

Duality for maximum entropy diffusion MRI

Pierre Maréchal

ISAE, Université de Toulouse, 10 avenue Edouard Belin, 31055 Toulouse, France

Abstract. We derive an entropy model for diffusion MRI and show that Fenchel duality techniques make this model tractable. This generalizes a model proposed by Daniel Alexander in 2005, in which the displacement of particles is confined to a sphere. In order to better suit the physics of the diffusion process, we propose to relax this constraint. The Kullback-Leibler relative entropy is used to measure the discrepancy between the probability to be inferred and some reference measure. The obtained optimization problem is then studied using tools from partially finite convex programming. The solution can be computed via the unconstrained maximization of a smooth concave function, whose number of variable is merely (twice) the number of Fourier samples.

Keywords: Diffusion MRI, Kullback Leibler relative entropy, Fenchel duality, partially finite convex programming

PACS: 07.05.Pj, 87.85.Pq, 02.30.Zz, 02.60.Pn

INTRODUCTION

Diffusion MRI (dMRI) has developed in the recent past as a method for non-invasive imaging of the diffusion process of water in biological tissues. Since diffusion is affected by obstacles such as membranes and fibers, diffusion MRI opens the way to mapping fiber structures (e.g. neurons or muscle fibers).

The diffusion MRI data are samples of the Fourier transform of the probability P of particle displacements in each voxel of the volume to be imaged within a given time interval. From these data, one estimates either *diffusion tensors*, under the assumption that P is Gaussian (this is referred to as *Diffusion Tensor Imaging*) or *Orientation Diffusion Functions* (ODF). On denoting by p the density of P (which entails absolute continuity of P with respect to the Lebesgue measure), the ODF may be defined as

$$\psi(\mathbf{s}) := \int_{\mathbb{R}^3} p(r\mathbf{s}) r^2 dr, \quad \mathbf{s} \in S^2.$$

Here, S^2 denotes the unit sphere in \mathbb{R}^3 . Clearly, $\psi(\mathbf{s})$ is the total probability of displacement in the direction corresponding to \mathbf{s} , regardless of the displacement length.

Fibers can then be imaged from either diffusion tensors or orientation diffusion functions. The reader may refer to the Human Connectome Project site [12] to see examples of imaged fiber structures. Notice that the Fourier sampling is generally poor: the probability must be inferred from a few dozens of Fourier samples.

The idea of using the maximum entropy principle in this context appeared, to the best of our knowledge, in a paper published in 2005 by Daniel Alexander [1]. In the latter paper, the displacement is assumed to be confined to a sphere centered at the initial point of the particle.

Our aim in this paper is to write a more flexible model. More precisely, we wish

- (1) to manipulate more realistic P : it seems reasonable to relax the constraint that P should be supported on a sphere;
- (2) to preserve the possibility to account for moment constraints;
- (3) to be able to use of more general entropies: in particular, the Kullback-Leibler relative entropy will be used to measure the discrepancy between the probability to be inferred and some reference measure, which is a reasonable way to introduce physical prior knowledge.

We stress that we only focus here on modelling aspects, regardless of all numerical aspects. The latter are deferred to a later publication.

THE MODEL

If γ is a function from \mathbb{R}^n to \mathbb{R}^m , we denote by $E_P[\gamma]$ the mathematical expectation of γ , that is, the vector of \mathbb{R}^m defined by

$$E_P[\gamma] := \int \gamma(\mathbf{x}) dP(\mathbf{x}),$$

provided the integral is well-defined.

Experimental constraints

The MRI device provides noisy values of the complex integrals (Fourier samples)

$$z_j = \int_{\mathbb{R}^3} e^{-2i\pi\langle \mathbf{q}_j, \mathbf{x} \rangle} dP(\mathbf{x}), \quad j = 1, \dots, m. \quad (1)$$

In real notation, these constraints read

$$y_j = \int_{\mathbb{R}^3} \gamma_j(\mathbf{x}) dP(\mathbf{x}), \quad j = 1, \dots, 2m,$$

with

$$\gamma_j(\mathbf{x}) = \begin{cases} \cos(2\pi\langle \mathbf{q}_{[(j+1)/2]}, \mathbf{x} \rangle) & \text{if } j \text{ is even,} \\ \sin(2\pi\langle \mathbf{q}_{[(j+1)/2]}, \mathbf{x} \rangle) & \text{if } j \text{ is odd.} \end{cases}$$

Here, $[a]$ denotes the integer part of a . The data vector $\mathbf{y} := (y_1, \dots, y_{2m})^\top \in \mathbb{R}^{2m}$ may be written as $\mathbf{y} = E_P[\gamma]$, where

$$\gamma(\mathbf{x}) := (\gamma_1(\mathbf{x}), \dots, \gamma_{2m}(\mathbf{x}))^\top.$$

Moment constraints

Since P is a probability measure, the *normalization equation*

$$\int_{\mathbb{R}^3} dP(\mathbf{x}) = 1$$

must be satisfied. Notice that the latter equation is nothing but (1), with $\mathbf{q} = \mathbf{0}$ and $z = 1$.

