
International Journal of Computer Vision 60(3), 225–240, 2004

Optimal Mass Transport for Registration and Warping

STEVEN HAKER
Surgical Planning Laboratory Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
haker@bwh.harvard.edu

LEI ZHU AND ALLEN TANNENBAUM
Departments of Electrical, Computer and Biomedical Engineering, 777 Atlantic Drive, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA
zlzl@ece.gatech.edu
tannenba@ece.gatech.edu

SIGURD ANGENENT
Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA
angenent@math.wisc.edu

Received September 11, 2002; Revised August 26, 2003; Accepted March 11, 2004

Abstract. Image registration is the process of establishing a common geometric reference frame between two or more image data sets possibly taken at different times. In this paper we present a method for computing elastic registration and warping maps based on the Monge–Kantorovich theory of optimal mass transport. This mass transport method has a number of important characteristics. First, it is parameter free. Moreover, it utilizes all of the grayscale data in both images, places the two images on equal footing and is symmetrical: the optimal mapping from image A to image B being the inverse of the optimal mapping from B to A. The method does not require that landmarks be specified, and the minimizer of the distance functional involved is unique; there are no other local minimizers. Finally, optimal transport naturally takes into account changes in density that result from changes in area or volume. Although the optimal transport method is certainly not appropriate for all registration and warping problems, this mass preservation property makes the Monge–Kantorovich approach quite useful for an interesting class of warping problems, as we show in this paper. Our method for finding the registration mapping is based on a partial differential equation approach to the minimization of the L^2 Kantorovich–Wasserstein or "Earth Mover's Distance" under a mass preservation constraint. We show how this approach leads to practical algorithms, and demonstrate our method with a number of examples, including those from the medical field. We also extend this method to take into account changes in intensity, and show that it is well suited for applications such as image morphing.

Keywords: elastic registration, image warping, optimal transport, mass-preservation, gradient flows

1. Introduction

1.1. Image Registration

Image registration and warping are key challenges that must be addressed for a number of practical imaging problems. Registration is the process of establishing a common geometric reference frame between two or more image data sets. In the context of medical imaging, registration allows for the incorporation of preoperative image information in a surgical setting to improve image-guided surgery and therapy. It also aids in diagnosis by allowing the concurrent use of information from multiple data sets, possibly taken at different times using different modalities and patient positions.

Registration processes proceed in several steps. Typically, a measure of similarity between the data sets is established, so that one can quantify how close an image is from another after transformations are applied. Such a measure may include the similarity between pixel intensity values, as well as the proximity of predefined image features such as implanted fiducials, anatomical landmarks, surface contours, and ridge lines. Next, the transformation that maximizes the similarity between the transformed images is found. Often this transformation is given as the solution of an optimization problem where the transformations to be considered are constrained to be of a predetermined class. Finally, once an optimal transformation is obtained, it is used to fuse the image data sets.

Registration has an extensive literature devoted to it with numerous approaches ranging from statistical to computational fluid dynamics to various types of warping methodologies. See Toga (1999) for a number of recent papers on the subject as well as an extensive set of references. We only review some of the more relevant ones here. Our method is in the class of warping strategies based on continuum and fluid mechanics, in which one tries to use properties of elastic materials to determine the deformation. One defines a (typically quadratic) cost functional that penalizes the mismatch between the deforming template and target (Christensen et al., 1993, 1996; Miller et al.; Bro-Nielsen and Gramkow, 1996; Thirion, 1995). In this sense, our method is closest to the registration philosophy of these works. In fact, the optimal warping map of the L^2 Monge-Kantorovich may be regarded as the velocity vector field which minimizes a standard energy integral subject to the Euler continuity (mass preservation) equation (Benamou and Brenier, 2000). For the exact statement, see Section 3.6 below. In particular, in the fluid mechanics framework, this means that the optimal Monge–Kantorovich solution is given as a *potential flow*.

1.2. Optimal Transport

The method described in this paper is designed for elastic registration, and is based on an optimization problem built around the L^2 Kantorovich–Wasserstein distance taken as the similarity measure. The constraint which we put on the transformations considered is that they obey a mass preservation property. Thus, we will be matching *mass densities* in this method, which may be thought of as weighted areas in 2D or weighted volumes in 3D. Optimal mass transport problems of this sort were first formulated by Gaspar Monge in 1781, and concerned finding the optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one site to another. Thus the Kantorovich–Wasserstein distance is also commonly referred to as the "Earth Mover's Distance." The problem was given a modern formulation in the work of Kantorovich (1948), and so is now known as the *Monge–Kantorovich problem.*

Our interest in the Monge–Kantorovich problem originally arose from our work in medical applications. The problem occurs in medical imaging, for example, in functional MR, where one may want to compare the degree of activity in various features deforming over time, and obtain a corresponding elastic registration map. A special case of this problem occurs in any application where volume or area preserving mappings are considered. For example, as we will show in Section 4, our method provides a means to obtain regular area-preserving surface diffeomorphisms. We have found this technique useful for applications such as brain surface and colon surface flattening (Angenent et al., 1999b; Haker et al., 2000). However, our optimal transport method may not be suitable under some circumstances when the mass preservation assumption is likely to be invalid, such as the matching of two different perspective projections of a spatial object, or the registration of MRI images to PET with image intensity treated as a mass density.

Optimal transport methods have appeared in econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimization, expert systems, and meteorology (Rachev and Rüschendorf, 1998). They also naturally fit into certain problems in

computer vision. In particular, for the general tracking problem, a robust and reliable object and shape recognition system is of major importance. A key way to carry this out is via *template matching*, which is the matching of some object to another within a given catalogue of objects. Typically, the match will not be exact and hence some *shape metric* is necessary to measure the "goodness of fit" or similarity between objects (Haralick and Shapiro, 1992; Fry, 1993).

The optimal transport problem has also been studied within the context of certain imaging applications, in particular for content-based image retrieval (Rubner, 1999; Rubner et al., 1998; Levina and Bickel, 2001). In this work, pixels in an image are divided into several bins (called "*signatures*") according to their positions in color and/or spatial locations. The Earth Mover's Distance (EMD) is calculated between the *signatures* of two images and then used for image retrieval. However, this EMD method does not give a warped grid or displacement defined at every pixel location, which is essential for image registration and image morphing.

Using the Kantorovich–Wasserstein distance for image registration and warping has a number of advantages. It is parameter free. It utilizes all of the grayscale data in both images, and places the two images on equal footing. It is thus symmetrical, the optimal mapping from image A to image B being the inverse of the optimal mapping from B to A. It is not necessary for using an extra constraint to guarantee the symmetrical property, as in Christensen and Johnson (2001). It does not require that landmarks be specified. The minimizer of the distance functional involved is unique; there are no other local minimizers. Finally, it is specifically designed to take into account changes in density that result from changes in area or volume. This last point is essential for some applications. As an example, we show in Section 4 how the method can be used to derive regular area-preserving surface diffeomorphisms from conformal mappings.

One of the key contributions of this paper is the introduction of an efficient partial differential equation approach for the computation of this metric and the warping map. Our solution of the problem allows one to compute the optimal warp from a rather simple first order partial differential equation, in contrast to higher order methods presented in such works as (Christensen et al., 1993, 1996) and to computationally complex discrete methods based on linear programming. This makes for ease of implementation and speed described below in Section 4. We give a precise formulation of the Monge–Kantorovich problem below (see Section 2), and then develop our algorithm. The key idea is to find the optimal mapping via the equivalent problem of finding the *polar factorization* of a mass preserving mapping (Gangbo, 1994; McCann, 2001). It will turn out that this may be done via a natural gradient descent technique. The details are given in Section 3. We illustrate our results on some synthetic densities and on real imagery in Section 4.

A comparison term can be added to the functional to penalize the change of intensity. The new functional is suitable for applications such as image morphing where *fade in* and *fade out* effects are undesired. A corresponding gradient descent technique is given in Section 3 and two examples based on real imagery are shown in Section 4. Finally, we note that our methods may be rigorously justified; see Angenent et al. (2003) for the mathematical details.

