Master research Internship

Bibliographic report

Compressed sensing in diffusion magnetic resonance imaging for the reconstruction of microstructure parameters of the cerebral white matter

Domain: Medical Imaging

Author:
Raphaël Truffet

Supervisor:
Emmanuel Caruyer
Univ Rennes, Inria, CNRS, IRISA
VisAGeS research team

Abstract: Magnetic resonance imaging (MRI) is a successful technique for the observation of the inside of the body. One of its modality, diffusion MRI, provides techniques to capture information about the movemement of water molecules. Thus, it can give information at a microscopic scale, while the typical resolution of clinical MRI is close to 1 millimeter. The acquisition sequences in diffusion MRI rely on time-dependent magnetic field gradients. Every gradient waveform provides one diffusion-weighted measurement, and the repetition of different measurements allows for the reconstruction of microstructure parameters and diffusion features. The problem is that, for the moment, the number of acquisitions needed to correctly estimate the microstructure parameters may be too high for in vivo imaging. Several families of gradient waveforms have been proposed, in order to optimize the precision of the reconstruction and to decrease the number of acquisitions. During the internship, we explored generalized gradient waveforms. In this report, we propose a method to predict the signal for a given set of gradient waveforms, using only a subset of the measurements. This method relies on compressed sensing, with the use of a dictionary that we learned on data generated with Monte-Carlo simulations. We then compare two different heuristics to select the measures to use for the prediction; we found that limiting the redundancy of the measures allows us to reduce the number of measurements, with minimum loss of precision of the prediction.

Contents

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is a medical imaging technique based on electromagnetic resonance properties of spin-bearing particles in a strong magnetic field. Hence, this technique is minimally invasive and gives picture of the inside of the body. Since the creation of this technique, many acquisition sequences have been developed to better capture the structural and metabolic complexity of the brain, the heart, and other parts of the body. An information that can be captured is the diffusion of molecules. The study of this is the aim of whole branch of MRI: diffusion MRI [14]. As the movement of the molecules is constrained by the tissue structure, diffusion MRI also provides information about the latter. This can be very useful to detect some diseases; for example, in multiple sclerosis, axons in brain and spine white matter are demyelinated, which modifies their apparent density; this can be observed and measured using diffusion MRI.

The problem is that the axons are at a microscopic scale, while the resolution of MRI is close to 1 millimeter. However, even if the microstructure cannot be observed directly, there are some techniques that give submillimetric information. This is comparable to light diffraction in optics, that allows to measure objects of small dimensions. In diffusion MRI, the observation is made by exciting the spin magnetic moments of the molecules, and measuring the signal attenuation due to the movement of the molecules which introduces phase incoherence in the spin magnetization.

This incoherence results from the position of the molecule in a magnetic field gradient. Hence, several different measures can be done using several different gradients. Taking more acquisitions yields to capturing more information and makes the reconstruction of diffusion features easier. However, the number of acquisitions should not be too high, because it would need too much time, with the subject inside the MRI scanner. It is actually assumed that the scan should not exceed one hour for in vivo human brain [2], and less in a clinical context. That's why some techniques emerged to reduce the number of acquisitions for reconstructing diffusion features or microstructure parameters [7, 19].

A promising method for acquiring fewer and recovering the original signal is called compressed sensing [6]. The idea of compressed sensing is to take advantage of the sparsity of the signal to recover the signal from only a few measurements. In this technique, the aim is to find a sparse representation of the signal. This is used in order to undersample the signal, the number of nonzero values tells us a minimum number of measures for a correct reconstruction of the signal.

Several studies used compressed sensing in diffusion MRI [11, 17, 19]. However, these studies do not take advantage of the degree of freedom that we have on the gradient waveforms, and are restricted to specific forms of gradients. The aim of the internship is to use compressed sensing techniques to find generalized gradient waveforms to efficiently reconstruct microstructure parameters.

In chapter 2, we first give some background about diffusion MRI, including a description of the sequence for a diffusion MRI acquisition and some diffusion features, then we describe some gradient waveforms that can be used in this sequence for a better estimation of the microstructure, and finally, we see how the number of acquisitions may be reduced, mainly using compressed sensing.

In chapter 3, we detail the contribution. We explain how we generate data for our experiments, then we describe the dictionary learning performed on this dataset, and how the resulting dictionary can be used to select gradient waveforms. Finally, experimental results are presented for evaluating the efficiency of the techniques.

Chapter 2

State of the art

Diffusion imaging allows to indirectly estimate microscopic features of the white matter. However, a large number of acquisitions remain necessary for these estimations. A promising technique to decrease this number of acquisitions is compressed sensing. In this chapter, we first explain the diffusion MRI techniques that rely on magnetic field gradients, and show how these gradients are an important acquisition parameter. Then, we present compressed sensing and give a quick review of studies that used compressed sensing in diffusion MRI.

2.1 Diffusion MRI background

Diffusion MRI and the acquisition sequence that is presented in this section dates back to 1965 [24]. Since, many ameliorations have been made, but the principle remains the same. Diffusion MRI can be used for the reconstruction of microstructure parameters and to detect some diseases. It can also be used to get some diffusion features, that are used to build a map of the connexions of the brain, called connectome.

2.1.1 Spin Echo sequence and diffusion-sensitizing gradients

The method described here is the most classical diffusion MRI sequence. It is based on spin echo in presence of a magnetic field gradient.

The classical sequence

This method consists in the study of the spin echo that occurs at a time $t = TE$ after two RF pulses (Fig. 2.2, on the left). For this kind of sequences, the subject is submitted to an intense magnetic field $\boldsymbol{B_0}$, of a few teslas. At steady state, the spin magnetic moment of each molecule angle is aligned with the magnetic field $\boldsymbol{B_0}$ (Fig. 2.1.A). At the time $t = 0$, the molecules are excited by an RF pulse in order to make the spin rotate by 90 degrees (Fig. 2.1.B). Hence, the spins are coherently aligned in the transverse plan. Then, for a duration of $TE/2$, the spins rotate in the transverse plan at a frequency ω. According to physics laws, and more precisely, to the Larmor precession, there is a constant γ, called the gyromagnetic ratio, such that $\omega(t) = \gamma B(t)$ (Fig. 2.1.C). At time $TE/2$, the molecules are excited by a second RF pulse that makes the spins

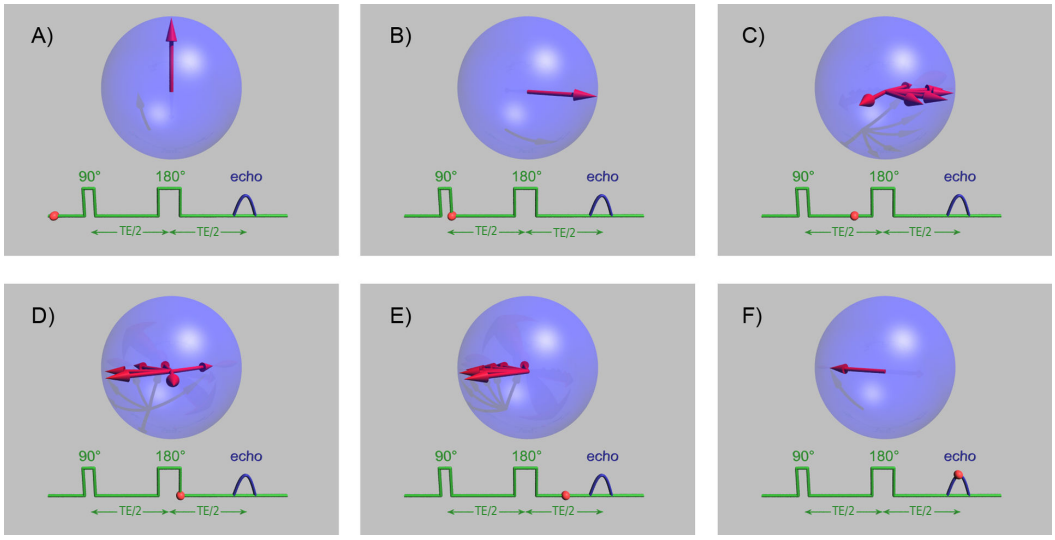


Figure 2.1: Influence of the spin echo sequence on spin magnetic moments. (Illustration of Gavin Morley)

rotate by 180 degrees (Fig. 2.1.D). As a consequence, the spins now rotate in the opposite direction (Fig. 2.1.E). Hence, for $t > TE/2$, $\omega(t) = -\gamma B(t)$. The phase-angle accumulated by the spin between $t = 0$ and $t = TE$ is:

$$\phi = \int_0^{TE} \omega(t)dt = \int_0^{TE/2} \gamma B(t)dt - \int_{TE/2}^{TE} \gamma B(t)dt \tag{2.1}$$

If $B(t)$ is constant, then, there is no phase-angle, and all the spins come back to alignment at the end of the sequence. This alignment yield to the observation of an echo at the end of the sequence (Fig.2.1.F).

