

Published in final edited form as:

SIAM J Sci Comput. 2010 ; 32(1): 197–211. doi:10.1137/080730238.

AN EFFICIENT NUMERICAL METHOD FOR THE SOLUTION OF THE L_2 OPTIMAL MASS TRANSFER PROBLEM*

Eldad Haber[†], Tauseef Rehman[‡], and Allen Tannenbaum[§]

[†]Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322 (haber@mathcs.emory.edu).

[‡]School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (tauseef@ece.gatech.edu).

[§]Schools of Electrical & Computer and Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (tannenba@ece.gatech.edu).

Abstract

In this paper we present a new computationally efficient numerical scheme for the minimizing flow approach for the computation of the optimal L_2 mass transport mapping. In contrast to the integration of a time dependent partial differential equation proposed in [S. Angenent, S. Haker, and A. Tannenbaum, *SIAM J. Math. Anal.*, 35 (2003), pp. 61–97], we employ in the present work a direct variational method. The efficacy of the approach is demonstrated on both real and synthetic data.

Keywords

optimal mass transport; sequential quadratic programming; variational formulation

1. Introduction

Optimal mass transport is an important problem with applications in econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimization, expert systems, and meteorology [30,36]. The problem was first formulated by the civil engineer 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. Much later the problem was extensively analyzed by Kantorovich [23], and so is now known as the Monge–Kantorovich (MK) problem.

There are several formulations of the problem [1,30,36] of varying degrees of generality. We recall here the formulation of the Monge–Kantorovich problem for smooth densities and domains in Euclidean space. For more general measures, see [1]. Let Ω_0 and Ω_1 be two

*This work was partially supported by grants from NSF, AFOSR, ARO, as well as by a grant from NIH (NAC P41 RR-13218) through the Brigham and Women’s Hospital. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from <http://nihroadmap.nih.gov/bioinformatics>.
<http://www.siam.org/journals/sisc/32-1/73023.html>

© 2010 Society for Industrial and Applied Mathematics

Current address: Department of Electrical Engineering, Technion, Haifa, Israel.

AMS subject classifications. 76M20, 76M25, 76M30, 90C20

diffeomorphic connected subdomains of \mathbb{R}^d , and let μ_0, μ_1 be Borel measures on Ω_0 and Ω_1 , each with a strictly positive density function $\mu_0(x) \geq \mu_{\text{low}}^0 > 0$ and $\mu_1 \geq \mu_{\text{low}}^1 > 0$, respectively. Assume

$$\mu_0(\Omega_0) = \mu_1(\Omega_1),$$

i.e.,

$$\int_{\Omega_0} \mu_0(x) dx = \int_{\Omega_1} \mu_1(x) dx,$$

so that the same total mass is associated with Ω_0 and Ω_1 .

The version of the Monge–Kantorovich problem of interest in this work may be expressed as follows:

$$\min M(u) := \frac{1}{2} \int_{\Omega} \mu_0(x) \rho(u, x) dx, \quad (1.1a)$$

$$\text{s.t. } c(u) = \det(\nabla u) \mu_1(u(x)) - \mu_0(x) = 0, \quad (1.1b)$$

where u is a $C^{1,\alpha}$ diffeomorphism from $\Omega \rightarrow \Omega$ and $\rho(u, x)$ is a distance function between x and u . Here we treat only the case

$$\rho(u, x) = |u - x|^2.$$

The constraint $c(u) = 0$ (the Jacobian equation) is often referred to as the mass preserving (MP) property.

Even with a simple, quadratic distance function, this is a highly nonlinear equality constrained optimization problem. There is extensive analysis as to the existence, uniqueness, and properties of the solution (see, for example, [1,11,36] and the references therein). However, while there is a large body of literature which deals with the analysis of the problem, there is a surprisingly small number of papers that deal with the solution of the problem, and even a smaller number of papers that deal with efficient *numerical* solutions of the problem [8,4,2,29,10].

Among the papers that deal with the numerical solutions is the paper of Benamou and Brenier [4]. Their paper reconstructs an optimal path from μ_0 to μ_1 by solving an optimization problem with a space-time transport partial differential equation (PDE) as a constraint. Their approach is particularly useful if the transportation path is needed. Its disadvantage is that it enlarges the dimension of the problem by introducing a space-time control problem. An interesting geometric method has also been formulated by Cullen and Purser [9].

A very different approach which is closer to that of the present work has been proposed by Angenent, Haker, and Tannenbaum (AHT) [2]. This approach reconstructs the transformation directly. The idea of the AHT method is to obtain an initial MP

transformation $u_0(x)$, and then replace the original problem with the following optimization problem:

$$\min M(s) = \frac{1}{2} \int_{\Omega} \mu_0(x) |u_0(s^{-1}) - x|^2 dx, \quad (1.2a)$$

$$\text{s.t. } c(s) = \det(\nabla s) \mu_0(s(x)) - \mu_0(x) = 0, \quad (1.2b)$$

where $s \in \mathcal{C}^{1,\alpha}$ is an MP mapping from μ_0 to itself. It is well-known that the optimization problems (1.1) and (1.2) are equivalent [25]. The iteration starts with $s = Id$, and since a composition of mass preserving maps is also mass preserving, the authors assume that the MP constraint is valid throughout the optimization process. This allows them to obtain a time dependent PDE with the MP constraint as its invariant, that converges (in functional space) to the solution of the problem. In order to implement their method in a discrete setting, the authors have employed a forward Euler scheme for time stepping and a first order finite difference discretization of the spatial derivatives.