2. Formulation of the Problem

The following notation will be used throughout the paper:

μ_0	density map in domain Ω_0, which is positive everywhere.
μ_1	density map in domain Ω_1, which is positive everywhere.
I_0	image intensity on Ω_0.
I_1	image intensity on Ω_1.
u^0	initial mapping function from (Ω_0, μ_0) to (Ω_1, μ_1)
u	$u = u^0 \circ s^{-1}$ MP mapping function at given time t; for simplicity we usually omit t.
$\tilde{u}$	$\tilde{u} = u^0 \circ \tilde{s}^{-1}$ optimal mapping function, which is u at time $t \to \infty$.
s	MP mapping from (Ω_0, μ_0) to (Ω_0, μ_0). The inverse function is denoted s^{-1}.
$\tilde{s}$	MP mapping from (Ω_0, μ_0) to (Ω_0, μ_0) at $t \to \infty$.
w	scalar field.
χ	divergence-free vector field, such that $u = \nabla w + \chi$.
ζ	divergence-free vector field.

We now give a modern formulation of the Monge–Kantorovich problem. Let Ω_0 and Ω_1 be two subdomains of $\mathbf{R}^d$, with smooth boundaries, each with a positive density function, μ_0 and μ_1, respectively.

We assume

$$\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1 \tag{1}$$

so that the same total mass is associated with Ω_0 and Ω_1. Here, and elsewhere in this paper where appropriate, we use the simplified notation

$$\int_{\Omega} f := \int_{\Omega} f(x)\, dx, \tag{2}$$

where dx is the standard Lebesgue measure on the domain Ω.

We consider diffeomorphisms $\tilde{u}$ from Ω_0 to Ω_1 which map one density to the other in the sense that

$$\mu_0 = |D\tilde{u}|\, \mu_1 \circ \tilde{u}, \tag{3}$$

which we will call the *mass preservation* (MP) property, and write $\tilde{u} \in MP$. Equation (3) is called the *Jacobian equation.* Here $|D\tilde{u}|$ denotes the determinant of the Jacobian map $D\tilde{u}$. In particular, Eq. (3) implies, for example, that if a small region in Ω_0 is mapped to a larger region in Ω_1, then there must be a corresponding decrease in density in order for the mass to be preserved. A mapping $\tilde{u}$ that satisfies this property may thus be thought of as defining a redistribution of a mass of material from one distribution μ_0 to another distribution μ_1.

There may be many such mappings, and we want to pick out an optimal one in some sense. Accordingly, we define the L^p Kantorovich–Wasserstein metric as follows:

$$d_p^p(\mu_0, \mu_1) := \inf_{\tilde{u} \in MP} \int \|\tilde{u}(x) - x\|^p \mu_0(x)\, dx. \tag{4}$$

An *optimal MP map*, when it exists, is an MP map which minimizes this integral. This functional is seen to place a penalty on the distance the map $\tilde{u}$ moves each bit of material, weighted by the material's mass. Hence, the Kantorovich–Wasserstein metric defines the distance between two mass densities, by computing the "cheapest" way to transport the mass from one domain to the other with respect to (4).

The case $p = 2$ has been extensively studied and will be the one used in this paper for registration and image morphing. The L^2 Monge–Kantorovich problem has been studied in statistics, functional analysis, and the atmospheric sciences; see Cullen and Purser (1984), and Benamou and Brenier (2000) and the references therein. A fundamental theoretical result (Knott and Smith, 1984; Brenier, 1991; Gangbo and McCann, 1996), is that there is a unique optimal $\tilde{u} \in MP$ transporting μ_0 to μ_1, and that this $\tilde{u}$ is characterized as the gradient of a convex function w, i.e., $\tilde{u} = \nabla w$. Note that from Eq. (3), we have that w satisfies the *Monge–Ampère* equation

$$|Hw|\, \mu_1 \circ (\nabla w) = \mu_0, \tag{5}$$

where $|Hw|$ denotes the determinant of the Hessian Hw of w.

In some applications of image registration and image morphing, large intensity discrepancy between corresponding pixels is not desirable. Hence, we can further add a comparison term penalizing the changing of intensity. Our extended energy function will have the following form:

$$M := C(I_0, I_1, u) + \alpha^2 \int \|\tilde{u}(x) - x\|^p \mu_0(x)\, dx, \tag{6}$$

where C stands for any comparison term, for example squared error, normalized correlation or mutual information. Although the optimal mapping $\tilde{u}$ for this functional is no longer the gradient of a convex function, it can still be solved using the iterative method we present here for optimal mass transport problems.

3. Algorithms for Computing the Transport Map

There have been a number of algorithms considered for computing an optimal transport map. For example, methods have been proposed based on linear programming (Rachev and Rüschendorf, 1998), and on Lagrangian mechanics closely related to ideas from the study of fluid dynamics (Benamou and Brenier, 2000). A clever geometric method has been formulated by Cullen and Purser (1984). Other interesting discrete computational techniques have also been proposed (Kaijser, 1998) and applied to images. An excellent summary of the recent research on this topic may be also found (Ambrosio, 2000).

One common method is to reduce the L^2 optimal transport to a linear programming problem. Thus one can approximate the densities μ_0 and μ_1 as sums of

delta functions, such as

$$\mu_0(x) = \sum_{i=1}^{N} \delta(x - x_i), \quad \mu_1(x) = \sum_{i=1}^{N} \delta(x - y_i), \quad (7)$$

for $2N$ given points $x_1, \ldots, x_N, y_1, \ldots, y_N \in \mathbf{R}^d$. The L^2 Kantorovich distance may then be taken to be

$$d_2^2(\mu_0, \mu_1) = \min_{\rho} \sum_{i,j=1}^{N} \|x_i - y_j\|^2 \rho_{ij} \quad (8)$$

where ρ denotes any $N \times N$ doubly-stochastic matrix, i.e., a matrix with non-negative entries and row and column sums equal to 1. In the special case where ρ is restricted to be a permutation matrix σ, this formulation becomes a standard discrete *point matching* problem. A fundamental difficulty with this approach is that even in the 2D case, for a 256×256 image we may have $256^2 \times 256^2$ for the dimensions of ρ, and the linear programming problem can get to be quite unwieldy (Kaijser, 1998). Remarkably, our approach can be understood as the continuous limit of discrete point matching, as the number of points N goes to *infinity*.

3.1. Polar Factorization and Rearrangement Maps

In this section, we will employ a natural solution to the optimal transport problem based on the equivalent problem of *polar factorization*; see Brenier (1991), Gangbo (1994) and McCann (2001) and the references therein. We will work with the general case of subdomains in $\mathbf{R}^d$, and point out some simplifications that are possible for the $\mathbf{R}^2$ case.

Let $\Omega_0, \Omega_1 \subset \mathbf{R}^d$ be convex subdomains with smooth boundaries, with corresponding positive density functions μ_0 and μ_1 satisfying $\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1$ (if not, we can always perform a normalization to make them equal). Let $u : (\Omega_0, \mu_0) \to (\Omega_1, \mu_1)$ be an initial diffeomorphic mapping with the mass preserving (MP) property. Then according to the generalized results of Brenier (1991) and Gangbo (1994), u has a unique decomposition of the form

$$u = (\nabla w) \circ s, \quad (9)$$

where w is a convex function and s is an MP mapping s: $(\Omega_0, \mu_0) \to (\Omega_0, \mu_0)$. This is the *polar factorization* of u with respect to μ_0. In Gangbo (1994), just the case of area preservation is considered, i.e., μ_0 is assumed constant, but the general case goes through as well.

We will find the polar factorization of the MP mapping u, according to the following strategy. We consider the family of MP mappings of the form $\tilde{u} = u \circ s^{-1}$ as s varies over diffeomorphic MP mappings from (Ω_0, μ_0) to itself. If we consider $\tilde{u}$ as a vector field, we can always find a function w and another vector field χ, with $\operatorname{div}(\chi) = 0$, such that

$$\tilde{u} = \nabla w + \chi, \quad (10)$$

i.e., we can decompose $\tilde{u}$ into the sum of a curl-free and a divergence-free vector field (Strang, 1986). Thus, what we try to do is find a mapping s which will yield a $\tilde{u}$ without any curl, that is, such that $\tilde{u} = \nabla w$. Once such an s is found, we will have $u = \tilde{u} \circ s = (\nabla w) \circ s$ and so we will have found the polar factorization (9) of our given function u.