Influence of a magnetic field gradient

In presence of a magnetic field gradient, we can observe an attenuation of the echo when the matter is submitted to a magnetic field gradient: the signal measured at $t = TE$ has a lower value (Fig. 2.2, on the right) than if the magnetic field was uniform (Fig. 2.2, on the left). This attenuation is due to the movement of the molecules of water in the matter. That's why this method permit to deduce some information about the diffusion, and as a consequence, about the microstructure in which the molecules are moving.

If there is a magnetic field gradient $\boldsymbol{g_r}(t)$, the magnetic field becomes $B(t) = B_0 + \boldsymbol{g_r}(t).\boldsymbol{r}(t)$ where $\boldsymbol{r}(t)$ is the position of the molecule at time t.

$$\phi = \int_0^{TE/2} \gamma(B_0 + \boldsymbol{g_r}(t).\boldsymbol{r}(t))dt - \int_{TE/2}^{TE} \gamma(B_0 + \boldsymbol{g_r}(t).\boldsymbol{r}(t))dt = \gamma \int_0^{TE} \boldsymbol{g}(t).\boldsymbol{r}(t)dt \tag{2.2}$$

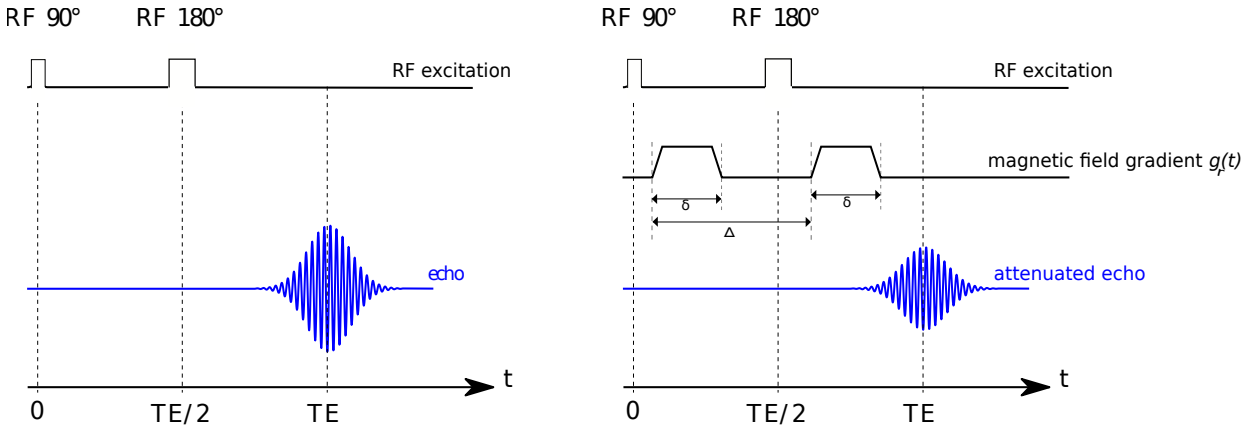


Figure 2.2: Spin echo sequence : on the left, the classical spin echo sequence ; on the right, the same sequence with a magnetic field gradient. The signal on the right is attenuated, it is diffusion weighted.

where $\boldsymbol{g}(t)$ is the effective gradient which is $\boldsymbol{g_r}(t)$ from 0 to $TE/2$ and $-\boldsymbol{g_r}(t)$ from $TE/2$ to TE. Remark that ϕ only depends on the trajectory $\boldsymbol{r}(t)$ of the molecule, and the diffusion-sensitizing gradient $\boldsymbol{g}$.

Mathematically, we can represent the normalised spin magnetization in the transverse plan by a dimensionless complex number. Therefore, if after the first RF pulse, every spin is the complex number 1, then at $t = TE$, the magnetic moment of a molecule m is $e^{i\phi_m}$. Hence, the signal attenuation can be computed by adding the spins [25]:

$$E = \frac{S(TE)}{S(0)} = \frac{1}{N}\left|\sum_{m=1}^{N} e^{i\phi_m}\right| \tag{2.3}$$

If the gradient is not specific, the phase-angles of the spins risk to to be completely uncorrelated, that may lead to vanish the resuling signal (as the mean of complex numbers with uniformly spread phases). In order to avoid this problem, we can force the integral of the effective between $t = 0$ and $t = TE$ to be zero. With this condition, every motionless spin will have the same phase-angle in the end, and the measured attenuation will only be due to spin motion. In general, the gradients present a symmetry at $t = TE/2$, that is to say $\boldsymbol{g_r}(TE/2 + t) = \boldsymbol{g_r}(TE/2 - t)$, like in [9].

2.1.2 Extracting microscopic information

The MRI resolution is at a millimetric scale, but we can collect statistical information about the microstructure. In the case of diffusion MRI, what is studied is the movement of the molecules. So the information that we recover in diffusion MRI is about this movement, and about the main direction of diffusion.

Diffusion features

One of the most widespread methods to model information about diffusion, is diffusion tensor imaging. This method relies on the computation of the diffusion tensor, which is a positive symmetric 3×3 matrix, characterizing the diffusion of molecules in the matter. A minimum of only 6 acquisitions is needed to reconstruct this tensor. The main direction of diffusion is then given by the eigen vector associated to the highest eigen value of the matrix.

In diffusion weighted imaging, we often try to reconstruct the ensemble average propagator (EAP) $P(\vec{r}|\vec{r}_0, \Delta)$. This propagator is the density probability for a molecule to move from a position $\vec{r}_0$ to a position $\vec{r}$ during a period Δ. This representation contains more information about the microstructure than the diffusion tensor and allows a better estimation of the microstructure, in particular for crossing fibers. In fact, the diffusion tensor contains a summary of the EAP, assuming that the EAP is a 3D Gaussian function, and computing the covariance matrix of this distribution. Contrary to the diffusion tensor imaging, the methods that permit to reconstruct the EAP need a lot more acquisitions.

Microstructure

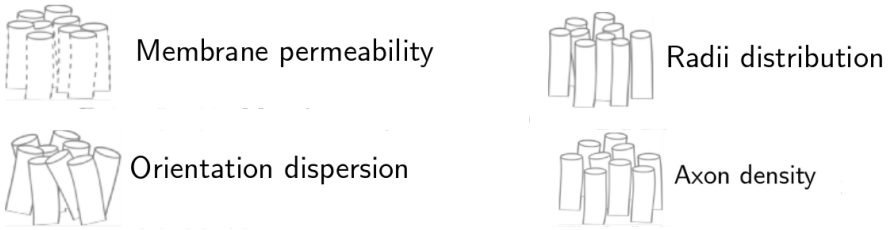
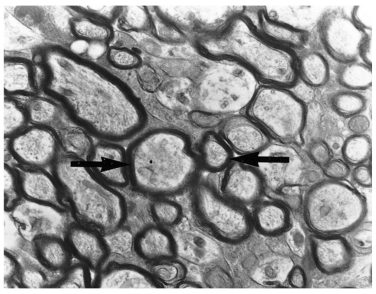


Figure 2.3: Microstructure parameters.

These two previous representations are very convenient, however they don't give concrete information about the microstructure, but only about the diffusion. To this end, we can also use parametric, biophysical models of microstructure [21]. Well known examples of such representation are a composite hindered and restricted model of diffusion (CHARMED) (Fig. 2.4) [4] and Axcaliber [3]. These kinds of representation rely on assumptions on the microstructure. We suppose that the matter corresponds to a model with only a few parameters about the fibers, such as membrane permeability, orientation dispersion, radii distribution, or axon density (Fig. 2.3). The aim of the acquisitions becomes the estimation of these parameters. For example, CHARMED models the water diffusion within "restricted" cylindrical axons to estimate their diameter, and diffusion around the axons as "hindered" by the latter to estimate the density (Fig. 2.4, right). AxCaliber is an extension of CHARMED that provides an estimation of a diameter distribution instead of a fixed diameter. AxCaliber also considers stationary water and free water in addition to the hindered and restricted diffusion. These representations are a very good approach to find concrete information about the microstructure.



