Total Variation Minimization for Measure-Valued Images with Diffusion Spectrum Imaging as Motivation

Davide La Torre[1,2], Franklin Mendivil[3], Oleg Michailovich[4], and Edward R. Vrscay[5(✉)]

[1] Department of Economics, Management, and Quantitative Methods, University of Milan, Milan, Italy
davide.latorre@unimi.it

[2] Department of Applied Mathematics and Sciences, Khalifa University, Abu Dhabi, UAE
davide.latorre@kustar.ac.ae

[3] Department of Mathematics and Statistics, Acadia University, Wolfville, NS, Canada
franklin.mendivil@acadiau.ca

[4] Deparment of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada
olegm@uwaterloo.ca

[5] Deparment of Applied Mathematics, University of Waterloo, Waterloo, ON, Canada
ervrscay@uwaterloo.ca

Abstract. In this paper, we present a notion of total variation for measure-valued images. Our motivation is Diffusion Spectrum Imaging (DSI) in which the diffusion at each voxel is characterized by a probability density function. We introduce a total variation denoising problem for measure-valued images. In the one-dimensional case, this problem (which involves the Monge-Kantorovich metric for measures) can be solved using cumulative distribution functions. In higher dimensions, more computationally expensive methods must be employed.

1 Introduction

Over the last few decades, Diffusion Magnetic Resonance Imaging (dMRI) has developed into an established tool of diagnostic medical imaging, which is nowadays used in a broad spectrum of clinical applications, including the diagnosis of oncological and neurodegenerative disorders [6]. In application to brain imaging, the advent of dMRI has made it possible to delineate the anatomical structure of white matter, thereby opening the possibility of quantitative exploration of *in vivo* connectivity of the brain by means of fibre tractography [1,3,7]. Naturally, the reliability and diagnostic value of such analysis depends on the quality of acquired diffusion data, which tend to be affected by various measurement artifacts, including noise [12].

A. Campilho and F. Karray (Eds.): ICIAR 2016, LNCS 9730, pp. 131–137, 2016.
DOI: 10.1007/978-3-319-41501-7_15

In a previous paper [10], we showed that *function-valued mappings* (FVM) provide a natural setting for a particular instance of dMRI known as *high angular resolution diffusion imaging* (HARDI), which excels in delineation of the orientational structure of crossing fibre bundles in the white matter of the brain. At each position/pixel $x \in \mathbb{R}^3$ and orientation $s \in \mathbb{S}^2 := \{u \in \mathbb{R}^3 \mid \|u\|_2 = 1\}$, HARDI provides a measurement which quantifies the apparent diffusivity of water molecules along the direction defined by s. Therefore, a HARDI signal u measured at position x, i.e., $u(x)$, can be assumed to be a square-integrable function supported over $\mathbb{S}^2$. In other words, the HARDI signal is considered to be an FVM, where $u : \mathbb{R}^3 \to \mathbb{L}_2(\mathbb{S}^2)$.

While useful for inferring the orientational characteristics of diffusion *in vivo*, the applicability of HARDI to estimation of microstructural characteristics of biological tissues remains limited. At the same time, much richer information can be obtained using a different variation of dMRI, known as *diffusion spectral imaging* (DSI) [15]. In fact, the latter allows estimation of the *ensemble averaged diffusion propagator* (EADP) $P(r)$–a probability measure that quantifies the likelihood of a water molecule to undergo displacement by $r \in \mathbb{R}^3$ in a given experimental time [13]. Moreover, considering the straightforward relation between DSI data and EADP through Fourier transformation, it is reasonable to assume that the datum of DSI measurements is equivalent to the datum of the estimates of EADP, which will be referred below to as *propagator signals*.

One could, in principle, also consider a propagator signal as a function-valued mapping. However, given the fact that the signal is, in fact, a probability density function, it may be beneficial to analyze and process it using mathematical methods that have been adopted for measures. In this paper, therefore, we consider propagator signals as *measure-valued mappings* (MVM): At each position/pixel/voxel x is associated a probability measure μ_x supported on $\mathbb{R}^3$, to be denoted as $\mu(x)$. Subsequently, to counteract the effect of both measurement and estimation noises, we then propose a novel *total variation* (TV) denoising method for MVMs. Of course, this requires a definition of total variation for measure-valued mappings which, in itself, is a nontrivial mathematical problem. Space limitations allow us only to outline our method below.