Now, here is the key point. As we discussed above, the unique optimal solution of the L^2 Monge–Kantorovich problem has the form $\tilde{u} = \nabla w$, and so the problem of finding the polar factorization of u and finding the optimal Monge–Kantorovich mapping $\tilde{u}$ are equivalent. The mathematical details describing this connection may be found in Ambrosio (2000), Gangbo and McCann (1996) and Brenier (1991). In essence, to solve the Monge–Kantorovich problem we create a *rearrangement* of an initial vector field u^0 using a map s, so that the resulting vector field $\tilde{u} = u^0 \circ s^{-1}$ has no curl. The mapping s is the continuous analogue of the permutation σ in the discrete point matching formulation of the Monge–Kantorovich problem described above. By construction, the resulting $\tilde{u} : \Omega_0 \to \Omega_1$ is a mass preserving mapping of the form $\tilde{u} = \nabla w$, with w convex since $\tilde{u}$ is a diffeomorphism. Uniqueness follows from the theory of the Monge–Ampère Eq. (5) (see Taylor, 1996, p. 251). We can now give the technical details of our construction.

3.2. Finding an Initial Mapping

We will now propose an explicit algorithm to solve the Monge–Kantorovich problem. We will try to do this by finding an initial MP mapping $u^0 = (a(x), b(x, y))$ and then minimizing over $\tilde{u} = u^0 \circ s^{-1}$ by varying s over MP mappings from Ω_0 to Ω_0, starting with s equal to the identity map. The initial mapping can be found in general domains using a method of Moser (1965)

and Dacorogna and Moser (1990), or for simpler domains using the following algorithm. For simplicity, we work in $\mathbf{R}^2$ and assume $\Omega_0 = \Omega_1 = [0, 1]^2$, the generalization to higher dimensions being straightforward. The idea of this construction is that we solve a family of one-dimensional mass transport problems. In one-dimension, the optimal transport map can be found by simple quadrature. We first transport mass along lines parallel to the x-axis, and then afterward transport mass along lines parallel to the y-axis.

Accordingly, we define a function $a = a(x)$ implicitly by the equation

$$\int_0^{a(x)} \int_0^1 \mu_1(\eta, y)\, dy\, d\eta = \int_0^x \int_0^1 \mu_0(\eta, y)\, dy\, d\eta \tag{11}$$

which gives by differentiation with respect to x

$$a'(x) \int_0^1 \mu_1(a(x), y)\, dy = \int_0^1 \mu_0(x, y)\, dy. \tag{12}$$

We may now define a function $b = b(x, y)$ implicitly by the equation

$$a'(x) \int_0^{b(x,y)} \mu_1(a(x), \rho)\, d\rho = \int_0^y \mu_0(x, \rho)\, d\rho, \tag{13}$$

and set $u^0(x, y) = (a(x), b(x, y))$. Since $a_y = 0$, $|Du| = a_x b_y$, and differentiating (13) with respect to y we find

$$a'(x)\, b_y(x, y)\, \mu_1(a(x), b(x, y)) = \mu_0(x, y)$$
$$|Du|\, \mu_1 \circ u = \mu_0,$$

which is the MP property we need. In practice, a and b can be found with simple numerical integration techniques. Given our assumption that μ_0 and μ_1 are positive everywhere, $a(x)$ is well-defined and strictly monotone increasing by (11). The function $b(x, y)$ is also well-defined and strictly monotone increasing with respect to y, by (13) and the fact that $a'(x)$ is always positive. Hence, there is no space folding problem, i.e. the mapping is bijective.

3.3. Removing the Curl

Once an initial MP u^0 is found, we need to apply the process which will remove its curl. A simple calculation verifies that the composition of two mass preserving (MP) mappings is an MP mapping, and the inverse of an MP mapping is an MP mapping. Since u^0 is MP, while $\tilde{u} = u^0 \circ \tilde{s}^{-1}$ we see that $\tilde{u}$ is an MP mapping if and only if $\tilde{s}$ is an MP map, that is, if and only if

$$\mu_0 = |D\tilde{s}|\, \mu_0 \circ \tilde{s}. \tag{14}$$

The same argument applies to s also.

Next, rather than working with s directly, we solve the polar factorization problem via gradient descent. We should note that the algorithm described below converges to a global optimum (Angenent et al., 2003). Accordingly, we will assume that s is a function of time, and then determine what $s_t := \frac{d}{dt} s$ should be to decrease the L^2 Monge–Kantorovich functional. This will give us an evolution equation for s and in turn an equation for u_t as well, the latter being the most important for implementation. In what follows the t subscript denotes differentiation with respect to time t, while the D, ∇ and div refer to spatial derivatives.

One can easily show (see the Appendix below for details) that s_t and $\tilde{u}_t$ should have the following forms in order to preserve the MP property:

$$s_t = \left(\frac{1}{\mu_0}\zeta\right) \circ s, \tag{15}$$

$$u_t = -\frac{1}{\mu_0} Du\, \zeta, \tag{16}$$

for some vector field ζ on Ω_0, with $\operatorname{div}(\zeta) = 0$ and $\langle \zeta, \vec{n} \rangle = 0$ on $\partial\Omega_0$, $\vec{n}$ being the normal to the boundary of Ω_0. This last condition ensures that s remains a mapping from Ω_0 to itself, by preventing the flow of s, given by (15), from crossing the boundary of Ω_0. This also means that the range of $u = u^0 \circ s^{-1}$ is always $u(\Omega_0) = \Omega_1$.

Consider now the problem of minimizing the Monge–Kantorovich functional:

$$M = \int_{\Omega_0} \|u - x\|^2 \mu_0 \tag{17}$$

By taking the derivative with respect to time t, we have (see the Appendix for details)

$$-\frac{1}{2} M_t = \int_{\Omega_0} \langle u, \mu_0(s_t \circ s^{-1}) \rangle \tag{18}$$

where s is the MP mapping such that $u = u^0 \circ s^{-1}$. Referring to (15), we can rewrite $\mu_0(s_t \circ s^{-1})$ as

ζ, with ζ divergence free, so that:

$$-\frac{1}{2} M_t = \int_{\Omega_0} \langle u, \zeta \rangle. \tag{19}$$

Now decomposing u as $u = \nabla w + \chi$, (Helmholtz decomposition) we have

$$-\frac{1}{2} M_t = \int_{\Omega_0} \langle \nabla w + \chi, \zeta \rangle \tag{20}$$

$$= \int_{\Omega_0} \langle \chi, \zeta \rangle, \tag{21}$$

by the divergence theorem. Thus, in order to decrease M, we can take $\zeta = \chi$ with corresponding formulas (15)–(16) for s_t and u_t, provided that we have $\operatorname{div}(\chi) = 0$ and $\langle \chi, \vec{n} \rangle = 0$ on $\partial\Omega_0$. Thus the remaining task is to decompose u as $u = \nabla w + \chi$.

Gradient Descent: $\mathbf{R}^d$:

We let w be a solution of the Neumann-type boundary problem

$$\Delta w = \operatorname{div}(u) \tag{22}$$

$$\langle \nabla w, \vec{n} \rangle = \langle u, \vec{n} \rangle \quad \text{on} \quad \partial\Omega_0, \tag{23}$$

and set $\chi = u - \nabla w$. It is then easily seen that χ satisfies the necessary requirements. Thus, by (16), we have the following evolution equation for u:

$$u_t = -\frac{1}{\mu_0} Du \, (u - \nabla \Delta^{-1} \operatorname{div}(u)). \tag{24}$$

where we have used $\Delta^{-1} \operatorname{div}(u)$ to signify the solution w to (22). This is a first order non-local scheme for u_t if we count Δ^{-1} as minus 2 derivatives. Note that this flow is consistent with respect to the Monge–Kantorovich theory in the following sense. If $\tilde{u}$ is the optimal MP mapping, then it is given as $\tilde{u} = \nabla w$, in which case $\tilde{u} - \nabla \Delta^{-1} \operatorname{div}(\tilde{u}) = \nabla w - \nabla \Delta^{-1} \operatorname{div}(\nabla w) = 0$ so that by (24), $\tilde{u}_t = 0$ and the process has achieved steady state.