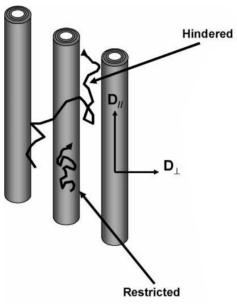


Figure 2.4: On the left: electron microscopy of a transverse section of a white matter axon bundle. On the right: CHARMED model. The modeling framework showing the two modes of diffusion in white matter, hindered outside the cylinders and restricted within the cylinders. Diffusion in the hindered part is characterized by a diffusion tensor.

One of the main advantages of these models, is that we now need only a few parameters to describe the microstructure. This low number of parameters is interesting, because in order to recover them, we need a number of measurements which is at least greater than the number of parameters to recover. In practice, the number of acquisitions is larger in order to get a higher precision and to compensate for the redundancy in the acquisitions.

2.2 Several existing families of diffusion gradients

One way to improve the precision of estimation without increasing too much the number of acquisitions, is to find magnetic field gradient waveforms which increase the sensitivity of the measured signal to microstructure parameters. In fact, there is a wide flexibility in the choice of the magnetic field gradient, since we can virtually choose any function of time for each of the 3 axes.

2.2.1 Constraints on the gradient waveforms

Yet, there are a few constraints to respect.

- The gradient magnitude is bounded and cannot exceed a given value. This limitation is due to hardware constraints, and physiological safety constraints. In fact, the subject can not endure too high magnetic field gradients.
- The slew rate: similarly to the magnitude, the gradient derivative also is limited. This limitation is also due to the hardware and the safety of the subject who cannot endure too fast variations of magnetic field.
- In order to get a high contrast between motionless and moving molecules, we also have to respect a symmetry constraint as explained in section 2.1.1

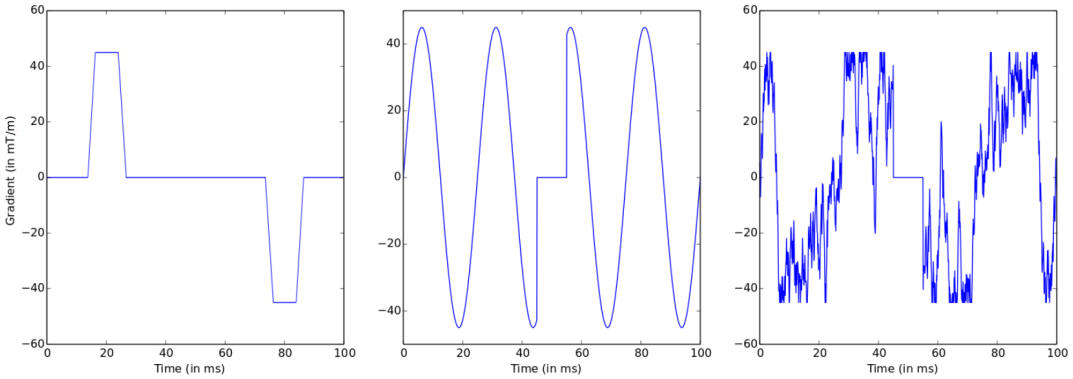


Figure 2.5: Different gradient waveforms. On the left: a pulsed gradient. In the middle: a sine oscillating gradient. On the right: a randomly generated gradient.

2.2.2 Pulsed Gradient Spin Echo (PGSE)

The Pulsed Gradient Spin Echo [24] is the sequence illustrated on the right of Fig. 2.2, and on the left of Fig. 2.5. In this sequence, the gradient waveform is zero almost everywhere but it has two pulses of the same constant direction, the same amplitude and the same duration δ. These two pulses are separated by a duration Δ. This sequence is one the most common, because it is quite simple, and in the approximation $\delta \ll \Delta$, we can express the signal attenuation as a Fourier transform of the EAP:

$$E(\boldsymbol{q}) = \int \rho(\boldsymbol{r_0}) \int P(\boldsymbol{r}|\boldsymbol{r_0}, \Delta) e^{i2\pi \boldsymbol{q}.\boldsymbol{r}} d\boldsymbol{r} d\boldsymbol{r_0} \tag{2.4}$$

where $\rho(\boldsymbol{r_0})$ is the density probability of finding a molecule at the position $\boldsymbol{r_0}$ when the first gradient pulse occurs, $P(\boldsymbol{r}|\boldsymbol{r_0}, \Delta)$ is the EAP, and $\boldsymbol{q} = (2\pi)^{-1}\gamma\delta\boldsymbol{g_{max}}$. Thus, we can reconstruct the EAP by computing the inverse Fourier transform of the signal attenuation. However, this computation needs to measure the signal attenuation for a lot of different $\boldsymbol{q}$ -vectors. This can be done by measuring the signal attenuation for several pulsed gradients, modifying the direction, δ, Δ and the amplitude.

2.2.3 Gradient waveforms specialized for microstructure reconstruction

PGSE is very convenient for the estimation of the EAP, but it may not be the case for the reconstruction of the microstructure. In [9], Dobnjak et al. try to find some sets of gradient waveform to optimize the estimation of microstructure parameters, and more precisely, the axon diameter. The search is made among all possible waveforms in a discretized time, and respecting the technical limits. The optimized gradient waveforms that result from the experiments allow a better estimation than PGSE for small diameters. It also appears that those optimized gradient waveforms are very similar to oscillating functions. They also noticed that the optimized frequency of the oscillations increases when the diameter decreases. More precisely, the characteristic period of the oscillations is close to the time for the molecule to travel the diameter of the axons.

Those results were then confirmed in [22]. While in [9], generalized gradient waveforms were optimized, in [22], gradients were limited to pulsed gradients, or oscillating gradients. In fact, Siow et al. [22] compare pulsed gradient with the following protocols: sine oscillating gradients (SNOGSE, Fig. 2.5, in the middle) , sine with arbitrary phase oscillating gradients (SPOGSE), and square oscillating gradients (SWOGSE). They optimized the length, duration, frequency and phase of the waveforms for a fixed gradient magnitude. They showed that the choice of the protocol does not influence a lot the estimation, the SWOGSE gives slightly better results than the others. Each optimized set of gradients reveals a mix of low and high gradient frequencies.

In the two previous examples, either there is a total freedom on the choice of the gradient waveforms, or there is only one parameter to optimize (the frequency). But none of them provide a gradient waveforms family that gives a summarized representation of the gradient waveforms without constraining too much the waveform.

2.2.4 Parameterization of the gradients

Other families of gradient waveforms have been proposed to give an even better estimation of the microstructure parameters. Many of them rely on making the orientation of the gradient changing [23, 26]. For example, Daniel Topgaard [26] uses pulsed gradients, but the direction of the gradient evolves during the pulse. This technique captures information about the diffusion in several direction, in only one acquisition.

Another proposition of gradient waveforms [20] is expressed in terms of cosine series:

$$g(t) = \Sigma_n c_n \cos(2\pi n t/TE) \tag{2.5}$$

where the c_n are the coefficients of the gradient in the corresponding basis. However, the optimization of these kind of gradient waveforms yields to the use of square oscillating gradients, as well as before. But finding a good parameterization of the gradients for the diffusion MRI is not easy, and there is no consensus on this question so far.

2.3 Compressed sensing

One of the biggest problem of the most advanced techniques in diffusion MRI is that they need a lot of acquisitions. Since each acquisition takes 5-10s, when a great number of acquisitions is required, the subject may be asked to stay a long time in the MRI machine. In this section, we will see how the number of measurements can be reduced using compressed sensing. The point of compressed sensing is to perform two tasks, acquisition and compression, at once, making some assumptions on the sparsity of the signal. In compressed sensing the number of measurements is the size of the compressed vector. After a presentation of the method of compressed sensing, we will see that it can be combined with dictionary learning, and then we will see some use of compressed sensing in MRI and diffusion MRI.