What is essentially done via the AHT methodology is to construct the so-called *polar factorization* [6,13,14,25] of the initial MP mapping u_0 , thus obtaining $u = \nabla \phi$, where ϕ is some scalar function. There are two main shortcomings to their numerical approach. First, a robust method to obtain an initial MP mapping is needed. The authors have used 1D integration which leads to an MP mapping which is highly irregular. See also [27] for an approach to constructing the initial MP mapping. Thus, the solution of the optimization problem may be far from the initial point, and as a result many iterations are needed to convergence. Second and more important, the AHT method updates the transformation in a space which is tangential to the linearized MP constraint. While such an update is tangential to the MP constraint for a small (infinitesimal) correction, it is not tangential to any finite step used in the *discrete* setting. Since the MP constraint is never specifically enforced during the *discrete* AHT minimization process it “drifts,” that is, the final mapping may not be mass preserving and therefore the discrete PDE may not converge to a discrete approximation of the problem.

In this paper we develop a new, straightforward technique that does not suffer from the aforementioned difficulties. First, we modify the objective function in order to obtain better attraction to the MP solution. Second, we employ a conservative discretization of the objective function and the constraint, and finally, we use the state-of-the-art numerical optimization procedure and linear algebra solvers to obtain the solution. The method is illustrated via some examples of medical image registration. These examples are included simply to elucidate the proposed approach on some actual data. It is not our intention to propose optimal mass transport as a general method to medical image registration. See also [21,20,39] for a much more extensive discussion of the use of optimal mass transport and elastic image registration.

The rest of the paper is organized as follows. In section 2 we review the mathematical derivation of our approach. In section 3, we discuss the discretization scheme. In section 4 we discuss the solution of the discrete problem. In section 5, we perform numerical experiments in two and three dimensions that demonstrate the effectiveness of our method. Finally, in section 6 we summarize our work.

2. Problem reformulation

We give now the key ideas underlying our approach to solving the L^2 Monge–Kantorovich problem. Accordingly, suppose that we have an initial transformation w that is not mass preserving, and that we would like to obtain a transformation u which is mass preserving from w . This can be done by projection, that is, by solving the optimization problem

$$\min_u \|u - w\|_{\mathcal{H}}^2, \quad (2.1)$$

$$\text{s.t. } c(u)=0, \quad (2.2)$$

where $\|\cdot\|_{\mathcal{H}}$ is some appropriate norm. It is interesting to note that the MK problem is nothing but a projection from the identity transformation $w(x) = x$ with respect to the μ_0 -weighted L_2 norm $\|\cdot\|_{\mu_0}$ (see (1.1a)–(1.1b)).

Thus if we simply project to the MP constraint with the μ_0 norm, we should be able to obtain a *local* minimizer for (2.1). However, such a projection may not be the global minimizer. Nevertheless, while we cannot guarantee *global* convergence of the projection process, we can increase the chances that the local minimizer of (2.1) is also a global minimizer of the MK problem. To do that, we recall that the global minimizer has the property that $u = \nabla\phi$ [1,36]; therefore, $\nabla \times u = 0$. Thus we propose to perform the projection by solving the following optimization problem:

$$\mathbf{PROJ} \quad \min_u \frac{1}{2} \int_{\Omega} (\mu_0(x)|u - w|^2 + \beta |\nabla \times u|^2) dx, \quad (2.3a)$$

$$\text{s.t. } c(u)=0, \quad (2.3b)$$

where $\beta > 0$. While the term $\beta |\nabla \times u|^2$ does not change the global minimum, it does give a bias towards a curl-free solution when solving for the projection, and thus has a better chance to converge to the global minimum.

One can check that the conditions for a minimum lead to the following nonlinear system of equations:

$$\mu_0(u - w) + \beta \nabla \times \nabla \times u + p \det(\nabla u) \nabla \mu_0(u) - \nabla \cdot (\det(\nabla u) (\nabla u)^{-\top} p), \quad (2.4a)$$

$$\det(\nabla u) \mu_0(u) = 0, \quad (2.4b)$$

where p is the Lagrange multiplier.

The optimization problem **PROJ** is a continuous constrained optimization problem. There are many techniques to obtain the solution to such problems. In the following, we will consider a version of sequential quadratic programming (SQP) which is a commonly used methodology for such constrained optimization problems [28]. The advantage of SQP type

methods is that they are mesh-independent. That is, the number of iterations to convergence is independent of the mesh size [37].

However, before we consider the explicit solution process, we should first describe two possible approaches. In the “discretize-then-optimize” approach, we discretize the projection (2.3), and then solve a discrete optimization problem. In the “optimize-then-discretize” approach, we first write the Euler–Lagrange equations (2.4) in a continuous setting, and then discretize them in order to obtain a system of discrete nonlinear equations. The two approaches are summarized in [17]. The advantage of the discretize-then-optimize approach is that common optimization algorithms can be used with little modification. This is particularly important if we wish to obtain convergence to a minimum from a possibly distant starting point. Recent work in the field [28,7] suggests several mechanisms to achieve this goal, which we explore in some detail below. Particular care must be taken so that the inner products and norms are discretized appropriately [22]. We therefore proceed with the discretization of the optimization problem.

We should note that although we use the discretize-then-optimize approach, the discrete nonlinear system and its linearization can be thought of as a discretization of the Euler–Lagrange equations (2.4). As such we need efficient solvers to systems which involve the **curl** operator and the linearization of the constraints.

3. Discretization

In order to discretize the problem, we define the Lagrangian

$$\mathcal{L}(u, p) = \frac{1}{2} \int_{\Omega} (\mu_0(x) |u - x|^2 + \beta |\nabla \times u|^2) dx + (p, c(u)), \quad (3.1)$$

where p is a Lagrange multiplier.

Certain key applications (for example, in medical image processing) have data which are discretized on a regular grid. We therefore construct our discretization based on a finite volume/difference approach. Rather than working with the deformation field u , we work with the update or displacement $v = u - x$. This yields numerically smaller perturbations. Note also that $\nabla \times u = \nabla \times v$, and thus the projection (2.3) does not change.