Gradient Descent: $\mathbf{R}^2$:

The situation is somewhat simpler in the $\mathbf{R}^2$ case, due to the fact that a divergence free vector field χ can in general be written simply as $\chi = \nabla^\perp h$ for some scalar function h, where $\perp$ represents rotation by 90 degrees, so that $\nabla^\perp h = (-h_y, h_x)$. In this case (21) becomes

$$-\frac{1}{2} M_t = \int_{\Omega_0} \langle \nabla^\perp f, \nabla^\perp h \rangle = \int_{\Omega_0} \langle \nabla f, \nabla h \rangle \tag{25}$$

where the decomposition of $\tilde{u}$ is $u = \nabla w + \nabla^\perp f$, and we can take $h = f$. The function f can be found by solving the Dirichlet-type boundary problem

$$\Delta f = -\operatorname{div}(u^\perp), \tag{26}$$

$$f = 0 \quad \text{on} \quad \partial\Omega_0, \tag{27}$$

which gives us the evolution equation

$$u_t = \frac{1}{\mu_0} Du \, \nabla^\perp \Delta^{-1} \operatorname{div}(u^\perp). \tag{28}$$

We may also derive a second order *local* evolution equation for u. For example, in the $\mathbf{R}^2$ case (see Angenent et al. (2003) for generalizations), we use the divergence theorem with (25) to get

$$u_t = -\frac{1}{\mu_0} Du \, \nabla^\perp \operatorname{div}(u^\perp), \tag{29}$$

where appropriate handling of the evolution at the boundary, as described in Section 4, is required.

3.4. Adding a Comparison Term

The L^2 Monge–Kantorovich metric has a penalty only on the "effort" required to move the mass from one configuration into another. For problems of image registration and image morphing where it is assumed that objects undergo changes in size but not intensity, a mapping that maps a small high-intensity region into a large low-intensity region is not desired. Hence, we allow for the addition of a comparison term to the energy function penalizing change of intensity. The idea is to minimize a functional of the following form over mass-preserving mappings $u : \Omega_0 \to \Omega_1$:

$$M_\alpha[u] := \int (I_1 \circ u - I_0)^2 \, dx + \alpha^2 \int \|u(x) - x\|^2 \mu_0 \, dx, \tag{30}$$

for a fixed $\alpha \in \mathbf{R}$. Here the first term controls the "goodness of fit" between the (intensity) images $I_0 : \Omega_0 \to \mathbf{R}$ and $I_1 : \Omega_1 \to \mathbf{R}$, and the second Monge-Kantorovich term controls the the warping of the map. The function μ_0 is the mass density of the source image defined on Ω_0, and could be the same as I_0 or a smoothed version of I_0. It could also be any scalar field that is appropriate for the underlying physical model. Similarly, μ_1 is assumed to be the mass density of the target image

defined on Ω_1. By adjusting α, a tradeoff between minimal mass transportation and minimal intensity change is achieved.

Taking the derivative of (30) respective to time t, we find

$$-\frac{dM_a}{dt} = \int \left\{ \left\langle \frac{1}{\mu_0^2}(I_1 \circ u - I_0)^2 \nabla \mu_0 + \frac{2}{\mu_0}(I_1 \circ u - I_0)\nabla I_0 + 2\alpha^2 u,\ \mu_0 \frac{\partial s}{\partial t} \circ (s)^{-1} \right\rangle \right\} dx \quad (31)$$

Let us denote the left hand term of the inner product by P. As with u above, P can be decomposed into a curl-free term and a divergence-free term as $P = \nabla w + \chi$, and the updating of u can be expressed as

$$u_t = -\frac{1}{\mu_0} Du\ \chi \quad (32)$$

or in the $\mathbf{R^2}$ case, it can be written as

$$u_t = \frac{1}{\mu_0}\, Du\, \nabla^\perp \Delta^{-1}\, \mathrm{div}(P^\perp) \quad (33)$$

for the non-local flow, and

$$u_t = -\frac{1}{\mu_0}\, Du\, \nabla^\perp\, \mathrm{div}(P^\perp) \quad (34)$$

for the local flow.

3.5. Methods

In this subsection we will outline the numerical details for finding the optimal mass preserve mapping. We limit ourselves here for simplicity to the 2D case, the higher dimensional cases being a straightforward generalization

(1) *Construct an initial MP mapping u^0 as described in Section 3.2*;
(2) *Calculate P, the left hand term in the inner product in (18) or (31). Specifically, take $P = u$ for the pure optimal mass transport problem as in Section 3.3, and*

$$P = \frac{1}{\mu_0^2}(I_1 \circ u - I_0)^2 \nabla \mu_0 + \frac{2}{\mu_0}(I_1 \circ u - I_0)\nabla I_0 + 2\alpha^2 u$$

for the functional with the additional comparison term as in Section 3.4;
(3) *Decompose P into a curl-free term ∇w and a divergence-free term χ as $P = \nabla + \chi$. For the nonlocal flow, this involves solving Poisson's equation with appropriate boundary conditions;*
(4) *Update u according to (16): $u_t = -\frac{1}{\mu_0} Du\ \chi$. In the 2D case this becomes*

$$u_t = \frac{1}{\mu_0} Du \nabla^\perp \Delta^{-1} div(P^\perp) \quad \textit{for the nonlocal flow},$$

or

$$u_t = -\frac{1}{\mu_0}\, Du\, \nabla^\perp div(P^\perp) \quad \textit{for the local flow};$$

(5) *Go to step (2).*

The optimal map is obtained as $t \to \infty$. In practice, the procedure iterates until the mean absolute curl is sufficiently small (in the case of no comparison term) or the energy is decreasing sufficiently slowly (in the case with the extra comparison term). We use standard techniques to solve the local and nonlocal flows. In particular we use an upwinding scheme when computing Du, and Matlab's Poisson equation solver, which uses sine transforms to invert the Laplacian on a rectangular grid. The time step dt can be chosen to be less than

$$\min_{x,i} \left| \frac{1}{\mu_0}(\nabla^\perp \Delta^{-1}\, \mathrm{div}(P^\perp))_i \right|^{-1}$$

for the nonlocal flow, where the subscript i stands for the component of the vector. Standard centered differences were used for the other spatial derivatives. Once we numerically solve for the right hand side of (24), (28), (29), (33) or (34), we use the result to update u. More details of the numerical implementation for solving the equations are given below in Section 4. The complexity of the method during each iteration scales as $N \log_2 N$ where N is the number of pixels in the image, due to the use of sine transforms.

3.6. Defining the Warping Map

Typically in elastic registration, one wants to see an explicit warping which smoothly deforms one image into the other. This can easily be done using the solution of the Monge–Kantorovich problem. Thus, we assume

now that we have applied our gradient descent process as described above and that it has converged to the Monge–Kantorovich mapping $\tilde{u}_{MK}$.

Following the work of Benamou and Brenier (2000) (see also Gangbo and McCann, 1996), we consider the following related problem:

$$\inf \int \int_0^1 \mu(t,x)\|v(t,x)\|^2 \, dt \, dx \tag{35}$$

over all time varying densities μ and velocity fields v satisfying

$$\frac{\partial \mu}{\partial t} + \operatorname{div}(\mu v) = 0, \tag{36}$$

$$\mu(0,\cdot) = \mu_0, \quad \mu(1,\cdot) = \mu_1. \tag{37}$$

It is shown in Benamou and Brenier (2000) that this infimum is attained for some $\mu_{\min}$ and $v_{\min}$, and that it is equal to the L^2 Kantorovich–Wasserstein distance between μ_0 and μ_1. Further, the flow $X = X(x,t)$ corresponding to the minimizing velocity field $v_{\min}$ via

$$X(x,0) = x, \; X_t = v_{\min} \circ X \tag{38}$$

is given simply as

$$X(x,t) = x + t\,(\tilde{u}_{MK}(x) - x). \tag{39}$$

Note that when $t = 0$, X is the identity map and when $t = 1$, it is the solution $\tilde{u}_{MK}$ to the Monge–Kantorovich problem. This analysis provides appropriate justification for using (39) to *define* our continuous warping map X between the densities μ_0 and μ_1. See McCann (1997) for applications and a detailed analysis of the properties of this *displacement interpolation*. Other interpolation schemes could also be used, such as the one proposed in Hinterberger and Scherzer (2001).