2.3.1 The principle

The Nyquist-Shannon sampling theorem states that if the signal's highest frequency is less than half of the sampling rate, then the signal can be reconstructed perfectly. The aim of compressed sensing [6] is to make the reconstruction possible even if there are fewer samples. This reconstruction can be done by making some assumptions about the sampled signal. First, the signal has to be *sparse*, that means it must have only a few nonzero values. In practice, the signal s, of dimension N, is rarely sparse, but we can generally find a transform ψ in which the signal is sparse. We can then consider the vector x, with only $K << N$ nonzero entries (x is said K-sparse), such that $s = \psi x$. Then, we are looking for a sensing matrix A of dimensions $M \times N$ such that the compressed signal is $y = Ax$ (Fig. 2.6). Each row of the sensing matrix can be seen as a measurement vector, and the vector y contains the measures.

Once the sensing matrix is fixed, we want to be able to reconstruct the vector x, by knowing the compressed vector y and the sensing matrix A. Since the equation $y = Ax$ has an infinite number of solution, we have to force the resolution to give a sparse result. This resolution can be considered as an optimization problem, in which we try to minimize the number of nonzero entries of the solution (it is the ℓ_0-norm of the solution). Unfortunately, this optimization problem in not convex, and the resolution is NP-hard. That's why the ℓ_1-norm is often used instead of the ℓ_0-norm. Then, the problem can be solved using linear programming.

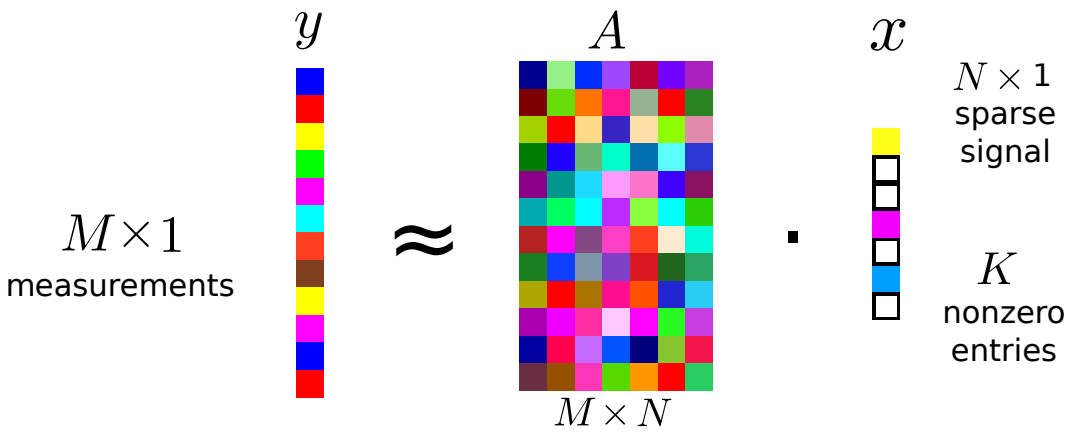


Figure 2.6: Compressed sensing main relation

The difficulty is to find a sensing matrix such that the resolution is possible. Candes and Walkin [6] points out several conditions that the sensing matrix must satify:

A spark greater than $2K$: The spark of a given matrix A is the smallest number of columns of A that are linearly dependent. It has been proved that for any vector $y \in \mathbb{R}^M$, there exists at most one K-sparse signal x such that $y = Ax$ if and only if the spark of A is greater than $2K$.

The Null Space Property (NSP): The NSP quantifies the notion that vectors in the kernel of the sensing matrix should not be too concentrated on a small subset of indices. If the sensing matrix satisfies the NSP, then the error between the vector to reconstruct, and the solution returned by the optimization is bounded.

The Restricted Isometry Property (RIP): The RIP of order K is a property that quantifies how much the ℓ_2-norm of a K-sparse signal can be modified by a multiplication by the matrix that satisfies the property. If a matrix A satisfies the RIP of order $2K$, then we can interpret the property as saying that A approximately preserves the distance between any pair of K-sparse vectors.

A low coherence: The coherence of a matrix A, $\mu(A)$, is the largest absolute inner product between any two columns a_i, a_j of A: $\mu(A) = \max_{i,j} \frac{|\langle a_i, a_j \rangle|}{||a_i||_2 ||a_j||_2}$. Intuitively, we understand that a low coherence limits the redundancy of the columns of A.

A particular attention in generally given to the low coherence constraint, because a low coherence gives some guarantees about the other ones. For example, the spark of a matrix is necessarily greater than $1 + \frac{1}{\text{coherence}(A)}$.

2.3.2 Link with dictionary learning

The dictionary learning is a method that solves a problem useful to compressed sensing, it can be used if there is no known space in which the signal is sparse (i.e if ψ is not known). For example, K-SVD [1] is an algorithm that, given a training set of signals y_i, tries to find a dictionary that leads to the best representation for each member in this set, under sparsity constraints. This is done by solving the equation $y_i = Dx_i$ where D, a $n \times K$ matrix, and x_i, a K-long vector, are both unknown. The algorithm used to solve this equation alternates between two phases of optimization. The first one tries to find the bet set of x_i with a fixed dictionary D, that minimizes the error under constraints of sparsity of the x_i. The second phase updates the dictionary D and the nonzero values of the x_i, using the Singular Value Decompisition (SVD, the generalization of the spectral theorem to nonsquare matrices) of the error matrix. These two phases are repeated until convergence. This learning algorithm provides a sparse representation of signals, with a low loss of information.

Merlet et al. [18] propose an adaptation of K-SVD for the reconstruction of two diffusion features: the Ensemble Average Propagator (EAP), and the Orientation Distribution Function (ODF). The dictionary is here characterized by a set of polynomial and scale parameters. This characterization is well adapted to sparsely and continuously model the diffusion signal. Once the dictionary learned, it is used as a sensing matrix, to reconstruct the sparse signal, in a known space. This signal can be used to recover the EAP and the ODF.

Gramfort et al [11], also used dictionary learning in diffusion MRI. In their study, they perform online dictionary learning to learn the structure of the diffusion-weighted signal in the brain. This dictionary is used for both denoising and undersampling. The learning relies on physical properties of the signal such symmetry and positivity. An encouraging result about dictionary learning in diffusion MRI comes from Bilgic et al [5] who showed that learning from acquisitions of one subjects generalizes well to other subjects.

2.3.3 Compressed sensing in diffusion MRI

As the number of acquisitions in limited in diffusion MRI, several techniques are already used to use as fewer measurements as possible. For example, q-space imaging is a method to find small sets of $\boldsymbol{q}$

vectors as described in section 2.1.1 such that we can perform the inverse Fourier transform. In [7], Caruyer et al. propose schemes of acquisitions in which the $\boldsymbol{q}$ are spread on balls. This approach gives a way of decreasing the number of acquisitions by utilizing the best angular distributions of the magnetic field gradients.

Several studies use compressed sensing in anatomical MRI [15]. In order to respect the necessary condition of sparsity, the signal is not expressed in the pixel domain. Thus, the Transform Point Spread Function (TPSF) is used for changing the basis. A particular attention is also given to the incoherence of the acquisitions. The measurements that are chosen, resulting from the choice of the sensing matrix, are very various, limiting their redundancy. Those results on anatomical MRI give promising perspectives on the use of compressed sensing for limiting the number of measurements in diffusion MRI.

Later work [17] studied how to use compressed sensing in diffusion MRI. Here, the acquisitions are made with the PGSE sequence, the compressed sensing is then used to choose the $\boldsymbol{q}$ vectors of the used gradients. For the q-space imaging, other basis are used to represent the signal sparsely. Merlet et al compared 4 of them and found that the SHORE basis seems to be the more appropriate for the compressed sensing recovery.

Michailovich et al [19] proposed a reconstruction method for High Angular Resolution Diffusion Imaging (HARDI) using compressed sensing simultaneously in the spatial and the diffusion domains. The sparsifying transform for the diffusion domain uses spherical ridgelets. This method lead to reliable reconstruction with only 16 measurements while between 60 and 100 are usually required.

Those studies showed that compressed sensing can be used to decrease the number of measurements, and to shorten the acquisition times. However, ameliorations can still be made. In fact, previous studies intend to use compressed sensig to optimize q-values [17], gradient directions [19], or to denoise the signal [11]. But no study using compressed sensing and different forms of gradient waveforms have been found. Generally, only PGSE sequences are used. There may be even more satisfying results by taking advantage of the degrees of freedom on the gradient waveforms.