We divide the domain, Ω , into $n_1 \times \cdots \times n_d$ cells, each of size $h_1 \times \cdots \times h_d$, where d is the dimension of the problem. For simplicity we discretize all the components of v at the nodes of each cell to obtain d grid functions $\hat{v}^1, \dots, \hat{v}^d$. The Lagrange multiplier p is discretized at the cell centers. In 2D, we denote by $\hat{v}_{j,k}^i$ the i th grid function discretized at the $(x_1)_j, (x_2)_k$ node, and by $\hat{p}_{j+\frac{1}{2}, k+\frac{1}{2}}$ the Lagrange multiplier discretized at the cell center. The discretized quantities are plotted in Figure 1.

We note that a staggered discretization for the displacement can also be used [19]. Similar to problems that are derived from computational fluid dynamics and electromagnetics, such a discretization has very nice numerical properties. However, it is less simple to implement and requires careful treatment of boundary conditions. A discussion about the staggered versus the unstaggered discretization can be found in [35].

We now consider the different numerical approximations needed in order to obtain a discrete approximation to the continuous system. A consistent discretization for Δ , $\nabla \times$ and a consistent discretization of $\nabla u = I + \nabla v$ are needed. There are a number of possible

discretizations that lead to a well-posed system. Here we derive the explicit discretization in 2D. The extension to 3D is straightforward.

3.1. Discretization of Δ and $\nabla \times$

In order to obtain a consistent discretization of the Laplacian, we use a standard discretization (5 point stencil in 2D and 7 point stencil in 3D) with Dirichlet boundary conditions.

To compute the discretization of the **curl**, we use “short differences” in one direction (see [3, p. 11]) and average them (or the differences) in the other direction. For example, in 2D we obtain

$$(C\widehat{v})_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{\widehat{v}_{i,j+1}^1 - \widehat{v}_{i,j}^1 + \widehat{v}_{i+1,j+1}^1 - \widehat{v}_{i+1,j}^1}{2h_1} - \frac{\widehat{v}_{i+1,j}^2 - \widehat{v}_{i,j}^2 + \widehat{v}_{i+1,j+1}^2 - \widehat{v}_{i,j+1}^2}{2h_2} + \mathcal{O}(h^2), \quad (3.2)$$

where we denote by C the **curl** matrix. It is easy to verify the second order accuracy of this discretization.

3.2. Discretization of the MP constraint

To discretize the MP constraint we use a finite volume discretization, and so integrating over a given region Ω_j

$$\int_{\Omega_j} c(v) dx = \int_{\Omega_j} \det(I + \nabla v) \mu_0(x+v) dx,$$

we see that two approximations are needed. First, we need to approximate $\det(I + \nabla v)$, and second, we need to approximate $\mu_0(x+v)$. There are a number of options for discretizing the constraint. In our application, μ_0 is typically discretized in cell centers; therefore we choose the domain Ω_j as a cell in our discretization.

In order to approximate $\mu_0(x+v)$ in cell centers, we use linear interpolation; see [26] for details. We denote this approximation by $\mu_0(x_h + \widehat{v})$, where x_h is the discretization of $x = [x_1,$

$x_2]$ on the cell centers, $(i+\frac{1}{2}, j+\frac{1}{2})$.

In order to approximate $\det(I + \nabla v)$ at the cell centers, we first compute all the derivatives in cell centers as follows:

$$(v_{x_1}^k)_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{\widehat{v}_{i+1,j}^k - \widehat{v}_{ij}^k + \widehat{v}_{ij+1}^k - \widehat{v}_{i+1,j+1}^k}{2h} + \mathcal{O}(h^2), \quad (3.3a)$$

$$(v_{x_2}^k)_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{\widehat{v}_{ij+1}^k - \widehat{v}_{ij}^k + \widehat{v}_{i+1,j+1}^k - \widehat{v}_{i+1,j}^k}{2h} + \mathcal{O}(h^2), \quad k=1, 2. \quad (3.3b)$$

We also denote the block diagonal matrix

$$\mathbf{GRAD}\mathbf{v} = \text{blkdiag} \left(\begin{array}{cc} (v_{x_1}^1)_{i+\frac{1}{2},j+\frac{1}{2}} & (v_{x_2}^1)_{i+\frac{1}{2},j+\frac{1}{2}} \\ (v_{x_1}^2)_{i+\frac{1}{2},j+\frac{1}{2}} & (v_{x_2}^2)_{i+\frac{1}{2},j+\frac{1}{2}} \end{array} \right),$$

$$i=1, \dots, n_1 - 1, j=1, \dots, n_2 - 1.$$

This matrix is the discrete analog to the matrix $\nabla \mathbf{v}$ at each cell center. For the MK problem to be well defined, the matrix $I + \nabla \mathbf{v}$ must be SPD everywhere.

Let D_j be the discretization of the j th derivative. The discrete MP constraint for \mathbf{v} then reads

$$(\mathbf{1} + D_1 \widehat{v}_1 + D_2 \widehat{v}_2 + (D_1 \widehat{v}_1) \odot (D_2 \widehat{v}_2) - (D_2 \widehat{v}_1) \odot (D_2 \widehat{v}_1)) \odot \mu_0(x_h + \widehat{\mathbf{v}}) = \mu_1(x_h), \quad (3.4)$$

where \odot is the Hadamard product. It is important to further study this discretization and its properties. Typically to study any discretization we study the consistency and the stability of our discretization.