4. Implementation and Examples

We illustrate our methods with the following examples. The first is the mapping of one synthetic density onto another. Figure 1 shows a mass distribution μ_0 on Ω_0, with dark regions representing little mass, lighter regions representing more. Similarly, Fig. 2 indicates the density μ_1 on Ω_1. Figure 3 represents the initial mapping u^0, which was obtained by the method described in Section 3. The shading in this figure represents the

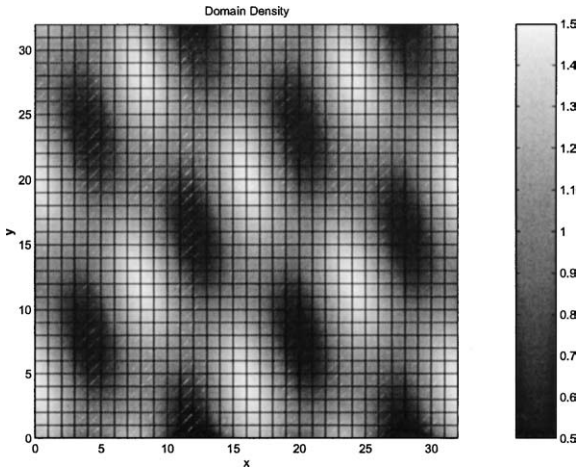

Figure 1. Density μ_0 on Ω_0.

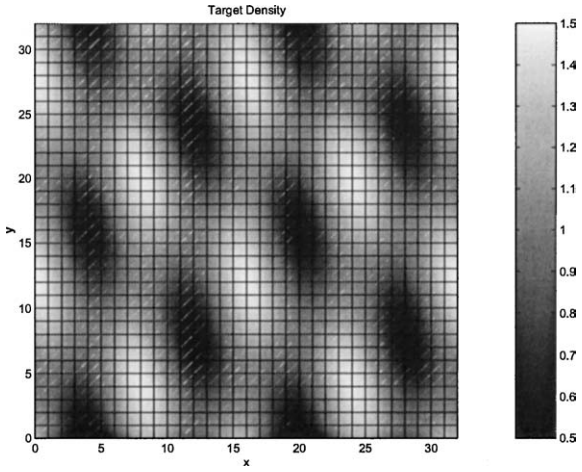

Figure 2. Density μ_1 on Ω_1.

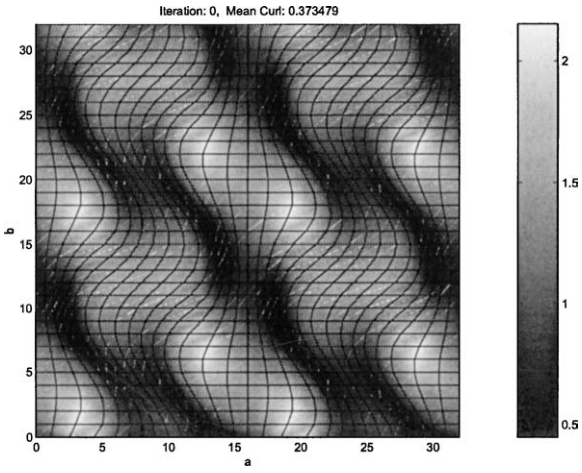

Figure 3. Initial mass—preserving mapping from Ω_0 to Ω_1.

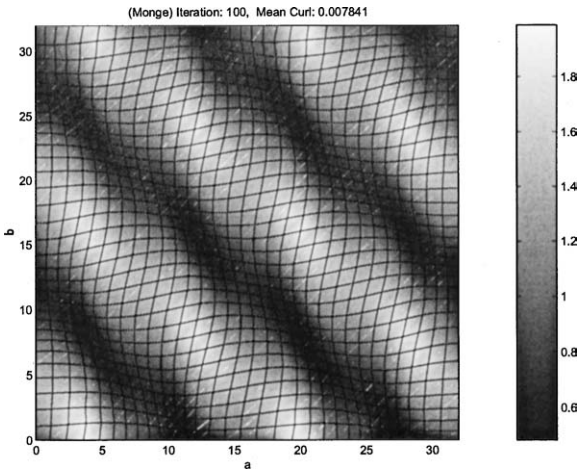

Figure 4. Final Monge–Kantorovich mapping from Ω_0 to Ω_1.

Jacobian of u^0. Figure 4 shows the optimal Monge–Kantorovich mapping $\tilde{u}_{MK}$, obtained using the non-local first order flow given in Eq. (28). One can see that the effect of removing the curl is to straighten out the grid lines somewhat. On a Sun Ultra10, this process took just a few seconds. Similarly, Figs. 5 through 9 show the result of applying our process to a pair of "cloud" densities. Figure 10 give the corresponding warped grid of the deformation.

In Fig. 11 we show an image morphing example of a flame sequence taken from the *Artbeats Digital Film Library*. Note that only the first and last images are

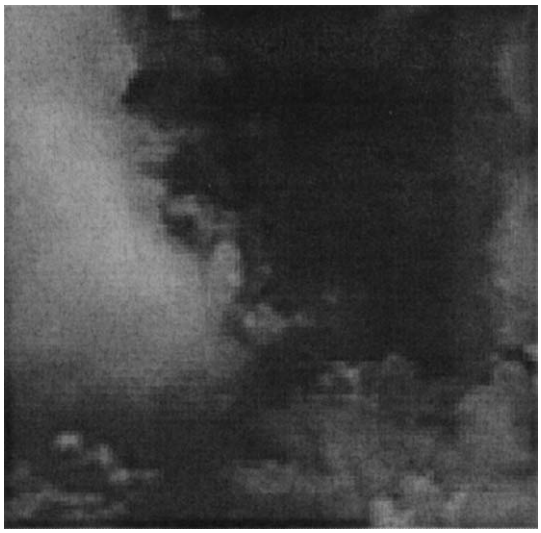

Figure 5. Cloud, time $t = 0.00$.

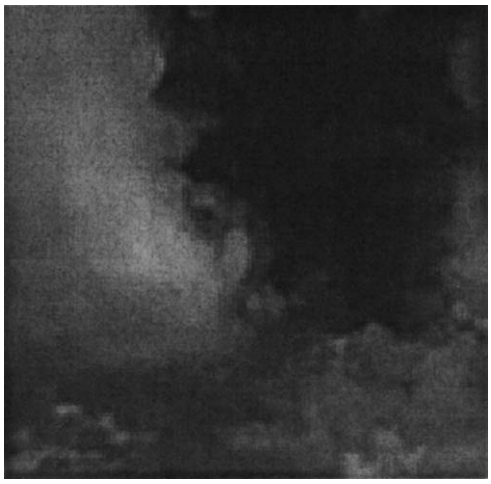

Figure 6. Cloud, time $t = 0.25$.

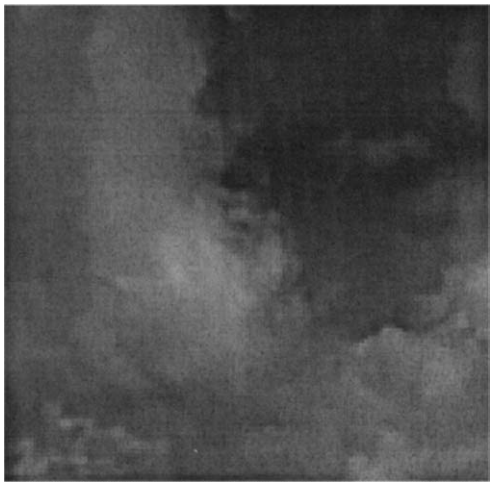

Figure 7. Cloud, time $t = 0.50$.

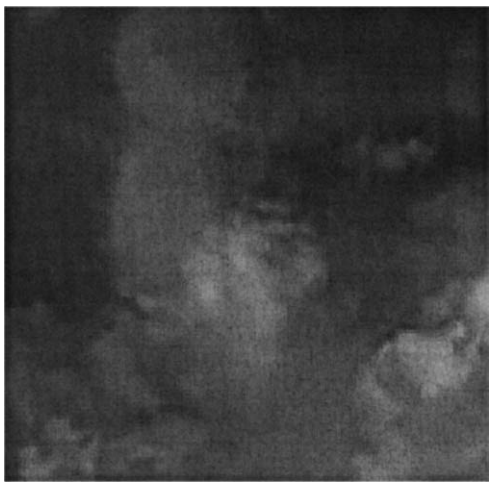

Figure 8. Cloud, time $t = 0.75$.