Chapter 3

Contribution

From the study of the state of the art, we can conclude that a promising idea to reduce the number of acquisitions is to combine compressed sensing with general magnetic field gradient waveforms. Thus, during the internship, we intended to build a dictionary that could allow us to predict the signal value for a given gradient waveform from a few samples. In this chapter, we present how we built this dictionary, beginning by explaining how we generated data to perform the learning. Then we present how we can use the dictionary to optimally select a subset of the samples that would give a good reconstruction of the signal. Finally, we present experimental results that compare the different techniques of gradient selection.

3.1 Data generation

In order to perform dictionary learning on signals, we need a large database of measurements. Since using MRI scanners is very expensive and time consuming, in order to get sufficiently large training set, we use Monte Carlo simulations instead of real acquisitions.

3.1.1 Monte Carlo simulations

The signal are obtained using the tool Camino [8]. The Monte Carlo simulations implementation is based on Hall and Alexander work [12]. The principle of these simulations is to compute trajectories of a large number of molecules. For a given gradient waveform, it is then possible to compute the phase-angle accumulated in the whole simulation for each molecule spin with equation 2.2. Finally, the signal is computed by addition of all the spins.

Microstructure variability

Since we want to perform dictionary learning on the signals, it is necessary not to have a biased dataset. Then the microstructure parameters must be various. Camino allows us to generate any kind of microstructure. For our experiments, we use the model "Irregularly Packed, Distributed Radius Cylinders". This model is made of parallel cylinders. Their radii follow a gamma distribution with two parameters (mean radius and shape). Another parameter that we vary in our simulations is the number of cylinders in the square in which the cylinders are uniformly distributed. This last parameter modifies the density of the cylinders. This model is very close to the AxCaliber model

(restricted and hindered diffusion, and distributed radii). The parameters of our experiments are the following:

- the mean radius is in [0.5 μm, 1 μm, 1.5 μm, 2 μm, 2.5 μm, 3 μm].
- the shape parameter of the gamma law is in [1.5, 3, 4.5, 6, 7.5, 9].
- the number of cylinders in a square of side 50 μm is in [50, 100, 150, 200, 250]. If the mean radius and the number of cylinder are both too high, it means that the simulation will not be able to fit all cylinders in the space, then the number of cylinders is automatically reduced.

This gives us 180 different microstructure parameter combinations. The minimum intracellular volume fraction is around 0.015, and it reaches a maximum around 0.8 when all cylinders cannot be placed.

Gradients

In order to reduce the search space, we decided to restrict ourselves to piecewise constant gradients with a fixed orientation. In what follows, we use an exhaustive list of gradients with four time steps and 3 possible values ($-g_{\max}$, 0, $+g_{\max}$). A symmetry is used to complete the gradient waveform after the RF180 pulse. We then filter them to keep only those which actually reach the value $+g_{\max}$. This leads to a set of 65 waveforms. Last, in order to better cover the space, we make the simulations for these waveforms in 40 directions. These directions are selected uniformly on the unit sphere [7]. We now have a total of 2600 gradients for our study.

The simulated signals are normalized by dividing the value of the simulation by the value of a simulation without magnetic field gradient (called b_0 signal). This gives us the measure of the attenuation E as in equation 2.3. Studying the attenuation instead of the signal without normalization does not make us lose information about the diffusion. In fact, the b_0 signal is not sensitive to the diffusion.

3.1.2 Data augmentation

An important feature to avoid having a biased dictionary is having rotation invariant data. In fact, if all microstuctures of the learning dataset have the same orientation, the gradients that are perpendicular to the fiber direction will have too much importance in the dictionaryb and the dictionary will not generalize to situations where the microstructure has a different orientation. Thus, we proceed a data augmentation to generate signals for microstructure made of fibers in several directions.

To have several directions of microstructure: we interpolate the signal using spherical harmonics of rank $L = 6$. For every waveform, we use the 40 signals that comes from the 40 gradients directions to build a continuous function. Then we simply sample the signal values that correspond to the rotated signal.

Finally, we have 100 directions uniformly spread on the unit sphere for every previous microstructure. We now have 18,000 microstructures in our dataset.

3.2 Dictionary learning

We used the implementation of Mairal et al algorithm [16] in SPAMS (SPArse Modeling Software). SPAMS is an optimization toolbox for solving various sparse estimation problems. It aims at solving:

$$\min_{D,x_i} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} ||y_i - Dx_i||_2^2 + \lambda ||x_i||_1 \tag{3.1}$$

where n is the number of vectors in the learning set and λ is a parameter that quantifies the constraint on the sparsity of the sparse representations x_i. As well as k-SVD, the algorithm used here by SPAMS alternatively updates the x_i (with the LARS-Lasso algorithm [10]) and the dictionary D. The optimization constraints the dictionary D to have normalized columns.

There are mainly two parameters to choose: the number N of atoms of the dictionary (i.e the number of columns of D) and the parameter λ. The number of atoms need to be big in order not to lose too much information by compression. After several tries, it has been chosen to be 200, that allows a good reconstruction of signals, and has a reasonable computation time. λ is fixed at 0.15 to have about 20 nonzero values according to SPAMS documentation. The dictionary is learned using 20% of the 18000 generated signals.

3.3 Gradients selection

In practice, only a few measures can be done. We must choose the gradients in order to reconstruct the full signal efficiently. In this section, we describe several ways for selecting the gradients from the dictionary.

3.3.1 Minimizing the redundancy of the measures

A first approach for selecting the gradients among the available ones is to select a set of lines of the dictionary that are uncorrelated. The idea behind this choice is that if two lines of the dictionary are similar, then the two corresponding gradients capture the same kind of information, they have the same response. Hence, minimizing the correlation of a subset of lines of the dictionary may limit the redundancy of the measures. The correlation is measured as followed: the lines of the dictionary D are centered and reduced to get the matrix $\widetilde{D}$, then the correlation matrix is $C = \widetilde{D}\widetilde{D}^T$, and for a subset Ω of the gradients, the correlation is the Frobenius norm of the matrix C restricted to the lines and columns of Ω. This defines the following correlation score:

$$f(\Omega) = \sum_{i,j \text{ in } \Omega} \left(\sum_k \widetilde{D}_{ik} \widetilde{D}_{jk} \right)^2 \tag{3.2}$$

In order to find the set Ω that minimizes the correlation score, we perform a local and discrete optimization. The algorithm consists in:

- Choosing an initial subset Ω_0 of gradients with a greedy algorithm.
- While the correlation decreases, at step t:

- we find the gradient $i \in \Omega_t$ that is the most correlated with the other gradients of Ω. If $C_{\Omega,k}$ is the column k of the matrix C restricted to the lines of Ω, we have:
  $i_t = \underset{k \in \Omega_t}{\text{argmax}}(f(\Omega_t \setminus \{k\})) = \underset{k \in \Omega_t}{\text{argmax}} ||C_{\Omega_t,k}||_2$
- then we find the gradient $j_t \notin \Omega_t$ that is the less correlated with Ω_t. We have:
  $j_t = \underset{k \notin \Omega_t}{\text{argmin}}(f((\Omega_t \setminus \{i_t\}) \cup \{k\})) = \underset{k \notin \Omega_t}{\text{argmin}} ||C_{\Omega_t \setminus \{i_t\},k}||_2$
- finally, we update $\Omega_{t+1} = (\Omega_t \setminus \{i_t\}) \cup \{j_t\}$

- We stop the optimization when the last change of index does not improve the objective function.

Since this algorithm is a local optimization (we change only one element of Ω at each step), there is a risk to fall in local minimum. Thus, the result may be far from the real minimum. Finding a theoretical lower bound of the objective function would give us more confidence in the result. A lower bound can be given by the result of the relaxed optimization problem. In fact, since the relaxed problem is a convex optimization, the minimum that is found is global and not local, so the result gives us a lower bound for the minimum that can be found in the discrete problem. In the relaxed problem, instead of having a subset of gradients Ω, we have a positive variable α_i for each available gradient. The sum of all the variables must be the size of Ω and every α_i has to be positive. The objective function becomes :

$$f(\alpha) = \sum_{i,j \text{ gradients}} \alpha_i \alpha_j \left(\sum_k \widetilde{D}_{ik} \widetilde{D}_{jk} \right)^2 \tag{3.3}$$

The difference between the continuous and discrete optimization is shown on Fig. 3.1. For comparison, we also show a plot of the correlation score when the indices are selected randomly (the plotted score is averaged on 100 sets of randomly selected indices). Since the blue curve is close to the orange one, we can conclude that the discrete minimization ran well, as the result is close to the theoretical minimum.