2 Measure-Valued Images

In what follows $X = [0,1]^n$, $n = 1, 2, 3$ will denote the "base space," i.e., the support of the images. $\mathbb{R}_g \subset \mathbb{R}^m_+$, $m = 1, 2, 3$, will denote a compact set of values that our image can assume at any $x \in X$ $\mathbb{B}$ will denote the Borel σ algebra on $\mathbb{R}_g$ and dx Lebesgue measure on X. Let $\mathscr{M}$ denote the set of all Borel probability measures on $\mathbb{R}_g$ and d_{MK} the Monge-Kantorovich metric (see [8]) on this set. That is, for $\alpha, \beta \in \mathscr{M}$,

$$d_{MK}(\alpha, \beta) = \sup_{\phi \in \mathrm{Lip}_1(\mathbb{R}_g)} \int_{\mathbb{R}_g} \phi(t)\; d(\alpha - \beta)(t), \tag{1}$$

where, as usual, $\text{Lip}_1(\mathbb{R}_g) = \{f : \mathbb{R}_g \to \mathbb{R} : |f(x) - f(y)| \leq \|x - y\|\}$. We can now define the following space of measure-valued images,

$$Y = \{\mu : X \to \mathscr{M}, \mu \text{ is Lebesgue integrable}\}, \tag{2}$$

with the following metric,

$$d_Y(\mu, \nu) = \int_X d_{MK}(\mu(x), \nu(x))dx. \tag{3}$$

Note that d_Y is well defined since μ and ν are Lebesgue-integrable functions, d_{MK} is continuous and X is compact, which implies that the function $\xi(x) = d_{MK}(\mu(x), \nu(x))$ is integrable on X. It can be shown that the space (Y, d_Y) is complete (see [9]).

3 Total Variation

In image analysis, the notion of total variation or total variation regularization has applications in noise removal. The basic idea relies on the fact that signals with spurious detail have high total variation or, more mathematically, the integral of the absolute gradient of the signal is high. It is well known that the process of reducing the total variation of the signal removes unwanted detail whilst preserving important details such as edges (see [14]). The total variation (TV) of a differentiable greyscale image $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is defined as follows

$$\|f\|_{TV} = \int_X \|\nabla f(x)\|_2 \, dx, \tag{4}$$

that is the integral of the $\|\cdot\|_2$ norm of the gradient. Other definitions of total variation are available in the literature–the reader is referred to [5] for an overview of many of the most recent ones.

In this section we introduce a notion of total variation for measure-valued images which, as usual, will involve derivatives. We define the *total variation* of $\mu \in Y$ to be

$$\|\mu\|_{TV} = \int_X \|D\mu\|_2 \, dx = \int_X \left(\sum_{i=1}^{n} |D_i\mu(x)|^2 \right)^{1/2} dx \, . \tag{5}$$

Here,

$$|D_i\mu(x)| := \sup_{\phi_i \in \text{Lip}_1(\mathbb{R}_g)} \limsup_{h_i \to 0^+} \frac{1}{h_i} \int_{\mathbb{R}_g} \phi_i(t) \; d(\mu(x + \hat{e}_i h_i) - \mu(x)) \, , \quad 1 \leq i \leq n,$$

are the analogues of the magnitudes of the directional derivative of μ at the point $x \in X$ in the directions of $\hat{e}_i$, the standard orthonormal basis vectors in $\mathbb{R}^n$. The TV norm in (5) may be used to define a *total variation distance* between μ and ν in Y as $\|\mu - \nu\|_{TV}$. A few additional comments are in order.

First, note that by taking the supremum over $\phi_i \in \mathrm{Lip}_1(\mathbb{R}_g)$, we are using a Monge-Kantorovich-type norm on measures (see [8]) to compute the magnitude of the directional derivative. By taking the supremum over $\phi_i \in L^\infty(\mathbb{R}_g)$, the magnitude of the directional derivative is computed using the *variation norm* on measures.

Next, let $f : X \to \mathbb{R}$ be a differentiable function and $\mu(x) = \delta_{f(x)}$. Then for a direction $\hat{d} \in \mathbb{R}^n$ with $\|\hat{d}\| = 1$ and any $\phi \in \mathrm{Lip}_1(\mathbb{R}_g)$,

$$\frac{1}{h}\int_{R_g} \phi(t)\ d(\mu(x+\hat{d}h) - \mu(x)) = \frac{\phi(f(x+\hat{d}h)) - \phi(f(x))}{h} \to \phi'(f(x))\ \nabla f(x)\cdot \hat{d}. \tag{6}$$

Taking the supremum over $\phi \in \mathrm{Lip}_1(\mathbb{R}_g)$, we obtain $|D_{\hat{d}}\mu(x)| = |\nabla f(x) \cdot \hat{d}|$ so that

$$\|\mu\|_{TV} = \int_X \left(\sum_i |\nabla f(x)\cdot \hat{e}_i|^2\right)^{1/2} dx = \int_X \|\nabla f(x)\|_2\ dx, \tag{7}$$

which agrees with the classical definition of TV for functions in Eq. (4). Our definition of total variation in (5) may therefore be viewed as a true extension of the classical definition to measure-valued mappings.