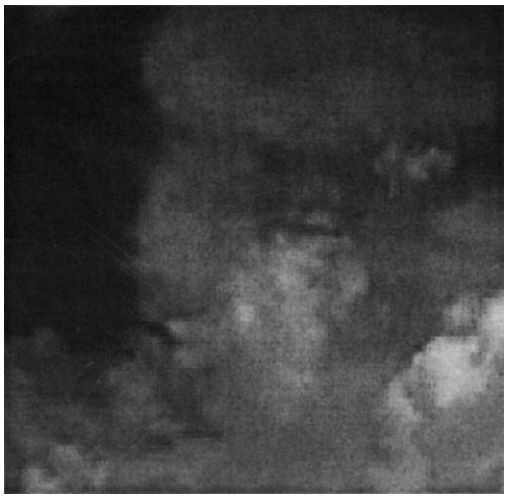

Figure 9. Cloud, time $t = 1.00$.

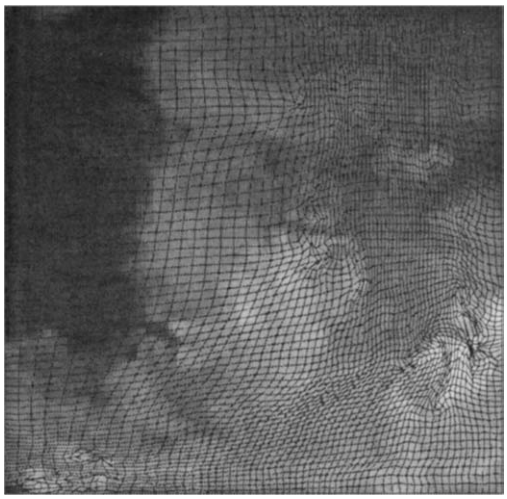

Figure 10. Deformation grid for cloud images.

given, and the three middle images are generated using our process. The corresponding deformation grid is shown in Fig. 12. In this example, as well as in the cloud example, an intensity comparison term was applied to reduce undesired *fade in* and *fade out* effects. In these two examples, the deformed grids were calculated from gray scale images and densities were normalized before finding the initial MP mapping. We should point out that by adding the intensity comparison term, there was some curl left in the resulting mapping. Hence, there is a trade-off between the remaining curl and the resulting *fade in* and *fade out* effects.

The final example shows an application of our method to the problem of constructing area preserving surface warpings. In particular, we consider the problem of brain flattening, i.e., mapping part of the surface of the brain onto a portion of the plane. We show a portion of the brain white matter surface obtained from the segmentation of a volumetric magnetic resonance imaging scan in Fig. 13.

The image on the left of Fig. 14 shows the flattened white matter surface obtained using a conformal flattening technique (Angenent et al., 1999b). It is well known that a surface of non-zero Gaussian curvature can not be flattened by any means without some metric distortion. The conformal mapping is an attempt to preserve the appearance of the surface after flattening through the preservation of angles. However, in some applications it is desirable to be able to preserve areas instead of angles, so that the sizes of surface structures are accurately represented in the plane. The Monge–Kantorovich approach allows us to find such an area-correct flattening. Specifically, once we have conformally flattened the surface, we define a "pseudo"

Figure 11. Flame interpolation: $t = 0.00, 0.25, 0.50, 0.75$ and 1.00 respectively.

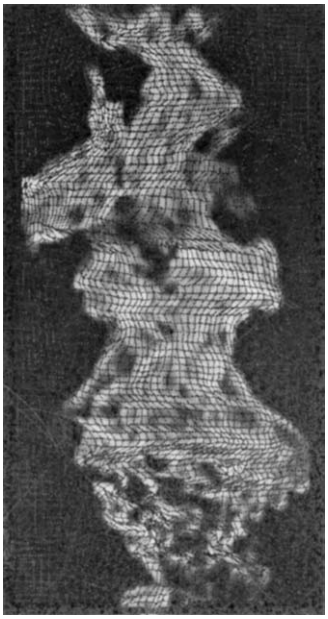

Figure 12. Deformation grid for flame images.

Figure 13. Prtion of the segmented brain white matter surface. Points are colored according to mean curvature with lighter colored points having curvature.

density μ_0 to be the Jacobian of the inverse of the flattening map, and set μ_1 to a constant. The Monge–Kantorovich optimal mapping is then area-corrected by (3). The image on the right indicates the area-corrected flattening obtained via the Monge–Kantorovich approach. The resulting map took just a few minutes to calculate. Although corrected for area, surface structures are still clearly discernible. The curl-free nature of the Monge–Kantorovich mapping avoids distortion effects often associated with area preserving maps. We have used a brain surface here, but the method can be applied to any surface, e.g., to the colon surface as part of a virtual colonoscopy or virtual pathology technique.

We have also successfully implemented the second order local flow as Eq. (29), with similar results. In this case, we require that a periodic boundary condition be enforced, specifically that $\tilde{u}(x) - x$ be periodic on the square image domain. We also used a standard upwinding scheme when calculating $D\tilde{u}$. While it may seem that this local flow should provide a faster method than the non-local flow (28), in practice this does not seem to be the case. Even though the non-local method requires that the Laplacian be inverted during each iteration, the problem has been set up to allow the use of fast numerical solvers which use FFT-type methods and operate on rectangular grids (Press et al., 1992). We have used the Matlab solver here, which uses sine transforms followed by the solution of a tri-diagonal system. Moreover, we have found that the functional is decreased substantially more during each iteration of the non-local method, using the maximum temporal step size allowed for stability in each case.

In general, the target domain Ω_1 need not be rectangular when using the non-local method. However, we note that if the periodic boundary condition described

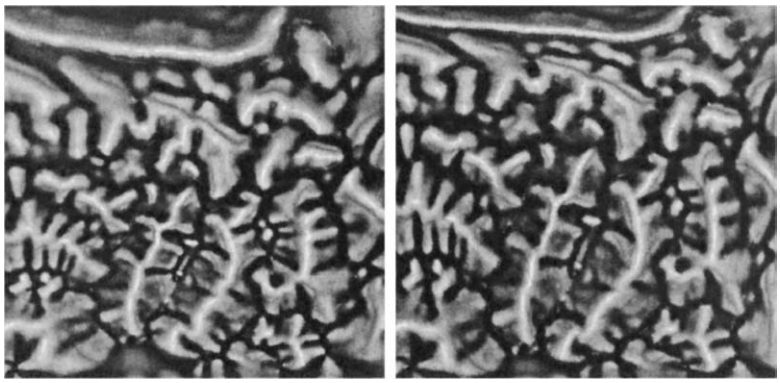

Figure 14. Conformal (left) and area corrected (right) flattenings. The shading scheme is the same as in Fig. 13.

above is used, then the Laplacian in (28) can be inverted using the FFT alone, without the need to solve a subsequent matrix system. For the warps shown, this reduced the processing time by 1/3.

5. Conclusion and Future Work

In this paper, we presented a natural method for finding registration mappings based on the classical problem of optimal mass transportation. We showed that for an L^2 version of the problem, one can derive easily-implementable gradient descent equations. Although applied here to the Monge–Kantorovich problem, the method used to enforce the mass preservation constraint is general and has other applications. In particular, we are currently working on using such constraints while minimizing standard Dirichlet-type energy functionals.

In particular, the concept of a harmonic mapping, defined as a minimizer of the Dirichlet integral, can be combined with a mass preservation constraint to obtain a new approach to mass-preserving diffeomorphisms (Angenent et al., 1999a). We state the results for Euclidean space even though they apply more generally to Riemannian surfaces. As above, let $\Omega_0,\ \Omega_1 \subset \mathbf{R}^2$ be subdomains equipped with positive densities μ_0 and μ_1, respectively, and consider the minimization of the Dirichlet integral over all MP maps:

$$\min_{u \in MP} \int_{\Omega_0} \|Du\|^2. \tag{40}$$

A minimizer (when it exists) is called an area-preserving map of *minimal distortion.* Non-local and local gradient descent methods for computing such a map can be derived in a manner very similar to that described above for the Monge–Kantorovich functional (Angenent et al., 1999a). These methods have applications to brain surface flattening and virtual colonoscopy as described in Angenent et al. (1999b) and Haker et al. (2000). The numerical procedure and applications will be presented in a future paper. Future work will also include the use of Monge–Kantorovich type functionals as a regularizer in standard problems from computer vision such as optical flow.

