Article

A Variational Surface-Evolution Approach to Optimal Transport over Transitioning Compact Supports with Domain Constraints

Anthony Yezzi

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA;
anthony.yezzi@ece.gatech.edu

Abstract: We examine the optimal mass transport problem in $\mathbb{R}^n$ between densities with transitioning compact support by considering the geometry of a continuous interpolating support boundary $\boldsymbol{\Gamma}$ in space-time within which the mass density evolves according to the fluid dynamical framework of Benamou and Brenier. We treat the geometry of this space-time embedding in terms of points, vectors, and sets in $\mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n$ and blend the mass density and velocity as well into a space-time solenoidal vector field $\mathbf{W} \mid \boldsymbol{\Omega} \to \mathbb{R}^{n+1}$ over a compact set $\boldsymbol{\Omega} \subset \mathbb{R}^{n+1}$. We then formulate a joint optimization for $\mathbf{W}$ and its support using the shaped gradient of the space-time surface $\boldsymbol{\Gamma}$ outlining the support boundary $\partial\boldsymbol{\Omega}$. This easily accommodates spatiotemporal constraints, including obstacles or mandatory regions to visit.

Keywords: optimal mass transport; fluid dynamical framework; active contours and surfaces; variational methods; Benamou and Brenier; obstacle avoidance; evacuation strategies

1. Introduction

Optimal mass transport (OMT) has a long history, beginning with Gaspard Monge [1] in 1781, and put into a more modern form solvable via linear programming by Leonid Kantorovich [2–7]. In recent years, OMT has undergone a huge surge, with many diverse applications including signal/image processing, computer vision, machine learning, data analysis, meteorology, statistical physics, quantum mechanics, and network theory [8–13].

Our interest in the present work is extending the Benamou and Brenier computational fluid dynamical (CFD) approach to OMT, where the domains have transitioning compact support and may also be subject to potential spatiotemporal constraints (obstacles to avoid or mandatory regions to visit). We recall that in the seminal work [14], Benamou and Brenier compute the Wasserstein 2-metric W_2 via the minimization of kinetic energy subject to a continuity constraint. They moreover compute the optimal path (i.e., geodesic) in the space of probability densities [15]. In image processing, this gives a natural interpolation path between two images, where the intensity is treated as a "generalized mass". This is important for problems in deformable registration and image warping; see [16,17] and the references therein.

Recently, this same fluid dyamical framework for OMT was related to the Schrödinger bridge problem in the continuum limit of Brownian particles exhibiting different empirical distributions at two different times. In [18], Chen et al. precisely elaborate this relationship, which they leverage in order to develop a stochastic control formulation of the fluid dynamic OMT problem which allows for an "a priori" evolution as in the Schrödinger bridge.

Novelties and Contributions Presented in This Work

The fluid dynamical formulation of Benamou and Brenier, however, does not deliberately take into account that the regions of interest in image-based applications are necessarily compact, and may even have non-convex regions of support when considering segmented subregions of interest in medical imaging applications (for example, the deformation of the cardiac ventricle as illustrated in an experiment in Section 5.1). In such cases,

boundary conditions must crucially be imposed, not only to maintain mass conservation but also to prevent *mass exchanges* (even if balanced) between identified subregions and their backgrounds. However, such boundary conditions impose constraints on the optimization problem, which may not be compatible with the classical unconstrained solution.

It is this precisely this issue that motivated the present work. To treat the problem of OMT on densities with compact support, we propose a synergy of methods from OMT as well as level set evolutions [19,20]. Level set methods are a powerful way of implementing interface (boundary) evolution problems and thus have become a standard for a number of approaches in computer vision for segmentation, so-called *active contour* and *active surface* methods; see [21] and the references therein. We may not only impose mass-preserving boundary conditions along such modeled interfaces but may also consider how the variation of such interfaces changes the constrained minimizing solution, thereby yielding a joint optimization of the mass transport along the interior of the support together with a geometric optimization of the support boundary itself. This synergy of level set shape optimization and OMT, we believe to be novel, and to potentially have a number of applications, particularly to digital pathology.