3.2.1. Consistency of the discretization—To study the consistency of the discretization, we employ the following lemma.

LEMMA 1. *Let $\widehat{\mathbf{v}} = (\widehat{v}^1, \dots, \widehat{v}^d)$ be grid functions of a sufficiently smooth field \mathbf{v} , and let $\mathbf{GRAD}\mathbf{v}$ be approximated for the matrix $\nabla \mathbf{v}$. Then, if $I + \nabla \mathbf{v}$ has positive real eigenvalues with $\text{Re}(\lambda_{\min}) \geq \rho > 0$, the matrix $\mathbf{GRAD}\mathbf{v}$ also has positive real eigenvalues for a sufficiently small h .*

Proof. To prove the lemma, we note that the matrix is a block 2×2 matrix. This implies that the eigenvalues of the matrix $I + \mathbf{GRAD}\mathbf{v}$ can be computed blockwise, where each of the blocks corresponds to a different cell in the discrete grid. Since the approximation of the derivatives is $\mathcal{O}(h^2)$, the eigenvalues of the matrix $I + \mathbf{GRAD}\mathbf{v}$ converge to the eigenvalues of $I + \nabla \mathbf{v}$.

It is important to note that the above lemma does not predict the order of convergence. Unfortunately, very small perturbations in the discrete grid function \mathbf{v} can lead to large perturbations in the eigenvalues of $\mathbf{GRAD}\mathbf{v}$. Consider, for example, the case in which the approximation to $\mathbf{GRAD}\mathbf{v}$ at some cell is

$$A = I + \mathbf{GRAD}\mathbf{v} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix},$$

and that we have a numerical error of the form εE where

$$E = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Then, it is easy to verify that

$$\lambda(A + \varepsilon E) = 1 \pm \sqrt{\varepsilon a}.$$

Thus, for large enough a the eigenvalues of the perturbed system can be very different from the eigenvalues of the continuous one.

Nevertheless, the above lemma is important because it implies that the overall (discrete) problem is consistent for sufficiently small h . It is also important because it gives certain hints about possible difficulties with the numerical solution. It is well known that the continuous problem can be ill-posed if the real parts of the eigenvalues of ∇u are not positive, that is, $\text{Re}(\nabla u) > 0$ (see [11]). Lemma 1 requires h to be sufficiently small in order to obtain this behavior. If the displacement field is not sufficiently smooth or h is not sufficiently small, then we may encounter negative eigenvalues in the discrete approximation of u . In this case the iteration can stall or even diverge. We therefore propose to check the eigenvalues of $I + \mathbf{GRAD} \mathbf{v}$ throughout the iteration. This can be done in $\mathcal{O}(n)$ work due to the block structure of the system.

There are a number of ways to correct for negative eigenvalues. Denote the Schur

decomposition of the gradient matrix at the $i+\frac{1}{2}, j+\frac{1}{2}$ cell by

$$(I + \mathbf{GRAD} \mathbf{v})_{i+\frac{1}{2}, j+\frac{1}{2}} = V_{i+\frac{1}{2}, j+\frac{1}{2}} \Lambda_{i+\frac{1}{2}, j+\frac{1}{2}} V_{i+\frac{1}{2}, j+\frac{1}{2}}^T,$$

where all the matrices are 2×2 (in 2D). For the sake of concreteness, assume that the first eigenvalue $\Lambda_{i+\frac{1}{2}, j+\frac{1}{2}}^1 \leq 0$. A simple modification of our method is simply to inflate Λ_{ij}^1 in such a manner that it is slightly larger than 0, for example, to h . We have taken precisely this approach to avoid vanishing or negative eigenvalues that may lead to divergence of the method.

3.2.2. Stability of the discretization—Although the discretization of the MP constraint is consistent and second order accurate, it is *not* a stable discretization. It is easy to verify that the derivatives have a nontrivial null-space known as a *checkerboard null-space*. This is not very surprising. Indeed, consider the simple case of a uniform mass in which one linearizes the perturbation around it. The Taylor expansion yields

$$\det(I + \nabla \delta v) \approx 1 + \nabla \cdot \delta v.$$

Thus, for the starting point $u = x$ and $v = 0$ the first iterate gives a system similar to the Stokes system. It is well known that nonstaggered discretizations of the Stokes equation are not stable [35]. Nevertheless, simple stabilization techniques can be used to obtain a well-posed discrete system. The most common approach is a penalty method [35], where an artificial viscosity is added to the Lagrange multiplier. This additional term stabilizes the system without affecting the overall accuracy of the problem. We return to this issue in the next section.

3.3. Discretization of the objective function

In order to discretize the objective function on a regular grid, two quantities are necessary. First, we need to approximate $\int \mu_0 |v|^2 dx$ on Ω_j , and second, the approximation of $\int |\nabla \times v|^2 dx$ is required.

Assuming μ_0 is given in cell centers, we obtain

$$\int_{\Omega_k} \mu_0 |v|^2 dx = h^2 \frac{1}{4} \mu_{0,(i+\frac{1}{2},j+\frac{1}{2})} \sum_{\ell=1}^2 (\widehat{v}_{ij}^\ell + \widehat{v}_{i+1,j}^\ell + \widehat{v}_{ij+1}^\ell + \widehat{v}_{i+1,j+1}^\ell)^2 + \mathcal{O}(h^2). \quad (3.5)$$

Summing over the cells, we obtain a second order accurate discretization that can be written in matrix form as

$$\int_{\Omega} \mu_0 |v|^2 dx = \widehat{v}^\top M_{\mu_0} \widehat{v} + \mathcal{O}(h^2), \quad (3.6)$$

where M_{μ_0} is a μ_0 weighted mass matrix.

Since the discretization of the **curl** is centered at the cell centers, it is straight-forward to use the **curl** matrix C to obtain

$$\int_{\Omega} |\nabla \times v|^2 dx = h^2 \widehat{v}^\top C^\top C \widehat{v} + \mathcal{O}(h^2). \quad (3.7)$$

4. Computation of a step

Given the properties of our discretization we now discuss the modification of the Lagrangian and the computation of a step.