Furthermore, from a physical viewpoint, it seems reasonable to assume that the random variable \mathbf{x} is centered or almost centered, which amounts to introducing constraints on the size of the first order moment

$$E_P[\mathbf{x}] = \int_{\mathbb{R}^3} \mathbf{x} dP(\mathbf{x}).$$

Formally, accounting for these moment constraints merely amounts to adding the three coordinate-functions of \mathbf{x} to the components of γ as well as three zero components to \mathbf{y} .

A general entropy model

In the DTI model, it is assumed that the probability P is Gaussian, and the problem is then that of identifying the covariance matrix, which is interpreted as a *diffusion tensor*. The main drawback of this approach is that it does not allow for imaging local structures such as fiber crossing.

Of course, the difficulty lies in the fact that P is a measure, while there are only finitely many constraints on P . The use of concepts from information theory is ubiquitous in such a situation. One infers a probability \bar{P} *via* the principle of minimum relative entropy, which consists in solving the following optimization problem:

$$\left| \begin{array}{l} \text{Minimize } \mathcal{H}(P||\nu) \\ \text{s.t. } 1 = E_P[1], \mathbf{y} = E_P[\gamma]. \end{array} \right.$$

Here, $\mathcal{H}(P||\nu)$ is the Kullback-Leibler relative entropy. The measure ν is a *reference* probability measure, and $\mathcal{H}(P||\nu)$ can be regarded as the amount of information to be introduced so as to replace ν by P . It is defined by

$$\mathcal{H}(P||\nu) := \begin{cases} \int u(\mathbf{x}) \ln u(\mathbf{x}) d\nu(\mathbf{x}) & \text{if } P \prec\prec \nu, \\ \infty & \text{otherwise,} \end{cases}$$

where u denotes the Radon-Nikodym derivative of P with respect to ν . Once a solution \bar{P} is inferred, the corresponding ODF may be computed.

Reformulating the problem

The above *minimum relative entropy* model inherently ignores probability measures that are not absolutely continuous with respect to ν . This leads us to reformulate constraints and functionals in terms of the Radon-Nikodym derivative u .

The moment and experimental constraints on P are naturally expressed on u by means of the following linear mappings:

$$\begin{aligned}\mathbb{A}u &= \mathbb{E}_P[\gamma] = \int_{\mathbb{R}^3} \gamma(\mathbf{x}) u(\mathbf{x}) d\nu(\mathbf{x}), \\ \mathbb{I}u &= \mathbb{E}_P[1] = \int_{\mathbb{R}^3} u(\mathbf{x}) d\nu(\mathbf{x}), \\ \mathbb{M}u &= \mathbb{E}_P[\mathbf{x}] = \int_{\mathbb{R}^3} \mathbf{x} u(\mathbf{x}) d\nu(\mathbf{x}).\end{aligned}$$

We stress here that \mathbb{M} is well-defined only on the subspace of $L^1(\mathbb{R}^3)$ of the functions having *well-defined finite* first order moments. This subspace is a particular example of what we call Köthe spaces, which are defined by some integrability condition against a family of measurable functions.

For the simplicity of the exposition, we now restrict attention to the formulation without moment constraints. The reader is invited to check that the model remains entirely tractable with constraints involving \mathbb{M} .

The above entropy problem can then be formulated as follows:

$$(\mathcal{P}) \quad \left| \begin{array}{l} \text{Minimize } K_\nu(u) \\ \text{s.t. } u \in L^1_\nu(\mathbb{R}^3), 1 = \mathbb{I}u, \mathbf{y} = \mathbb{A}u, \end{array} \right.$$

where

$$K_\nu(u) := \int k(u(\mathbf{x})) d\nu(\mathbf{x}), \quad \text{with } k(t) := \begin{cases} t \ln t & \text{si } t > 0 \\ 0 & \text{si } t = 0 \\ \infty & \text{si } t < 0. \end{cases}$$

Minimizing $K_\nu(u)$ corresponds to the desire to introduce a reasonable amount of information, under the form of a prior probability measure, in order to compensate for the lack of experimental data. The reference measure ν may be chosen as an isotropic Gaussian measure, the one we would have in an isotropic medium with no fiber.