3.3.2 Optimizing the properties of the sensing matrix

Another strategy to choose the gradients it to minimize the coherence as defined in section 2.3.1. The idea is to find a sensing matrix that is supposed to give a sparse representation according to the matrix properties descried by Candes and Wakin. Here, we want to find:

$$\widehat{\Omega} = \underset{\Omega, |\Omega|=n}{\text{argmin}}(\mu(\Omega)) = \underset{\Omega, |\Omega|=n}{\text{argmin}} \left(\max_{i,j \text{ atoms}} \frac{|\langle d_{\Omega,i}, d_{\Omega,j} \rangle|}{||d_{\Omega,i}||_2 ||d_{\Omega,j}||_2} \right) \tag{3.4}$$

where $d_{\Omega,i}$ is the column i of the dictionary restricted to the lines described by Ω. The algorithm to find $\widehat{\Omega}$ is similar to the algorithm to minimize the lines correlation.

- We choose an initial set Ω_0
- While the objective function decreases, at step t:

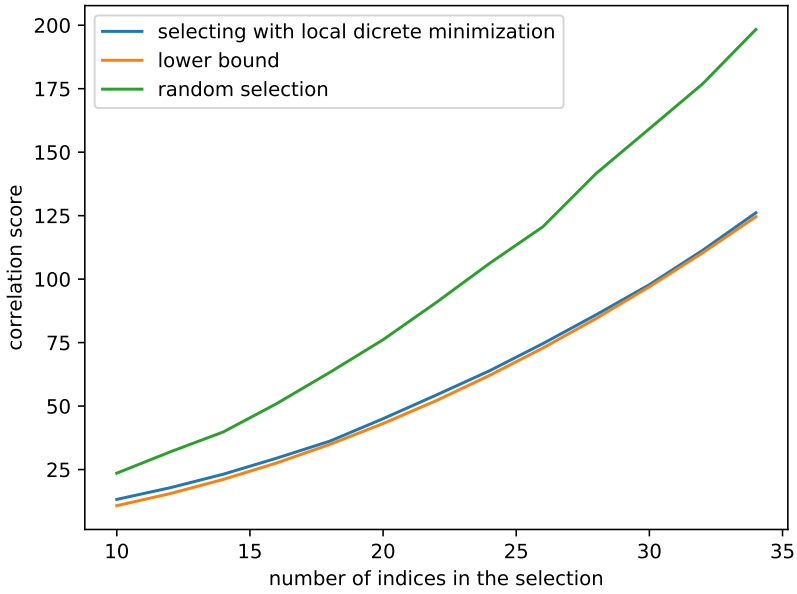


Figure 3.1: Comparing the discrete and local optimization with theoretical lower bound.

  - $i_t = \underset{k \in \Omega_t}{\operatorname{argmin}}(\mu(\Omega_t \setminus \{k\}))$
  - $j_t = \underset{k \notin \Omega_t}{\operatorname{argmax}}(\mu((\Omega_t \setminus \{i_t\}) \cup \{k\}))$
  - $\Omega_{t+1} = (\Omega_t \setminus \{i_t\}) \cup \{j_t\}$
- We stop the optimization when the last change of index does not improve the objective function.

Even if this algorithm is very similar to the algorithm minimizing the lines correlation, it costs a lot more in computation. In fact, here we have to compute $\mu(\Omega)$ for every considered set Ω, and this computation is long because it has to compute N norms, and N^2 scalar product of vectors of size n. At each iteration, $\mu(\Omega)$ is evaluated M times. Then the complexity of one iteration is $O(MnN^2)$. To minimize the lines correlation, if C is pre-computed, the complexity of one iteration is $O(Mn)$.

Another thing that makes this optimization harder than the previous one, is that it is not possible to find a relaxed problem. Then we do not have a theoretical lower and we can not check that our local optimization did not reach a local minimum far from the real minimum.

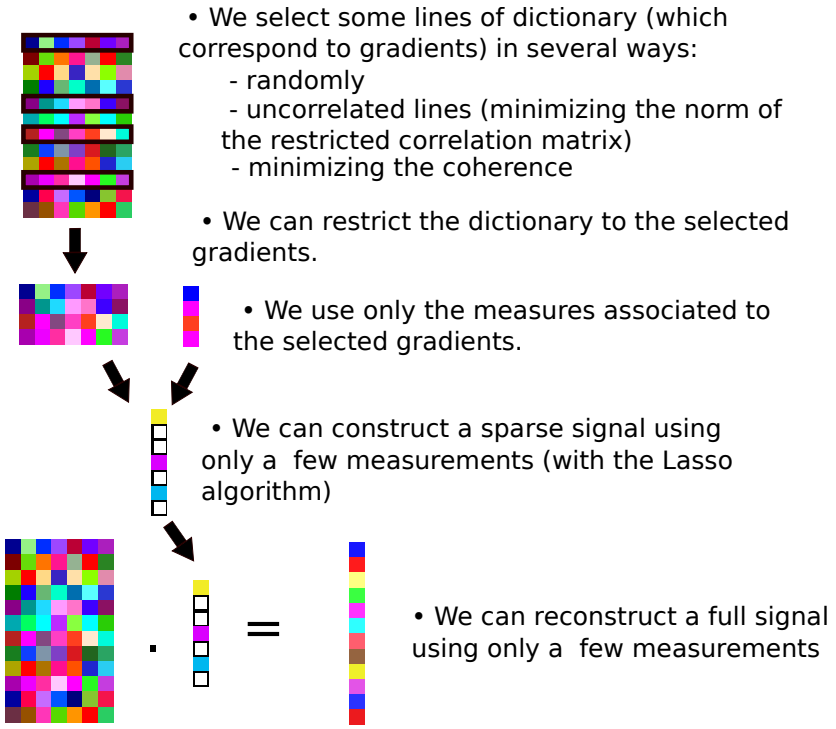


Figure 3.2: Reconstruction of a full signal from a few samples.

3.3.3 Evaluation of a selection

In order to evaluate how good is our selection of measures, we need to study how efficient are this measures to predict the unseen data. Thus, we first evaluate how good is the reconstruction of the full signal using only these measures. We proceed as follows:

- Select a subset Ω of gradients using one of the previous algorithms (from sections 3.3.1 and 3.3.2).
- Find a sparse representation with the LARS-Lasso algorithm [10] (for the ℓ_1 minimization) using only the measures corresponding to the gradients of Ω and the dictionary restricted to the lines corresponding to the gradients of Ω.
- Multiply the sparse representation x with the complete dictionary to get a reconstructed signal $\widehat{y}$.
- Compute a distance between y and $\widehat{y}$ (in our case, we use the Euclidean distance). We compute this distance for signals that are not in the learning set.

An illustration of the reconstruction pipeline is given on Fig. 3.2. The computed distance allows us to evaluate the quality of the reconstruction. The smaller this distance, the better the reconstruction. Then, we can evaluate the quality of a selection technique by measuring the quality of the reconstruction it provides.

On Fig. 3.3, we can observe an example of a reconstruction. On this figure, the abscissa axis does not have a particular meaning, except the fact that every set of 40 gradients with the same waveform but different directions are gathered together. We can see that there are as expected only a few nonzero values in the sparse representation. The atom 50, that is predominant in the representation of this signal, is also the atom that is the more used in the representation of all the signals. This atom is isotropic, it has the same value for every gradient with the same waveform, wathever the direction is. The other atoms that appear in the representation are less general, and anisotropic.