Appendix

In this section we give some mathematical details omitted in the main part of the paper. It will mainly focus on the proof of the properties of MP mapping and the deduction of our gradient descent method. Full details of the mathematics underlying our methodology (including convergence to thee optimal solution) may be found in Angenent et al. (2003).

A.1. *Optimal Mass Preserve Mapping is Symmetrical*

The optimal mass preserve mapping places the two images on equal footing and is symmetrical: the optimal mass preserve mapping from (Ω_0, μ_0) to (Ω_1, μ_1) is the inverse of the optimal mapping from (Ω_1, μ_1) to (Ω_0, μ_0).

Assume u is a MP mapping from (Ω_0, μ_0) to (Ω_1, μ_1) and $y = u(x)$, we have

$$\begin{aligned}
&\int_{\Omega_0} \|u(x) - x\|^p \mu_0(x)\, dx \\
&\quad = \int_{\Omega_1} \|u \circ u^{-1}(y) - u^{-1}(y)\|^p \mu_0 \circ u^{-1}(y)\, du^{-1}(y) \\
&\quad = \int_{\Omega_1} \|y - u^{-1}(y)\|^p \mu_0 \circ u^{-1}(y) |Du^{-1}(y)|\, dy \\
&\quad = \int_{\Omega_1} \|y - u^{-1}(y)\|^p \mu_1(y)\, dy,
\end{aligned}$$

where we used the MP property of u^{-1}: $\mu_1 = |Du^{-1}|\, \mu_0 \circ u^{-1}$.

Since this holds for any $u \in MP$, we have that

$$d_p(\mu_0, \mu_1) = d_p(\mu_1, \mu_0),$$

and that if $\tilde{u}_{MK}$ denotes the optimal mass-preserving map from Ω_0 to Ω_1 then the optimal from Ω_1 to Ω_0 is precisely $\tilde{u}_{MK}^{-1}$.

A.2. *Properties of Mass Preserve Mappings*

As we have mentioned in Section 3.3, in order to preserve the MP property, the updating of u and s should have the following forms:

$$s_t = \left(\frac{1}{\mu_0}\zeta\right) \circ s, \tag{41}$$

$$u_t = -\frac{1}{\mu_0} Du\, \zeta, \tag{42}$$

Now we will prove this assertion. Since s is MP mapping from (Ω_0, μ_0) to itself, we have that $\mu_0 =$

$|Ds|\,\mu_0 \circ s$. By differentiating this equation with respect to time, we get

$$\begin{aligned}
0 &= |Ds|_t\,\mu_0 \circ s + |Ds|\,(\mu_0 \circ s)_t \\
&= |Ds|(\operatorname{div}(s_t \circ s^{-1}) \circ s)\,\mu_0 \circ s \\
&\quad + |Ds|\langle(\nabla\mu_0) \circ s, s_t\rangle, \\
0 &= (\mu_0 \operatorname{div}(s_t \circ s^{-1})) \circ s \; + \langle(\nabla\mu_0) \circ s, s_t\rangle, \\
0 &= \mu_0 \operatorname{div}(s_t \circ s^{-1}) + \langle\nabla\mu_0, s_t \circ s^{-1}\rangle \\
&= \operatorname{div}(\mu_0\, s_t \circ s^{-1}),
\end{aligned}$$

Hence s_t should have the following form

$$s_t = \left(\frac{1}{\mu_0}\zeta\right) \circ s, \tag{43}$$

for some vector field ζ on Ω_0, with $\operatorname{div}(\zeta) = 0$ and $\langle\zeta, \vec{n}\rangle = 0$ on $\partial\Omega_0$, $\vec{n}$ being the normal to the boundary of Ω_0. Since u and s satisfy $u \circ s = u^0$, by taking the derivative of it respective to t, we derive

$$\begin{aligned}
(Du \circ s)\,s_t + u_t \circ s &= 0 \\
u_t \circ s &= -(Du \circ s)\,s_t \\
u_t &= -Du\; s_t \circ s^{-1}.
\end{aligned}$$

By (43) we have

$$u_t = -\frac{1}{\mu_0} Du\,\zeta, \tag{44}$$

A.3. Gradient Descent Method for Pure Monge–Kantorovich Problem

Consider now the problem of minimizing the Monge–Kantorovich functional:

$$M = \int_{\Omega_0} \|u - x\|^2 \mu_0 \tag{45}$$

$$= \int_{\Omega_0} \|u\|^2 \mu_0 - 2\int_{\Omega_0} \langle u, x\rangle \mu_0 + \int_{\Omega_0} \|x\|^2 \mu_0. \tag{46}$$

The last term is obviously independent of time. Interestingly, so is the first. To see this, we set $y = s^{-1}(x)$, and using the MP property of s and s^{-1}, we find

$$\begin{aligned}
\mu_0(x)\,dx &= \mu_0 \circ s(y)\,ds(y) = \mu_0 \circ s(y)\,|\,Ds(y)\,|\,dy \\
&= \mu_0(y)\,dy,
\end{aligned} \tag{47}$$

since $\mu_0(y) = \mu_0 \circ s(y)|Ds(y)|$. Now we have

$$\begin{aligned}
\int_{\Omega_0} \|u(x)\|^2 \mu_0(x)\,dx &= \int_{\Omega_0} \|u^0(x) \circ s^{-1}(x)\|^2 \mu_0(x)\,dx \\
&= \int_{\Omega_0} \|u^0(y)\|^2\,\mu_0(y)\,dy
\end{aligned}$$

which is a constant for all time. Turning now to the middle term, we do the same trick by setting $y = s^{-1}(x)$. Now we have

$$\begin{aligned}
&\int_{\Omega_0} \langle u(x), x\rangle \mu_0 \\
&\quad = \int_{\Omega_0} \langle u^0 \circ s^{-1}(x), s \circ s^{-1}(x)\rangle\,\mu_0(x)\,dx \\
&\quad = \int_{\Omega_0} \langle u^0(y), s(y)\rangle\,\mu_0(y)\,dy,
\end{aligned}$$

and taking $s_t = (\frac{1}{\mu_0}\zeta) \circ s$, we compute

$$\begin{aligned}
-\frac{1}{2} M_t &= \int_{\Omega_0} \langle u^0(y), s_t(y)\rangle\,\mu_0(y)\,dy \\
&= \int_{\Omega_0} \langle u \circ s(y), \left(\frac{1}{\mu_0}\zeta\right) \circ s(y)\rangle\,\mu_0(y)\,dy \\
&= \int_{\Omega_0} \langle u(x), \frac{1}{\mu_0(x)}\zeta(x)\rangle\,\mu_0(x)\,dx \\
&= \int_{\Omega_0} \langle u(x), \zeta(x)\rangle\,dx.
\end{aligned}$$

Now decomposing u as $u = \nabla w + \chi$, we have

$$\begin{aligned}
-\frac{1}{2} M_t &= \int_{\Omega_0} \langle\nabla w + \chi, \zeta\rangle \\
&= \int_{\Omega_0} \langle\nabla w, \zeta\rangle + \int_{\Omega_0} \langle\chi, \zeta\rangle \\
&= \int_{\Omega_0} (\operatorname{div}(w\zeta) - w \operatorname{div}(\zeta)) + \int_{\Omega_0} \langle\chi, \zeta\rangle \\
&= \int_{\partial\Omega_0} w\,\langle\zeta, n\rangle + \int_{\Omega_0} \langle\chi, \zeta\rangle \\
&= \int_{\Omega_0} \langle\chi, \zeta\rangle,
\end{aligned}$$

where we've used the divergence theorem, $\operatorname{div}(\zeta) = 0$, and $\langle\zeta, \vec{n}\rangle = 0$ on $\partial\Omega_0$. And then we choose $\zeta = \chi$ for updating u.