In order to overcome the problem of stability in the discrete constraint, we use a well-known strategy utilized in computational fluid dynamics [35] and add a small penalty term to the discrete Lagrangian, defined as

$$\mathcal{L} = \frac{1}{2} \widehat{v}^\top M_{\mu_0} \widehat{v} + \frac{\beta}{2} \widehat{v}^\top C^\top C \widehat{v} + p^\top c(v) - \frac{\gamma}{2} \|\nabla_h \widehat{p}\|^2, \quad (4.1)$$

where γ is an $\mathcal{O}(h^2)$ parameter. The latter term is added to the problem to overcome the stability problem. For a simple constraint $c(v) = \nabla \cdot v$, it is possible to obtain an analytic expression for the optimal γ (see, for example, [35]). However, when the constraint depends on v , such an optimal expression is not attainable. We therefore set $\gamma = h_1 h_2$, which now yields a consistent and stable approximation to the Lagrangian.

To solve the problem, we use a version of inexact SQP [7]. The bottleneck in this methodology is the solution of the so-called Karush–Kuhn–Tucker (KKT) system

$$\begin{pmatrix} H & c_u^\top \\ c_u & -S \end{pmatrix} \begin{pmatrix} \delta \widehat{v} \\ \delta \widehat{p} \end{pmatrix} = \text{RHS}, \quad (4.2)$$

where H is an approximation to the second derivatives of the Lagrangian and c_u is the Jacobian of the constraint. Here we used $H = M_\mu + \beta C^\top C$. This approximation to the Hessian yields a semipositive definite (SPD) approximation to the (1,1) block and therefore guarantees a descent direction. We have modified standard SQP algorithms and introduced the regularization matrix $S = h_1 h_2 \Delta_h > 0$ in our numerical procedure.

To solve the KKT system, we have taken a number of steps. First, we have used an augmented Lagrangian approach [15] and generated the modified system

$$\begin{pmatrix} H_{ag} & c_u^\top \\ c_u & -S \end{pmatrix}, \quad (4.3)$$

where $H_{ag}=H+c_u^\top c_u$. The augmentation helps to deal with the null-space of the **curl** operator and is commonly used in computational electromagnetics [18,16].

Next we use a block preconditioner similar to the one discussed in [33]. It is well known that an optimal preconditioner for the KKT system is [5]

$$\begin{pmatrix} H_{ag} & c_u^\top \\ 0 & S_c \end{pmatrix}, \quad (4.4)$$

where S_c is the Schur complement

$$S_c=c_u H_{ag}^{-1} c_u^\top + S.$$

The matrix H_{ag} is derived from an elliptic operator; therefore it is possible to approximate its inverse using standard multigrid methods. To approximate the Schur complement matrix we drop the matrix S , which is $\mathcal{O}(h^2)$, and use the pseudoinverse of c_u to obtain

$$S_c^{-1} \approx c_u^\dagger H_{ag} c_u^\ddagger.$$

The computation of c^\dagger times a vector z requires the solution of a system of the form

$$c_u c_u^\top q = z.$$

Since $c_u c_u^\top$ is a discretization of an elliptic operator, this can be done using a standard multigrid technique. The numerical properties and performance of this approximation will be studied elsewhere. In our numerical experiments we have noticed that the number of iterations of the generalized minimal residual method (GMRES) [32] that was needed to solve the KKT system was almost mesh independent.

5. Numerical experiments

To test the performance of our method we considered three illustrative examples. First, we used a known deformation field to deform a synthetic image, and we verified that our algorithm could recover the deformation field accurately. We will present results both for 2D and 3D cases. We also tested the algorithm for the registration of 3D brain magnetic resonance image (MRI) datasets as an example application in medical imaging. Regarding this latter application, we certainly do not propose optimal mass transport as a general approach to image registration. Optimal mass transport may be regarded as an intensity-based approach to elastic registration, and as such is closely related to certain optical flow based methods [34,38]. It makes sense for there to be imagery in which the intensity can be related somehow to some density, which is the case for MRI data. Here the intensity is proportional to the proton density (basically water content) of the given tissue.

5.1. Synthetic examples

Since the solution of the problem is also the solution of the Monge–Ampère equation [11], it is easy to construct an analytic example. In order to do this, we consider the following function:

$$\phi(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) + c \cdot e^{-\frac{1}{2}(x_1 - \frac{1}{2})^2 / \sigma_1^2} \cdot e^{-\frac{1}{2}(x_2 - \frac{1}{2})^2 / \sigma_2^2} \in [0, 1]^2,$$

where c , σ_1 , and σ_2 are parameters chosen to create a unique deformation field. Differentiating ϕ with respect to $x = (x_1, x_2)$, we obtain $u = (u_1, u_2)$,

$$\begin{aligned} u_1 &= x_1 - c \cdot ((x_1 - 0.5)/\sigma_1^2) \cdot e^{-\frac{1}{2}(x_1 - \frac{1}{2})^2 / \sigma_1^2} \cdot e^{-\frac{1}{2}(x_2 - \frac{1}{2})^2 / \sigma_2^2}, \\ u_2 &= x_2 + c \cdot ((x_2 - 0.5)/\sigma_2^2) \cdot e^{-\frac{1}{2}(x_1 - \frac{1}{2})^2 / \sigma_1^2} \cdot e^{-\frac{1}{2}(x_2 - \frac{1}{2})^2 / \sigma_2^2}. \end{aligned}$$

We employed a standard 2D phantom image from MATLAB to serve as $\mu_1(x_1, x_2)$, and define $\mu_0(x_1, x_2)$ by

$$\mu_0 := \det(\nabla u) \mu_1(u). \quad (5.1)$$

We computed $\mu_0(x_1, x_2)$ via linear interpolation. The deformed grid, deformation vector field, μ_0 , and μ_1 thus obtained are shown in Figure 2.

We next inputted the μ_1 and μ_0 obtained in this manner into our solver to find the transformation u using our algorithm.

We also performed the same experiment for a 3D brain MRI dataset. The deformation field from the 2D example was symmetrically extended in the third dimension and the deformed image was computed in the same way. For both experiments we use $\beta = 10^2$. We terminated our algorithm after 100 iterations or when the curl of the solution was 4 orders of magnitudes smaller than its initial size (in the infinity norm).