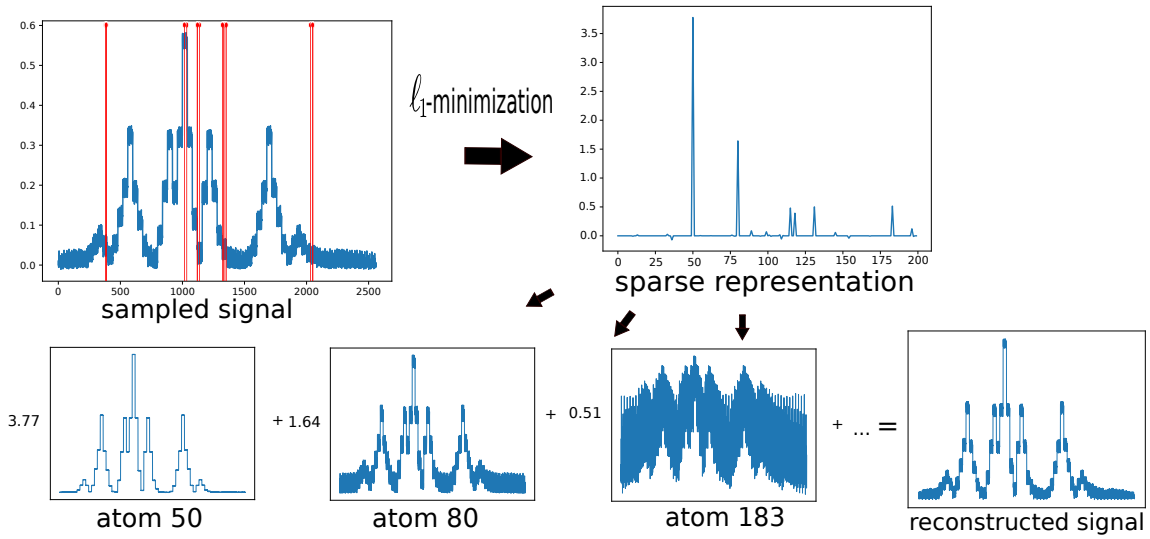


Figure 3.3: Example of a reconstruction of a full signal from a few samples.

We made several experiments to find which method provide a better reconstruction. For the experiments of figures 3.4 end 3.5, the number of gradients used for the reconstruction. For the experiments of figures 3.6 and 3.7, λ was set to 5E-5. To have significant results, we plotted the sum of the distance for the signals that were not in the learning set.

Figures 3.4 and 3.6 compare the efficiency of the reconstruction depending on the technique to choose the gradients. To have significant results, we plotted the sum of the distance for the signals that were not in the learning set. We can notice on Fig. 3.6 that the 3 techniques give a similar score when the number of gradients is higher than 25, but if the number of authorized measurements is lower, the best technique consists in minimizing the correlation of the lines of the restricted dictionary. However, Fig. 3.4 shows that we have to choose λ carefully. If it's too small, there is over-fitting because the LARS-Lasso algorithm gives too much importance to the data fidelity and is not able to generalize for gradient waveforms that are not used to find the representation. On the contrary, if it's too high, the algorithm finds a representation that is too sparse, then the representation does not contains enough information. We can see on figures 3.5 and 3.7 that, according to what Candes and Wakin stated, the sensing matrix with a low coherence gives a sparser representation. The two other methods have a similar sparsity.

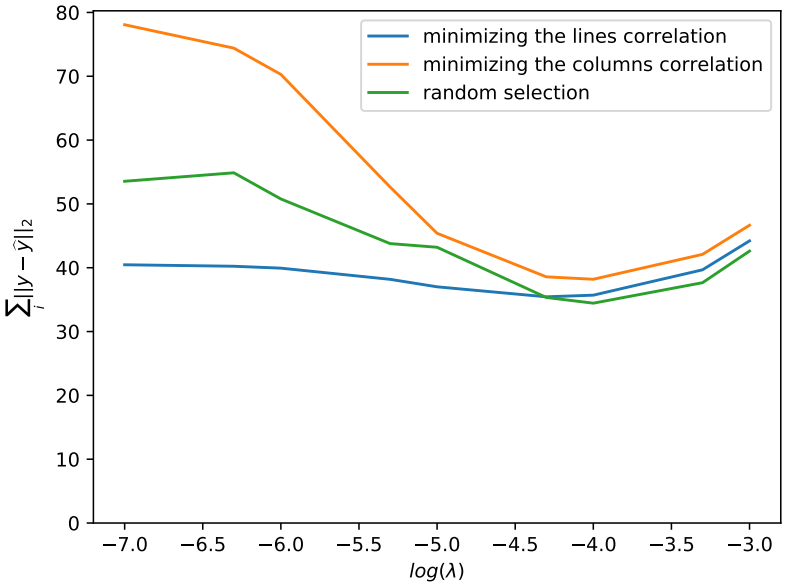


Figure 3.4: Fidelity of the reconstruction depending on λ (using 30 measures for the reconstruction).

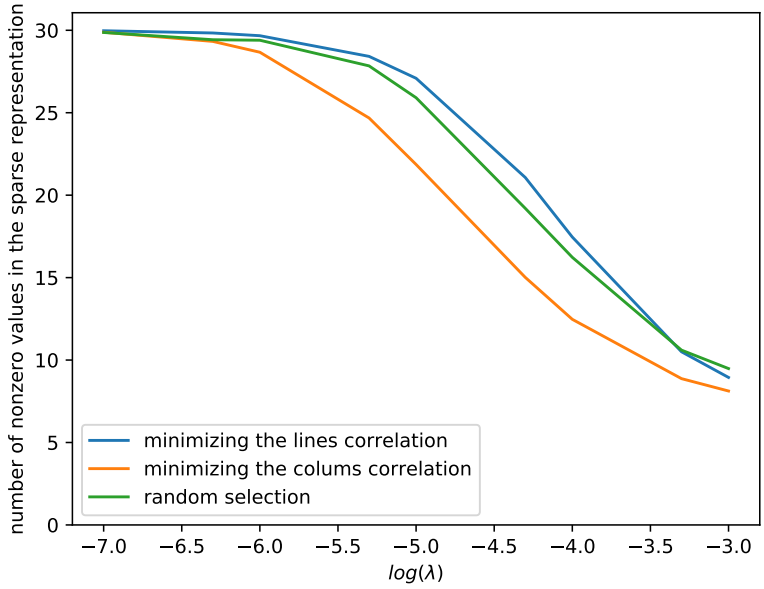


Figure 3.5: Sparsity depending on λ (using 30 measures for the reconstruction).

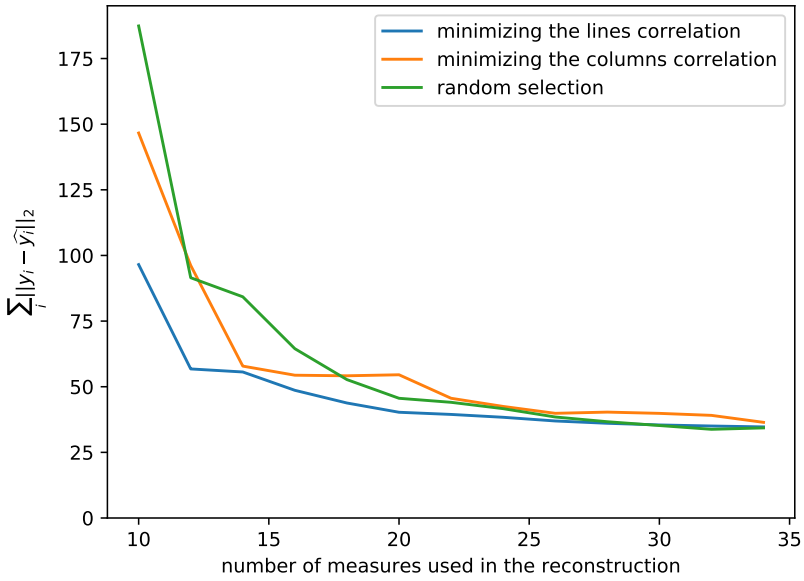


Figure 3.6: Fidelity of the reconstruction depending on the number of measures used. For a small number of measures, minimizing the lines correlation offers a better prediction.