Of course, since the data are noisy, the constraint should be relaxed, and the optimization problem to be solved should take the following form:

$$(\mathcal{P}_\alpha) \quad \left| \begin{array}{l} \text{Minimize } K_\nu(u) + \frac{1}{2\alpha} \|\mathbf{y} - \mathbb{A}u\|^2 \\ \text{s.t. } u \in L^1_\nu(\mathbb{R}^3), 1 = \mathbb{I}u. \end{array} \right.$$

Let g and g_\circ be the functions respectively defined on \mathbb{R}^{2m} and \mathbb{R}^{1+2m} by

$$g(\eta) = -\frac{1}{2\alpha} \|\eta - \mathbf{y}\|^2 \quad \text{and} \quad g_\circ(\eta_\circ, \eta) = g(\eta) - \delta(\eta_\circ | \{1\}),$$

where $\eta \mapsto \delta(\eta|\{1\})$ denotes the *indicator function* of the singleton $\{1\}$. Recall that the indicator function of a set $S \subset \mathbb{R}^d$ is defined (on \mathbb{R}^d) by

$$\delta(\mathbf{x}|S) = \begin{cases} 0 & \text{if } \mathbf{x} \in S, \\ \infty & \text{otherwise.} \end{cases}$$

The functions g and g_\circ are *closed proper concave*. Finally, Problem (\mathcal{P}_α) can then be written as

$$(\mathcal{P}_\alpha) \quad \left| \begin{array}{l} \text{Minimize } K_\nu(u) - g_\circ(\mathbb{A}_\circ u) \\ \text{s.t. } u \in L_\nu^1(\mathbb{R}^3), \end{array} \right.$$

in which $\mathbb{A}_\circ u = (\mathbb{I}u; \mathbb{A}u)$.

DUALITY

In order to deal with Problems such as (\mathcal{P}_α) , we use a dual strategy. The next subsection provides general results on partially finite convex programming with integral functionals. The particular case of Problem (\mathcal{P}_α) will be considered subsequently. The reader is assumed to be familiar with the language and notation of Convex Analysis. Our reference books in Convex Analysis are [8, 4]. Results on *convex integral functionals* can be found in [7, 9, 10, 11]. The reader interested in details may also consult [6], where a digest of the following results is provided.

Fenchel duality and primal-dual relationship

The following theorem [2, 5] is a partially finite version of Fenchel's duality theorem.

Theorem 1. *Suppose we are given*

- L and L^* , two vector spaces, algebraically paired by $\langle \cdot, \cdot \rangle$;
- $A: L \rightarrow \mathbb{R}^d$, a linear mapping;
- $A^*: \mathbb{R}^d \rightarrow L^*$ its (formal) adjoint;
- $H: L \rightarrow (-\infty, \infty]$, a proper convex functional;
- $H^*: L^* \rightarrow (-\infty, \infty]$, its convex conjugate;
- $g: \mathbb{R}^d \rightarrow [-\infty, \infty)$, a proper concave function;
- $g_\star: \mathbb{R}^d \rightarrow [-\infty, \infty)$, its concave conjugate.

If the constraint qualification condition

$$(CQ) \quad \text{ri}(A \text{ dom} H) \cap \text{ri}(\text{dom } g) \neq \emptyset$$

is satisfied, then

$$\eta := \inf_{u \in L} \{H(u) - g(Au)\} = \max_{\lambda \in \mathbb{R}^d} \{g_\star(\lambda) - H^*(A^*\lambda)\}.$$

The function $D := g_\star - H^ \circ A^*$ appearing in the right hand side is referred to as the dual function.*

This theorem is a powerful tool for convex programming in infinite dimension whenever it is possible to compute the conjugate functions H^* and g_* . We shall get back later to conjugacy issues. A key result for primal-dual relationship is the following.

Theorem 2. *With the notation and assumptions of the previous theorem, assume in addition that*

$$(CQ^*) \quad \text{ri dom } g_* \cap \text{ri dom}(H^* \circ A^*) \neq \emptyset$$

and that

- (a) $H^{**} = H$ and $g_{**} = g$;
- (b) there exist $\bar{\lambda}$, a dual solution, and \bar{u} in $\partial H^*(A^*\bar{\lambda})$ such that $H^* \circ A^*$ admits $A\bar{u}$ as gradient at $\bar{\lambda}$.