Table 1 shows the ∞ -norm of the error between the known and compute deformation fields at different grid sizes. The tables clearly demonstrate quadratic convergence of our method to the true solution, which is expected from the discretization error used in our numerical approximations.

To demonstrate the efficiency of our algorithm, we present in Table 2 the number of projection steps used for each of the examples on each level. Note that we have not used a multilevel procedure that could speed up calculations. This was done in order to demonstrate mesh independence of the method. The incorporation of a multilevel method is beyond the scope of the present paper and will be done in future work.

Table 2 clearly demonstrates the mesh-independence property of the modified SQP algorithm.

5.2. A 3D brain example

We also tested our approach for multimodal registration of a 3D brain atlas to an MRI. Our goal is the identification of cortical structures by mapping a publicly available atlas [24] to the scan of a patient. The MRI of the atlas is a spoiled gradient recalled image acquired on a 1.5-Tesla General Electric Signa System (GE Medical Systems, Milwaukee) with 256×256

$\times 124$ voxels and voxel dimension of $0.92 \times 0.92 \times 1.5$ mm. The patient scan is an MPRAGE acquired on a Siemens 3T long bore machine using an 8 channel head coil. The resolution of the scan is $256 \times 256 \times 144$ with voxel dimension $0.54 \times 0.54 \times 1.0$ mm (see Figure 3(b)).

The parcellation of the cortex can be encoded by partitioning the boundary between cortex and white matter into anatomical regions [12]. The label map of cortical structures can then be inferred from this partition by propagating the labeling along the boundary to the entire cortex. The pipeline described below will apply this concept for the parcellation of the cortex to the high resolution scan.

The input of the pipeline consists of the atlas, the high resolution scan, as well as a segmentation of the scan into the major tissue classes. In the first step, we coarsely align the atlas to the image data using the B-spline implementation by Rohlfinger [31] with a final spacing of the grid nodes of 2.5 mm. This results in a coarse alignment of the scans. The algorithm has difficulties in mapping the folds of the white matter due to the inherent constraints of the B-spline representation. We then reduce the atlas to the white matter including the parcellation of the cortex along the boundary between gray and white matter (see Figure 2(a)). Afterwards, we refine the alignment of this new atlas to the white matter of the high resolution scan using our optimal mass transport registration approach.

Registration using optimal mass transport is a highly flexible approach that is, unlike B-splines, not constrained to a set of control points. The intensities in the two input datasets are first normalized and rescaled to make sure that both have the same total mass. The white matter registration with the proposed algorithm took just 12 iterations to converge with 2 iterations of the projection to constraint per iteration. The $\nabla \times u$ (convergence metric) was reduced to an order of 10^{-3} indicating an optimal map. Figure 2(c) shows the resampled images with 3D views of the corresponding deformation grid in Figure 4. The difference (see Figure 2(d)) between the target indicated in Figure 2(b) and the resampled image shows that our approach accurately aligned the folds.

After this local alignment, the folds of the atlas should perfectly align with those of the high resolution scan. The parcellation of the folds of the atlas, therefore, also encodes the parcellation of the same region in the high resolution scan. We then complete the cortex parcellation of the high resolution scan by confining the Voronoi diagram of the aligned atlas to the gray matter mask of the high resolution scan. The results in Figure 5 show the corresponding segmentation when applying the deformation map of the B-spline registration and our approach to the label map of [24], and propagating the labels to the cortex via the Voronoi diagram.

6. Summary

In this paper, we have investigated a new method for the computation of the optimal L_2 mass preserving mapping, and derived a novel numerical framework for its efficient computation. We proposed a direct variation approach, and showed how one may obtain a descent direction. A second order accurate discretization to the problem was presented. Further, we presented several illustrative numerical experiments in both two and three dimensions which indicated the effectiveness of our method. There are a number of key directions for future research. In particular, we plan to improve our preconditioners to the KKT system, which is the bottleneck of the computation, as well as use adaptive multilevel mesh refinement. Finally, there are several possible applications of our optimal mass transport technique which we intend to explore. These include texture mappings in computer graphics, and visual tracking in controlled active vision.

Acknowledgments

This author's research was supported by NSF grants DMS 0724759, CCF-0728877, and CCF-0427094, and DOE grant DE-FG02-05ER25696.