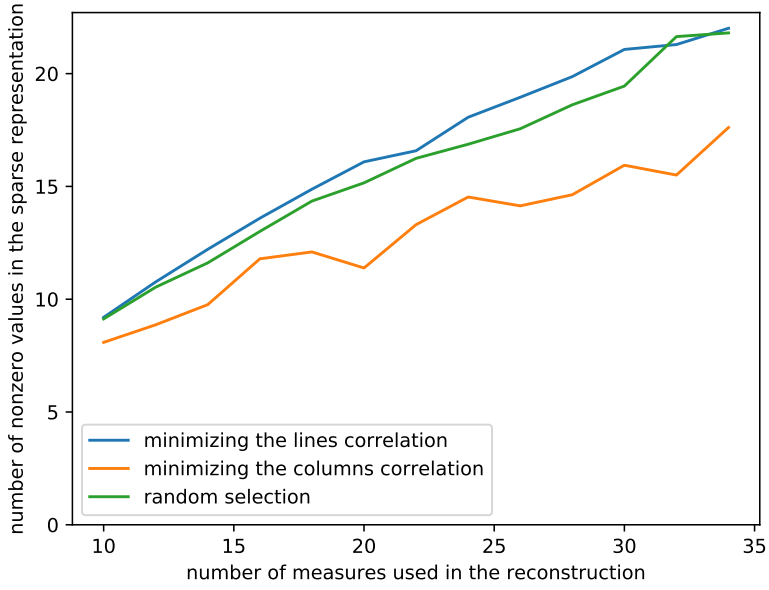


Figure 3.7: Sparsity depending on the number of measures used. Minimizing the columns correlation provides a sparser representation.

Chapter 4

Conclusion

Discussion

Diffusion MRI is a successful technique to reconstruct microstructure parameters and diffusion features, but many ameliorations remain possible. In fact, there is a wide flexibility in the choice of acquisition parameters such as the magnetic field gradient waveforms. However, only a few families of gradient waveforms have been proposed, and there is no convincing parameterization of the gradients. For the moment, the most efficient gradient waveforms are simple: pulsed gradients or oscillating gradients.

Since compressed sensing already gives good results in anatomical MRI and in diffusion MRI to optimize pulsed gradient, the idea for finding gradient waveforms better-suited to microstructure reconstruction was utilizing the sparsity of the signal. We proposed a method based on dictionary learning and compressed sensing to predict the signal for a given set of magnetic field gradients. The first results are encouraging and the signal seems to have a sparse representation.

The proposed technique still relies a lot on the set of samples that are used. Even if a random selection of the samples already gives interesting results, when the number of samples is too small, the random selection may choose redundant measures. In fact, selecting the samples by minimizing the correlation of the measures leads to a better production for a small number of samples. Another tested technique of selection is based on the minimization of the coherence of the sensing. As expected, this technique leads to a sparser representation. However, it does not provide a better prediction.

Further work

As explained before, our method allows us to predict the signal for a given set of magnetic field gradients. For the moment, this set of gradients is quite simple, and could be made more general. Then, we would like to find a parameterization of the magnetic field gradient waveforms to decrease the number of gradients used in the simulations, in order to decrease the number of lines of the dictionary. The parameterization should take advantage of the gradients waveforms properties.

We should also filter the gradients with their b-value. A too high b-value would be responsible of a too high attenuation of the signal and as a result a poor signal-noise ratio. Conversely, a too small b-value would not have enough impact on the signal. In both cases, the signal value does not vary enough between a microstructure and another.

Another possibility to find interesting waveforms is to see them as function of the time in $\mathbb{R}^3$. Then, we would try to find a good way to sample this space of functions to avoid the redundancy of the measures. An idea to do so is minimizing $\sum \frac{1}{d_{i,j}}$ under the constraint that the maximum value of the gradient is reached and where $d_{i,j}$ is the (integral) distance between two waveforms. This idea is based on electrostatic repulsion [13], in which we use the integral distance instead of a spatial distance.

However, even if we have an efficient way to predict the signal, we still have to check that the predicted signals are good enough for estimating the model parameters. Future studies will compare the reconstruction of the microstructure parameters depending on the techniques and on the algorithm parameters.

For further work, we would also like to acquire real data using phantoms, with known parameters instead of using Monte-Carlo simulations.

Bibliography

[1] M. Aharon, M. Elad, and A. Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". In: *Trans. Sig. Proc.* 54.11 (Nov. 2006), pp. 4311–4322.

[2] Daniel C Alexander. "A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features". In: *Magnetic Resonance in Medicine* 60.2 (2008), pp. 439–448.

[3] Yaniv Assaf et al. "Axcaliber: A method for measuring axon diameter distribution from diffusion MRI". In: *Magnetic Resonance in Medicine* 59.6 (2008), pp. 1347–1354.

[4] Yaniv Assaf et al. "New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter". In: *Magnetic Resonance in Medicine* 52.5 (2004), pp. 965–978.

[5] Berkin Bilgic et al. "Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries". In: *Magnetic Resonance in Medicine* 68.6 (2012), pp. 1747–1754.

[6] E. J. Candes and M. B. Wakin. "An Introduction To Compressive Sampling". In: *IEEE Signal Processing Magazine* 25.2 (Mar. 2008), pp. 21–30.

[7] Emmanuel Caruyer et al. "Design of multishell sampling schemes with uniform coverage in diffusion MRI". In: *Magnetic Resonance in Medicine* 69.6 (Apr. 2013), pp. 1534–1540.

[8] PA Cook et al. "Camino: open-source diffusion-MRI reconstruction and processing". In: *14th scientific meeting of the international society for magnetic resonance in medicine*. Vol. 2759. Seattle WA, USA. 2006, p. 2759.

[9] Ivana Drobnjak, Bernard Siow, and Daniel C. Alexander. "Optimizing gradient waveforms for microstructure sensitivity in diffusion-weighted MR". In: *Journal of Magnetic Resonance* 206.1 (2010), pp. 41–51.

[10] Bradley Efron et al. "Least angle regression". In: *The Annals of statistics* 32.2 (2004), pp. 407–499.

[11] Alexandre Gramfort, Cyril Poupon, and Maxime Descoteaux. "Denoising and fast diffusion imaging with physically constrained sparse dictionary learning". In: *Medical image analysis* 18.1 (2014), pp. 36–49.

[12] Matt G Hall and Daniel C Alexander. "Convergence and parameter choice for Monte-Carlo simulations of diffusion MRI". In: *IEEE transactions on medical imaging* 28.9 (2009), pp. 1354–1364.

[13] Derek K Jones, Mark A Horsfield, and Andrew Simmons. "Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 42.3 (1999), pp. 515–525.

[14] Denis Le Bihan and E. Breton. "Imagerie de diffusion in-vivo par résonance magnétique nucléaire". In: *Comptes-Rendus de l'Académie des Sciences* 93.5 (Dec. 1985), pp. 27–34.

[15] Michael Lustig et al. "Compressed sensing MRI". In: *IEEE SIGNAL PROCESSING MAGAZINE*. 2007.

[16] Julien Mairal et al. "Online learning for matrix factorization and sparse coding". In: *Journal of Machine Learning Research* 11.Jan (2010), pp. 19–60.

[17] Sylvain Merlet. "Compressive sensing in diffusion MRI". Theses. Université Nice Sophia Antipolis, Sept. 2013.

[18] Sylvain Merlet, Emmanuel Caruyer, and Rachid Deriche. "Parametric dictionary learning for modeling EAP and ODF in diffusion MRI". In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Vol. 7512. Lecture notes in computer science. Nice, France: Springer, Oct. 2012, 8 p.

[19] Oleg Michailovich, Yogesh Rathi, and Sudipto Dolui. "Spatially regularized compressed sensing for high angular resolution diffusion imaging". In: *IEEE transactions on medical imaging* 30.5 (2011), pp. 1100–1115.

[20] Markus Nilsson et al. "Resolution limit of cylinder diameter estimation by diffusion MRI: The impact of gradient waveform and orientation dispersion". In: *NMR in Biomedicine* (2017).

[21] Eleftheria Panagiotaki et al. "Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison". In: *Neuroimage* 59.3 (2012), pp. 2241–2254.

[22] BMCW Siow et al. "Axon radius estimation with oscillating gradient spin echo (OGSE) diffusion MRI". In: *Diffusion Fundamentals* 18.1 (2013), pp. 1–6.

[23] Jens Sjölund et al. "Constrained optimization of gradient waveforms for generalized diffusion encoding." In: *Journal of Magnetic Resonance* 261 (2015), pp. 157–168.

[24] E. O. Stejskal and J. E. Tanner. "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient". In: *The Journal of Chemical Physics* 42.1 (1965), pp. 288–292.

[25] Aaron Szafer, Jianhui Zhong, and John C Gore. "Theoretical model for water diffusion in tissues". In: *Magnetic resonance in medicine* 33.5 (1995), pp. 697–712.

[26] Daniel Topgaard. "Multidimensional diffusion MRI". In: *Journal of Magnetic Resonance* 275 (2017), pp. 98–113.