Then \bar{u} is a primal solution.

We now establish explicit primal-dual relationship in the case of entropy-like functionals. By entropy-like functionals, we mean convex integral functionals of the form

$$u \mapsto H(u) := \int h(u(x), x) \, d\nu(x).$$

The function K_ν is a standard example of such functionals, whose integrand shows no dependence on the second argument.

Theorem 3. *With the notation and assumptions of Theorem 1 assume in addition that $\text{dom } D$ has nonempty interior; that H is an integral functional of integrand h (see the appendix) such that conjugacy through the integral sign is permitted (which involve both the functional and the space on which it is considered). Assume that, as in Theorem 2, $H^{**} = H$ and $g_{**} = g$. Assume finally that the conjugate integrand h^* is differentiable over \mathbb{R} , and that there exists some dual-optimal vector $\bar{\lambda}$ in $\text{int dom } D$. If*

$$\bar{u}(\mathbf{x}) := h^{*'}([A^*\bar{\lambda}](\mathbf{x}), \mathbf{x}) \in L,$$

then \bar{u} is a primal solution.

Dealing with our particular entropy problem

On the one hand, we have:

$$\begin{aligned} g_*(\lambda) &:= \inf \left\{ \langle \lambda, \eta \rangle + \frac{1}{2\alpha} \|\eta - \mathbf{y}\|^2 \mid \eta \in \mathbb{R}^{2m} \right\} \\ &= \langle \lambda, \mathbf{y} \rangle + \inf \left\{ \langle \lambda, \eta' \rangle + \frac{1}{2\alpha} \|\eta'\|^2 \mid \eta' \in \mathbb{R}^{2m} \right\}, \end{aligned}$$

using the change of variable $\eta' = \eta - \mathbf{y}$. On writing that the gradient of the function in the infimum should be zero at the minimum, we find that

$$g_*(\lambda) = \langle \lambda, \mathbf{y} \rangle - \alpha \|\lambda\|^2 + \frac{\alpha}{2} \|\lambda\|^2 = \langle \lambda, \mathbf{y} \rangle - \frac{\alpha}{2} \|\lambda\|^2.$$

Next,

$$(g_\circ)_*(\lambda_\circ, \lambda) = \inf \{ \eta_\circ \lambda_\circ + \langle \lambda, \eta \rangle - g_\circ(\eta_\circ, \eta) \mid (\eta_\circ, \eta) \in \mathbb{R}^{1+2m} \} = \lambda_\circ + g_*(\lambda).$$

On the other hand, the functional K_ν is an integral functional, whose integrand is the previously defined function k . It is well-defined on $L^1_\nu(\mathbb{R}^3)$, therefore on any of its subspaces, in particular the subspace of L^1 -functions having well-defined finite first order moment. An easy computation shows that

$$k^*(\tau) = \exp(\tau - 1)$$

for every $\tau \in \mathbb{R}$. Consequently, for every $w \in L^\infty_\nu(\mathbb{R}^3)$,

$$K_\nu^*(w) = \int_{\mathbb{R}^3} \exp(w(\mathbf{x}) - 1) d\nu(\mathbf{x}).$$

Finally, it is easy to check that the adjoint of \mathbb{A}_\circ is the mapping

$$\begin{aligned} \mathbb{A}_\circ^*: \quad \mathbb{R}^{1+2m} &\longrightarrow L^\infty_\nu \\ (\lambda_\circ, \lambda) &\longmapsto \lambda_\circ + \langle \lambda, \gamma(\mathbf{x}) \rangle. \end{aligned}$$

The dual problem associated to (\mathcal{P}_α) then reads:

$$(\mathcal{D}_\alpha) \quad \left| \begin{array}{l} \text{Maximize } D(\lambda_\circ, \lambda) \\ \text{s.t. } (\lambda_\circ, \lambda) \in \mathbb{R}^{1+2m}. \end{array} \right.$$

where

$$D(\lambda_\circ, \lambda) := \lambda_\circ + \langle \lambda, \mathbf{y} \rangle - \frac{\alpha}{2} \|\lambda\|^2 - \exp(\lambda_\circ - 1) \int \exp\langle \lambda, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}).$$