REFERENCES

1. Ambrosio, L. Mathematical Aspects of Evolving Interfaces, Lecture Notes in Math. 1812. Berlin: Springer-Verlag; 2003. Lecture notes on optimal transport theory; p. 1-52.
2. Angenent S, Haker S, Tannenbaum A. Minimizing flows for the Monge-Kantorovich problem. *SIAM J. Math. Anal* 2003;35:61–97.
3. Ascher, UM. Numerical Methods for Evolutionary Differential Equations. Philadelphia: SIAM; 2008.
4. Benamou J-D, Brenier Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math* 2000;84:375–393.
5. Benzi M, Golub GH, Liesen J. Numerical solution of saddle point problems. *Acta Numer* 2005;14:1–137.
6. Brenier Y. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math* 1991;64:375–417.
7. Byrd RH, Curtis FE, Nocedal J. An inexact SQP method for equality constrained optimization. *SIAM J. Optim* 2008;19:351–369.
8. Chartrand R, Vixie K, Wohlberg B, Boltt E. A gradient descent solution to the Monge-Kantorovich problem. *Appl. Math. Sci* 2009;3:1071–1080.
9. Cullen MJP, Purser RJ. An extended Lagrangian theory of semigeostrophic frontogenesis. *J. Atmospheric Sci* 1984;41:1477–1497.
10. Dean EJ, Glowinski R. Numerical methods for fully nonlinear elliptic equations of the Monge-Ampère type. *Comput. Methods Appl. Mech. Engrg* 2006;195:1344–1386.
11. Evans, LC. Current Developments in Mathematics. Boston: Int. Press; 1999. Partial differential equations and Monge-Kantorovich mass transfer; p. 65-126.
12. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. Automatically parcellating the human cerebral cortex. *Cerebral Cortex* 2004;14:11–22. [PubMed: 14654453]
13. Gangbo W. An elementary proof of the polar factorization of vector-valued functions. *Arch. Rational Mech. Anal* 1994;128:381–399.
14. Gangbo W, McCann RJ. The geometry of optimal transportation. *Acta Math* 1996;177:113–161.
15. Golub GH, Greif C. On solving block-structured indefinite linear systems. *SIAM J. Sci. Comput* 2003;24:2076–2092.
16. Greif C, Schötzau D. Preconditioners for the discretized time-harmonic Maxwell equations in mixed form. *Numer. Linear Algebra Appl* 2007;14:281–297.
17. Gunzburger, MD. Adv. Des. Control 5. Philadelphia: SIAM; 2003. Perspectives in Flow Control and Optimization.
18. Haber E, Ascher U. Fast finite volume simulation of 3D electromagnetic problems with highly discontinuous coefficients. *SIAM J. Sci. Comput* 2000;22:1943–1961.
19. Haber E, Modersitzki J. Numerical methods for volume preserving image registration. *Inverse Problems* 2004;20:621–638.
20. Haker, S.; Tannenbaum, A.; Kikinis, R. MICCAI 2001, Lecture Notes in Comput. Sci. 2208. Berlin-Heidelberg: Springer; 2001. Mass preserving mappings and surface registration; p. 120-127.
21. Haker S, Zhu L, Tannenbaum A, Angenent S. Optimal mass transport for registration and warping. *Int. J. Comput. Vis* 2004;60:225–240.
22. Heinkenschloss M, Vicente LN. Analysis of inexact trust-region SQP algorithms. *SIAM J. Optim* 2001;12:283–302.
23. Kantorovich LV. On a problem of Monge. *Uspekhi Mat. Nauk* 1948;3:225–226.

24. Kikinis R, Shenton ME, Iosifescu DV, McCarley RW, Saiviroonporn P, Hokama HH, Robatino A, Metcalf D, Wible CG, Portas CM, Donnino RM, Jolesz FA. A digital brain atlas for surgical planning, model-driven segmentation, and teaching. *IEEE Trans. Vis. Comput. Graph* 1996;2:232–241.
25. McCann RJ. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal* 2001;11:589–608.
26. Modersitzki, J. *Numer. Math. Sci. Comput.* Oxford: Oxford University Press; 2004. *Numerical Methods for Image Registration*.
27. Moser J. On the volume elements on a manifold. *Trans. Amer. Math. Soc* 1965;120:286–294.
28. Nocedal, J.; Wright, S. *Springer Ser. Oper. Res.* New York: Springer-Verlag; 1999. *Numerical Optimization*.
29. Oberman AM. Wide stencil finite difference schemes for the elliptic Monge-Ampère equation and functions of the eigenvalues of the Hessian. *Discrete Contin. Dyn. Syst. Ser. B* 2008;10:221–238.
30. Rachev, S.; Rüschendorf, L. *Probab. Appl. (N.Y.)*. New York: Springer-Verlag; 1998. *Mass Transportation Problems, Vol. I, Theory*.
31. Rohlfing T, Maurer CR Jr. Nonrigid image registration in shared-memory multi-processor environments with application to brains, breasts, and bees. *IEEE Trans. Inform. Technol. Biomed* 2003;7:16–25.
32. Saad Y, Schultz MH. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput* 1986;7:856–869.
33. Silvester D, Elman H, Kay D, Wathen A. Efficient preconditioning of the linearized Navier-Stokes equations for incompressible flow. *J. Comput. Appl. Math* 2001;128:261–279.
34. Toga, A. *Brain Warping*. San Diego: Academic Press; 1999.
35. Trottenberg, U.; Oosterlee, C.; Schuller, A. *Multigrid*. Orlando: Academic Press; 2001.
36. Villani, C. *Grad. Stud. Math.* 58. Providence, RI: AMS; 2003. *Topics in Optimal Transportation*.
37. Volkwein S. Mesh-independence for an augmented Lagrangian-SQP method in Hilbert spaces. *SIAM J. Control Optim* 2000;38:767–785.
38. Warfield, S.; Guimond, A.; Roche, A.; Bharatha, A.; Tei, A.; Talos, F.; Rexilius, J.; Ruiz-Alzola, J.; Westin, CF.; Haker, S.; Angenent, S.; Tannenbaum, A.; Jolesz, FA.; Kikinis, R. Advanced nonlinear registration algorithms for image fusion. In: Toga, A.; Mazziotta, J., editors. *Brain Mapping: The Methods*. 2nd ed.. San Diego: Academic Press; 2003. p. 661–690.
39. Zhu L, Yang Y, Haker S, Tannenbaum A. Optimal mass transport for registration and warping. *Int. J. Comput. Vis* 2004;60:225–240.

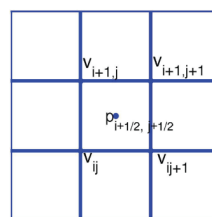


FIG. 1.
Discretization of the displacement and Lagrange multipliers.

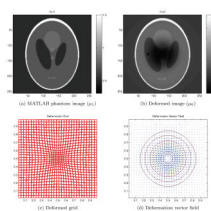


FIG. 2. 2D Synthetic Example. Two image pairs (a) and (b) with a known deformation, mapping one to the other. (c) shows the deformed grid, and (d) shows the deformation vector field with superimposed gradient contours of the determinant of the Jacobian.

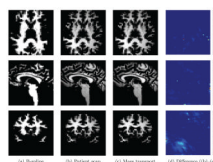


FIG. 3.
Registration results.

