EAPs of CS-DWTCDF9/7, CS-DWTHaar, CS-TV on real data and obtain results consistent with our findings on synthetic experiment. Furthermore, the EAPs were compared to low-pass DSI and the CS-DSI technique performed better in terms of NMSE. We showed the differences between the transforms and the SC on a slice of human brain data using the Pearson Correlation coefficient.

From the results, one transform performs best: the CDF 9/7-based DWT. The CS-DWTCDF9/7 method leads to accurate estimation of the EAP, kurtosis, and derived features from the ODF. We observed that the CDF 9/7 was a more appropriate wavelet to use in a CS reconstruction than the Haar wavelet used in (12). It was also
found more to be more accurate and more robust to noise than the TV-based technique of (10). We also found a robust and efficient sampling protocol exploiting the fact that the angular coverage of the q-space should be uniform. From these studies, we finally found that the combination of the CDF9/7-based DWT with the HA-SC was the best way to solve the CS-DSI problem with approximately 64 measurements, which represents an acceleration factor of approximately 4. This is seen in both synthetic and real dataset and agrees with acceleration factors reported before by recent CS works (8,10,12,14).

We have seen that the orthonormal property of the DWT was found to be more important than the particular and attractive property of the SWT (shift invariance) and the DTWT (directional property). Future works could be focused on finding an orthonormal and fast transform that also has properties such as shift-invariance or directionality. As far as we know, there are no such transforms in the literature. Thus, at the moment, the CDF 9-7 based DWT remains the most appropriate choice of sparse transform for the CS-DSI problem. Moreover, the CS-DWT/CDF9/7 method leads to more accurate angular information and kurtosis than the low-pass DSI reconstructions considering the same number of measurements.

Several points arising from this work could be investigated further. First, in this work, we use predefined

FIG. 9. Pearson correlation coefficient between EAPs reconstructed with CS-DWTHaar, CS-DWT/CDF9/7, and CS-TV at N = 64 with RU-SC, RG-SC, and HA-SC. It is clear that the reconstruction using the HA-SC is more accurate than their RU-SC and RG-SC counterpart. The DWT methods outperform the TV method and DWT/CDF9/7 shows very high and uniform correlation map. The values are bounded at a minimum of 0.8 to enhance visibility. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
sparse transforms to sparsely fit the EAP. Recent articles have focused on learning sparse dictionaries by adapting their contents to the raw diffusion data, as for example, in (12,14,16). Learning dictionaries instead of using pre-defined sparse transforms is very popular at the moment and these works have recently reported better acceleration factors than 4 for CS-DSI. However, dictionary learning techniques have limitations such as being slow techniques and computationally heavy techniques as well as being quite hard to constrain and optimize properly. Moreover, it is not clear if they can be generalized and useful for pathological brains and thus, useful in clinical settings. A thorough comparison between these learned dictionaries and the DWT based on the CDF 9/7 wavelet is worth to be done in future.

CONCLUSIONS
We have described an improved technique to accurately perform CS DSI recovery of the EAP from undersampled q-space measurements. The best reconstruction is based on CDF 9/7 discrete wavelet sparsifying transform and a new uniform angular and random radial q-space SCs. Extensive experiments on synthetic and human brain data demonstrate that 64 measurements (acceleration factor of 4) preserve the angular features of the ODF (DNC = 0.152, AE = 7.972), the kurtosis (NMSE = 0.064), and full 3D EAP (NMSE = 0.22) with reasonable accuracy.

REFERENCES
1. Le Bihan D, Breton E. Imagier de diffusion in vivo par resonance magnetique nucleaire. CR AcadSci 1985;301:1109–1112.