The function D is of course concave, with effective domain \mathbb{R}^{1+2m} . It is also differentiable on \mathbb{R}^{1+2m} . The stationaty points of D satisfy the system

$$(S) \quad \begin{cases} 0 &= 1 - \exp(\bar{\lambda}_\circ - 1) \int_{\mathbb{R}^3} \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}), \\ \mathbf{0} &= \mathbf{y} - \alpha \lambda - \exp(\bar{\lambda}_\circ - 1) \int_{\mathbb{R}^3} \gamma(\mathbf{x}) \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}), \end{cases}$$

which reduces to

$$(\tilde{S}) \quad \mathbf{0} = \mathbf{y} - \alpha \lambda - \left(\int_{\mathbb{R}^3} \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}) \right)^{-1} \int_{\mathbb{R}^3} \gamma(\mathbf{x}) \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}).$$

Observe that the latter system is also the optimality system of the problem

$$(\tilde{\mathcal{D}}) \quad \left| \begin{array}{l} \text{Maximize } \tilde{D}(\lambda) := \langle \lambda, \mathbf{y} \rangle - \frac{\alpha}{2} \|\lambda\|^2 - \ln \int \exp\langle \lambda, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}) \\ \text{s.t. } \lambda \in \mathbb{R}^{2m}. \end{array} \right.$$

Proposition 1. *The above defined function \tilde{D} is concave and smooth.*

One may check that (CQ) is satisfied as soon as $\{\mathbf{q}_j\}_{j=1,\dots,m}$ spans \mathbb{R}^3 . The function $h^*(\tau) = \exp(\tau - 1)$ obviously meets the requirements of Theorem 3 and, provided that we can solve the dual problem, the optimal density satisfies

$$\bar{u}(\mathbf{x}) = \exp[\bar{\lambda}_o - 1 + \langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle] = \left(\int \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}) \right)^{-1} \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle,$$

where $\bar{\lambda}$ maximizes the function \tilde{D} .

The above developments can be summarized as follows:

- (1) Find $\bar{\lambda}$ which maximizes $\tilde{D}(\lambda)$;
- (2) Compute

$$\exp(\bar{\lambda}_o - 1) = \left(\int \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle d\nu(\mathbf{x}) \right)^{-1};$$

- (3) Compute the ODF from the density

$$\bar{u}(\mathbf{x}) = \exp(\bar{\lambda}_o - 1) \exp\langle \bar{\lambda}, \gamma(\mathbf{x}) \rangle.$$

The optimization problem in step (1) Any standard optimization routine for is unconstrained and smooth. The only difficulty lies in that the evaluations of \tilde{D} and its gradient require some numerical integration.

Finally, it is interesting to notice that The optimal density \bar{u} is searched for in a smooth manifold of dimension $2m$ in $L^1_{\nu}(\mathbb{R}^3)$, the coordinates of which being the dual variable λ .

REFERENCES

1. D.C. ALEXANDER, *Maximum Entropy Spherical Deconvolution for Diffusion MRI*, in G.E. CHRISTENSEN & M. SONKA (Eds.), *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, **3565**, pp. 76-87, 2005.
2. J.M. BORWEIN and A.S. LEWIS, *Partially finite convex programming, Part I: quasi relative interiors and duality theory*, *Mathematical Programming*, **57**, pp. 15-48, 1992.
3. S. KULLBACK and R.A. LEIBLER, *On information and sufficiency*, *Annals of Mathematical Statistics*, **22** (1), pp. 79-86, 1951.
4. J.-B. HIRIART-URRUTY and C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms, I and II*, Springer-Verlag, Berlin, 1993.
5. P. MARÉCHAL, *Sur la régularisation des problèmes mal posés*, PhD. thesis, Université Paul Sabatier (Toulouse III), 1997.
6. P. MARÉCHAL, J.J. YE and J. ZHOU, *K-optimal design via semi-definite programming and entropy optimization*, to appear in *Mathematics of Operations Research*.
7. R.T. ROCKAFELLAR, *Integrals which are convex functionals*, *Pacific Journal of Mathematics*, **24**(3) pp. 525-539, 1968.
8. R.T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, N. J., 1970.
9. R.T. ROCKAFELLAR, *Convex integral functionals and duality*, In E. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, Academic Press, 1971.
10. R.T. ROCKAFELLAR, *Integrals which are convex functionals, 2*, *Pacific Journal of Mathematics*, **39**(2) pp. 439-469, 1971.
11. R.T. ROCKAFELLAR, *Integral functionals, normal integrands and measurable selections*, in Lucien Waelbroeck, editor, *Nonlinear Operators and the Calculus of Variations*, Springer-Verlag series Lecture Notes in Math., 1976.
12. <http://www.humanconnectomeproject.org/gallery/>