Finally, if $\mu(x)$ has a density $\rho_\mu(x,\cdot)$ for each $x \in X$, then a standard calculation shows that $D_{\hat{e}_i}\mu$ is a signed measure with density $\frac{\partial \rho_\mu}{\partial x_i}$. This is important for models where one fits data with a parametric form of $\mu(x)$ for each x.

4 The Inverse Problem and Total Variation Minimization

The total variation denoising problem for measure-valued images can now be formulated as follows: Given a noisy image $\tilde{\mu}$ (the "observed data") we seek a solution to the following optimization problem,

$$\min_{\mu\in Y} d_Y(\tilde{\mu},\mu) + \lambda\|\nu\|_{TV}, \tag{8}$$

where λ is a trade-off or regularization parameter.

We now show how to solve the optimization problem (8) practically in the special case that the pixel space $X = [0,1]$ and the greyspace $\mathbb{R}_g = [0,1]$. (Admittedly, this case is not very interesting from a practical perspective, but it does provide insight into some of the mathematical aspects of this problem. A discussion of the higher-dimensional situation $\mathbb{R}_g \subseteq \mathbb{R}^n$ for $n > 1$ would be much more complicated, requiring more space and details.) In this special case one can rely on the following characterization of the Monge-Kantorovich distance in terms of cumulative distribution functions. For more on the practical issue of computing the Monge-Kantorovich distance, see [11].

Theorem 1 [2,4]**.** *Let μ and ν two probability measures defined on $\mathbb{R}$ and $F_\mu(x) = \mu((-\infty,x])$ and $F_\nu(x) = \nu((-\infty,x])$ be the two corresponding cumulative distribution functions. Then*

$$d_{MK}(\mu,\nu) = \|F_\nu - F_\mu\|_1. \tag{9}$$

We now use this theorem to approximate the total variation integral $\|\mu\|_{TV}$ in (5) as follows: For a small, fixed value of $h \in \mathbb{R}$,

$$\|\mu\|_{TV} = \int_0^1 |D\mu(x)|\, dx \approx \int_0^1 \frac{\|F_{\mu(x+h)} - F_{\mu(x)}\|_1}{h}\, dx. \tag{10}$$

The total variation minimization problem in (8) can now be reformulated as follows,

$$\min_{\mu \in Y} \int_0^1 \left[\|F_{\tilde{\mu}(x)} - F_{\mu(x)}\|_1 + \frac{\lambda}{h}\|F_{\mu(x+h)} - F_{\mu(x)}\|_1 \right] dx. \tag{11}$$

It is worth noticing that the space Y is convex. It is also straightforward to show that the map $T : \mu \in Y \to \mathbb{R}_+$ (or $T : F \in Y \to \mathbb{R}_+$), defined as

$$TF := \int_0^1 \int_{\mathbb{R}_g} \left[|\tilde{F}(x,t) - F(x,t)| + \frac{\lambda}{h}|F(x+h,t) - F(x,t)| \right] dt\, dx, \tag{12}$$

is convex. Equation (11) is therefore a convex minimization problem so that classical algorithms from convex programming can used to determine an optimal solution. Note also that if $\mu(x)$ has density $\rho_\mu(x,\cdot)$ for each $x \in X$, then from Theorem 1 above,

$$\|\mu\|_{TV} = \int_0^1 \int_{T \in \mathbb{R}_g} \left| \int_{-\infty}^{T} \frac{\partial \rho_\mu}{\partial x}(z,t)\; dt \right| dT\, dz. \tag{13}$$

One way to solve Eq. (11) practically is to employ a finite subset of elements, $F_i \in Y$, $i = 1...n$. We now solve the minimization problem in (11) over the convex subset spanned by the F_i. Any function in this subset can be written as a convex combination of the F_i, that is,

$$F(x,t) = \sum_{i=1}^{n} \alpha_i F_i(x,t), \tag{14}$$

where $\alpha_i \in [0,1]$, $\sum_{i=1}^{n} \alpha_i = 1$. The minimization problem in Eq. (11) can then be approximated by

$$\min \int_0^1 \int_{\mathbb{R}_g} \left(\left| \tilde{F}(x,t) - \sum_{i=1}^{n} \alpha_i F_i(x,t) \right| + \frac{\lambda}{h} \left| \sum_{i=1}^{n} \alpha_i (F_i(x+h,t) - F_i(x,t)) \right| \right) dt\, dx. \tag{15}$$