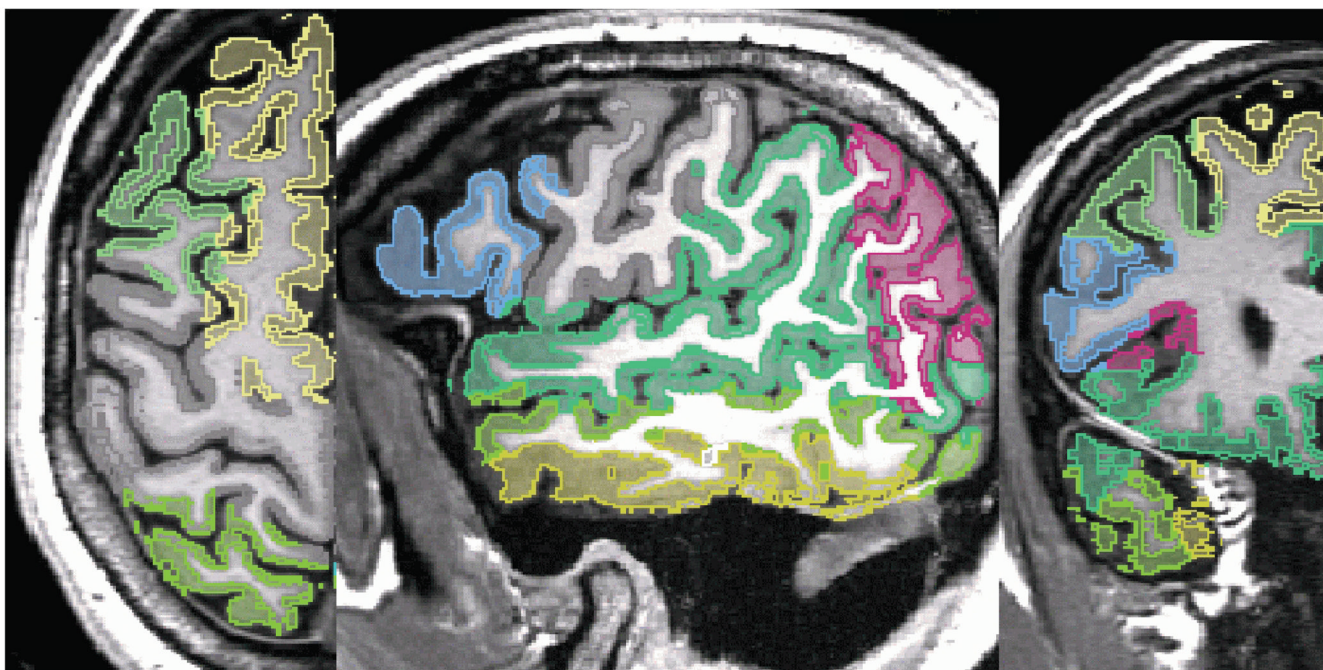


FIG. 4.
Parcellation results.

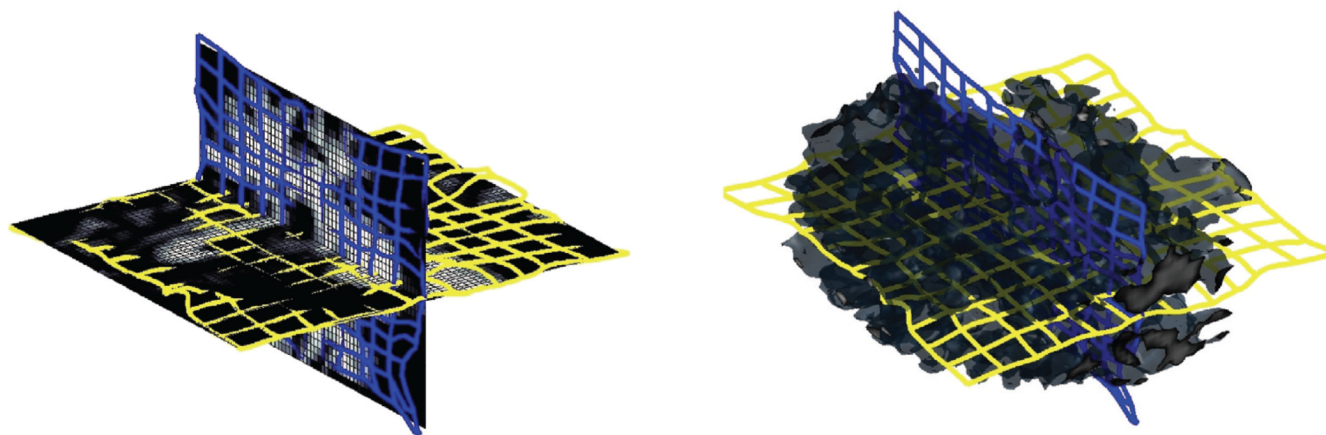


FIG. 5.
Deformed grid on white matter slices (left) and 3D volume (right).

TABLE 1

Convergence for the 2D and 3D experiments.

Grid size	Error (infinity norm) in 2D	Error (infinity norm) in 3D
2^{-3}	1.1×10^{-4}	1.2×10^{-2}
2^{-4}	2.5×10^{-5}	3.8×10^{-3}
2^{-5}	6.0×10^{-6}	9.0×10^{-4}
2^{-6}	2.4×10^{-7}	2.2×10^{-5}
2^{-7}	5.9×10^{-8}	5.1×10^{-6}
2^{-8}	1.2×10^{-8}	

TABLE 2

Number of projection iterations for the 2D and 3D experiments.

Grid size	# proj iter (2D)	# proj iter (3D)
2^{-3}	56	67
2^{-4}	32	46
2^{-5}	28	34
2^{-6}	27	35
2^{-7}	28	36
2^{-8}	26	35