A.4. Gradient Descent Method for Functional with Comparison Term

For convenience, we rewrite the our functional for image interpolation as following,

$$M_\alpha[u] := \int (I_1 \circ u - I_0)^2 \, dx + \alpha^2 \int \|u(x) - x\|^2 \mu_0 \, dx, \quad (48)$$

The derivative of the second part is exactly the same as we discussed above. Now we focus on the first part:

$$M_1(u) = \int (I_1 \circ u - I_0)^2 \; dx, \quad (49)$$

Taking the time derivative of M_1 and setting $y = s^{-1}(x)$, we get

$$\begin{aligned}
M_1 &= \int (I_1 \circ u(x) - I_0(x))^2 \, dx \\
&= \int \left(\frac{1}{\mu_0(x)} (I_1 \circ u(x) - I_0(x))^2 \right) \mu_0(x) \, dx \\
&= \int \left(\frac{1}{\mu_0 \circ s(y)} (I_1 \circ u^0(y) - I_0 \circ s(y))^2 \right) \\
&\quad \times \mu_0(y) \, dy \\
M_{1t} &= \int \Big\langle -\frac{1}{\mu_0^2 \circ s(y)} (I_1 \circ u^0(y) \\
&\quad - I_0 \circ s(y))^2 \nabla \mu_0 \circ s(y) \\
&\quad - \frac{2}{\mu_0 \circ s(y)} (I_1 \circ u^0(y) \\
&\quad - I_0 \circ s(y)) \nabla I_0 \circ s(y), \frac{\partial s^t}{\partial t} \Big\rangle \mu_0(y) \, dy \\
-M_{1t} &= \int \Big\langle \frac{1}{\mu_0^2(x)} (I_1 \circ u(x) - I_0(x))^2 \nabla \mu_0(x) \\
&\quad + \frac{2}{\mu_0(x)} (I_1 \circ u(x) - I_0(x)) \nabla I_0(x), \\
&\quad \mu_0 \frac{\partial s}{\partial t} \circ s^{-1}(x) \Big\rangle dx
\end{aligned}$$

By setting $P = \frac{1}{\mu_0^2}(I_1 \circ u - I_0)^2 \nabla \mu_0 + \frac{2}{\mu_0}(I_1 \circ u - I_0) \nabla I_0 + 2\alpha^2 u$ and decomposing $P = \nabla w + \chi$, we now get the $\zeta = \chi$ for updating u.

Acknowledgments

The authors would like to thank Professor Rob McCann of the University of Toronto, Professor Mike Cullen of the ECMWF, and Professors Wilfrid Gangbo and Anthony Yezzi of Georgia Tech for some very interesting conversations about the Monge–Kantorovich problem. This work was supported in part by grants from the National Science Foundation ECS-9700588, NSF-LIS, by the Air Force Office of Scientific Research AF/F49620-98-1-0168, by the Army Research Office DAAG55-98-1-0169, a MURI Grant, and NIH grants including a training grant, R01 AG19513, and NAC through Brigham and Women's Hospital.

References

Ambrosio, L. 2000. Lecture notes on optimal transport problems. Lectures given at Euro Summer School. Available on http://cvgmt.sns.it/papers/amb00a/.

Angenent, S., Haker, S., and Tannenbaum, A. 2003. Minimizing flows for the Monge–Kantorovich problem. *SIAM J. Math. Analysis*, 35:61–97.

Angenent, S., Haker, S., Tannenbaum, A., and Kikinis, R. 1999a. On area preserving maps of minimal distortion. In *System Theory: Modeling, Analysis, and Control*, T. Djaferis and I. Schick (Eds.), Kluwer: Holland, pp. 275–287.

Angenent, S., Haker, S., Tannenbaum, A., and Kikinis, R. 1999b. Laplace-Beltrami operator and brain surface flattening. *IEEE Trans. on Medical Imaging*, 18:700–711.

Aruliah, D.A., Aschery, U.M., Haberz, E., and Oldenburgx, D. 2001. A method for the forward modelling of 3D electromagnetic quasi-static problems. *Mathematical Models and Methods in Applied Sciences*, 11:1–21.

Benamou, J.-D. and Brenier, Y. 2000. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393.

Brenier, Y. 1991. Polar factorization and monotone rearrangement of vector-valued functions. *Com. Pure Appl. Math.*, 64:375–417.

Bro-Nielsen, M. and Gramkow, C. 1996. Fast fluid registration of medical images. In *Visualization in Biomedical Imaging*, K. Höhne and R. Kikinis (Eds.), Lecture Notes in Computer Science, vol. 1131, Springer-Verlag: New York, pp. 267–276.

Christensen, G.E., Rabbit, R.D., and Miller, M. 1996. Deformable templates using large deformation kinetics. *IEEE Trans. on Image Processing*, 5:1435–1447.

Christensen, G.E., Rabbit, R.D., and Miller, M. 1993. A deformable neuroanatomy handbook based on viscous fluid mechanics. In *27th Ann. Conf. on Inf. Sciences and Systems*, pp. 211–216.

Christensen, G.E. and Johnson, H.J. 2001. Consistent image registration. *IEEE Trans. on Medical Imaging*, 20:568–582.

Cullen, M. and Purser, R. 1984. An extended Lagrangian theory of semigeostrophic frontogenesis. *J. Atmos. Sci.*, 41:1477–1497.

Dacorogna, B. and Moser, J. 1990. On a partial differential equation involving the Jacobian determinant. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 7:1–26.

Fry, D. 1993. Shape recognition using metrics on the space of shapes. Ph.D. Thesis, Harvard University.

Gangbo, W. 1994. An elementary proof of the polar factorization of vector-valued functions. *Arch. Rational Mechanics Anal.*, 128:381–399.

Gangbo, W. and McCann, R. 1996. The geometry of optimal transportation. *Acta Math.*, 177:113–161.

Gangbo, W. and McCann, R. 1999. Shape recognition via Wasserstein distance. Technical Report, School of Mathematics, Georgia Institute of Technology.

Haker, S., Angenent, S., Tannenbaum, A., and Kikinis, R. 2000. Nondistorting flattening maps and the 3D visualization of colon CT images. *IEEE Trans. of Medical Imaging*.

Haralick, R. and Shapiro, L. 1992. *Computer and Robot Vision*. Addison-Wesley: New York.

Hinterberger, W. and Scherzer, O. 2001. Models for image interpolation based on the optical flow. *Computing*, 66:231–247.

Kaijser, T. 1998. Computing the Kantorovich distance for images. *Journal of Mathematical Imaging and Vision*, 9:173–191.

Kantorovich, L.V. 1948. On a problem of monge. *Uspekhi Mat. Nauk.*, 3:225–226.

Knott, M. and Smith, C. 1984. On the optimal mapping of distributions. *J. Optim. Theory*, 43:39–49.

Levina, E. and Bickel, P. 2001. The earth mover's distance is the Mallow's distance: Some insights from statistics. In *Proceedings IEEE Int. Conf. on Computer Vision*, vol. 2, pp. 251–256.

McCann, R. 2001. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608.

McCann, R. 1997. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179.

Miller, M., Christensen, G., Amit, Y., and Grenander, U. 1992. Mathematical textbook of deformable neuroanatomies. *Proc. National Academy of Science*, 90:11944–11948.

Moser, J. 1965. On the volume elements on a manifold. *Trans. Amer. Math. Soc.*, 120:286–294.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn., Cambridge University Press: Cambridge U.K.

Rachev S. and Rüschendorf, L. 1998. *Mass Transportation Problems*, vols. I and II, *Probability and its Applications*. Springer: New York.

Rubner, Y. 1999. Perceptual metrics for image database navigation. Ph.D. Thesis, Stanford University.

Rubner, Y., Tomasi, C., and Guibas, J. 1998. The earth mover's distance as a metric for image retrieval. Technical Report STAN-CS-TN-98-86, Department of Computer Science, Stanford University.

Strang, G. 1986. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press: Wellesley, MA.

Taylor, M. 1996. *Partial Differential Equations III*. Springer-Verlag: New York.

Thirion, J.-P. 1995. Fast non-rigid matching of non-rigid images. In *Medical Robotics and Computer Aided Surgery (MRCAS' 95)*, Baltimore, p. 4754.

Toga, A. 1999. *Brain Warping*. Academic Press: San Diego.