An upper bound for the optimal value can be found by noticing that

$$\begin{aligned} &\left| \tilde{F}(x,t) - \sum_{i=1}^{n} \alpha_i F_i(x,t) \right| + \frac{\lambda}{h} \left| \sum_{i=1}^{n} \alpha_i (F_i(x+h,t) - F_i(x,t)) \right| \\ &= \left| \sum_{i=1}^{n} \alpha_i \tilde{F}(x,t) - \sum_{i=1}^{n} \alpha_i F_i(x,t) \right| + \frac{\lambda}{h} \left| \sum_{i=1}^{n} \alpha_i (F_i(x+h,t) - F_i(x,t)) \right| \\ &\leq \sum_{i=1}^{n} \alpha_i \left(\left| \tilde{F}(x,t) - F_i(x,t) \right| + \frac{\lambda}{h} \left| F_i(x+h,t) - F_i(x,t) \right| \right) \end{aligned} \tag{16}$$

and then solving the linear programming model,

$$\min \sum_{i=1}^{n} \alpha_i \int_0^1 \int_{\mathbb{R}_g} \left(\left| \tilde{F}(x,t) - F_i(x,t) \right| + \frac{\lambda}{h} \left| F_i(x+h,t) - F_i(x,t) \right| \right), \tag{17}$$

subject to the constraint,

$$\sum_{i=1}^{n} \alpha_i = 1, \quad \alpha_i \in [0,1]. \tag{18}$$

A particularly simple choice has each F_i of the form

$$F(x,t) = \begin{cases} 0 & \text{if } t < \psi(x), \\ 1 & \text{if } t \geq \psi(x), \end{cases}$$

for some function $\psi : X \to \mathbb{R}_g$ (this corresponds to a point mass at $\psi(x)$ for each x). For instance, if $\{0 = t_0 < t_1 < \cdots < t_n = 1\}$ is a partition of $\mathbb{R}_g$, then $\psi_i(x) = t_i$ allows an estimation of the empirical CDF for $\tilde{F}$, though this approximation will be constant in x. Polynomial ψ give non-constant (in x) approximations and increasing the degree allows for arbitrarily close approximations.

The methods of this section can be generalized to higher dimensional X and $\mathbb{R}_g$, though $\dim(\mathbb{R}_g) > 1$ is more computationaly expensive. For direction fields in 2D, [4] provides an efficient computation of the Monge-Kantorovich distance on the circle. Research is ongoing for efficient computational algorithms for $\mathbb{S}^2$ and $\mathbb{R}^3$. Another alternative is to replace the Monge-Kantorovich metric on measures with another metric.

Acknowledgements. This work has also been supported in part by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (NSERC): FM (238549-2012) and ERV (Grant No. 106270-2012).

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**(1), 259–267 (1994)
2. Brandt, J., Cabrelli, C., Molter, U.: An algorithm for the computation of the Hutchinson distance. Inform. Process. Lett. **40**, 113–117 (1991)
3. Bihan, D.L., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M.: MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurological disorders. Radiology **161**, 401–407 (1986)
4. Cabrelli, C.A., Molter, U.M.: The Kantorovich metric for probability measures on the circle. J. Comput. Appl. Math. **57**, 345–361 (1995)
5. Goldluecke, B., Strekalovskiy, E., Cremers, D.: The natural vectorial total variation which arises from geometric measure theory. SIAM J. Imaging Sci. **5**(2), 537–563 (2012)
6. Johansen-Berg, H., Behrens, T.E.J.: Diffusion MRI: From Quantitative Measurements to In-Vivo Neuroanatomy, 1st edn. Academic, New York (2009)

7. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Neural tractography using an unscented kalman filter. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 126–138. Springer, Heidelberg (2009)
8. Kunze, H., La Torre, D., Mendivil, F., Vrscay, E.R.: Fractal-Based Methods in Analysis. Springer, Heidelberg (2012)
9. La Torre, D., Vrscay, E.R., Ebrahimi, A., Barnsley, M.: Measure-valued images, associated fractal transforms, and the affine self-similarity of images. SIAM J. Imaging Sci. **2**(2), 470–507 (2009)
10. Michailovich, O., La Torre, D., Vrscay, E.R.: Function-valued mappings, total variation and compressed sensing for diffusion MRI. In: Campilho, A., Kamel, M. (eds.) ICIAR 2012, Part II. LNCS, vol. 7325, pp. 286–295. Springer, Heidelberg (2012)
11. Mendivil, F.: Computing the Monge-Kantorovich distance. Comp. Appl. Math. (2016, to appear)
12. Brion, V., Poupon, C., Ri, O., Aja-Fernandez, S., Tristan-Vega, A., Mangin, J.F., Le Bihan, D., Poupon, F.: Noise correction for HARDI and HYDI data obtained with multi-channel coils and sum of squares reconstruction: An anisotropic extension of the LMMSE. Magn. Reson. Imaging **31**, 1360–1371 (2013)
13. Callahan, P.T.: Principles of Nuclear Magnetic Resonance Microscopy. Clarendon Press, Oxford (1991)
14. Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. Inverse Prob. **19**, 165–187 (2003)
15. Wedeen, V.J., Hagmann, P., Tseng, W.-Y., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. **54**, 1377–1386 (2005)