Furthermore, with direct modeling and control of the support boundaries not only for the initial and final mass configurations but also along its intervening transport, we may readily handle intermediate spatiotemporal constraints with no additional modeling or computational machinery. While not the original motivation that led to the present work, it nonetheless yields a powerful and highly useful extension of the optimal mass transport problem that is naturally handled with the joint framework of fluid dynamics and geometric interface evolution. We may, for example, employ it to solve a variation of the original problem posed by Monge, where a pile of dirt must be moved from one location to another but with a river passing in between with a set of bridges which constrain the transport paths in between. We may even, as illustrated in an experiment in Section 5.2.2, readily apply it to solve yet another extension of the OMT problem where the initial density is known but where only the final support is known, not the final density itself.

Finally, while the focus of this work is to develop and computationally demonstrate a novel joint variational formulation for OMT combining the fluid dynamical framework of Benamou and Brennier with active geometric methods for the evolution of interfaces, we present two mathematical conjectures in Section 3.4.6, supported by our developed variational formulas, regarding the existence of solutions to the OMT problem in the presence of spatiotemporal constraints.

2. Spatiotemporal Hypersurface

We use **bold notation** exclusively to denote space-time points, vectors, and sets and non-bold notation for similar entities in time or space only, as well as for scalar variables. Accordingly, $\mathbf{X}$ will represent an arbitrary spatiotemporal point in $\mathbb{R}^{n+1}$

$$\mathbf{X} = (t, x) = (\underbrace{X_0}_{t}, \underbrace{X_1, X_2, \ldots, X_n}_{x})$$

which pairs a temporal coordinate $t \in \mathbb{R}$ with a spatial coordinate $x \in \mathbb{R}^n$. We will denote the spatiotemporal basis vectors by $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ so that

$$\mathbf{X} = \underbrace{X_0\,\mathbf{e}_0}_{t} + \underbrace{X_1\,\mathbf{e}_1 + X_2\,\mathbf{e}_2 + \cdots + X_n\,\mathbf{e}_n}_{x}$$

2.1. Assumptions

2.1.1. Compact Support

We consider a spatiotemporal density ρ in the form of a positive scalar function $\mathbf{X} \to \rho(\mathbf{X})$ in $\mathbb{R}^{n+1}$ whose restriction to the $t=0$ hyperplane matches a given initial spatial density ρ_0 and whose restriction to the $t=1$ hyperplane matches a given spatial target density ρ_1. We assume that the initial spatial density ρ_0 has compact support $\Omega_0 \subset \mathbb{R}^n$

and that the target spatial density ρ_1 has compact support $\Omega_1 \subset \mathbb{R}^n$. We further assume that the full spatiotemporal density ρ also has compact support $\Omega \subset \mathbb{R}^{n+1}$ in space-time, sandwiched between the hyperplanes $t = 0$ and $t = 1$, which may be constructed by a continuous family of intermediate compact spatial supports $\Omega_{[t]} \subset \mathbb{R}^n$ with $\Omega_{[0]} = \Omega_0$ and $\Omega_{[1]} = \Omega_1$ as follows:

$$\Omega = \left\{ \mathbf{X} = (t, x) \mid 0 \leq t \leq 1, \, x \in \Omega_{[t]} \right\} \tag{1}$$

2.1.2. Balanced Density

We assume that the initial and target spatial densities ρ_0 and ρ_1 both have unit mass. We impose a similar constraint on the spatiotemporal density ρ, summarizing these assumptions as follows.

$$\int_{\Omega_0} \rho_0(x) \, dx = \int_{\Omega_1} \rho_1(x) \, dx = \int_{\Omega} \rho(\mathbf{X}) \, d\mathbf{X} = 1 \tag{2}$$

(The reason we do not start with the stronger constraint $\int_{\Omega_{[t]}} \rho(x, t) \, dx = 1$ for all $0 \leq t \leq 1$ is that both this as well as the weaker total spatiotemporal mass constraint are global mass preservation constraints that will automatically be satisfied later when we impose the much stronger local mass preservation constraint. The main point in even presenting the total spatiotemporal constraint here is to reinforce the embedded space-time interpretation of the problem, thereby allowing us to interpret ρ as a unit spatiotemporal mass density directly in $\mathbb{R}^{n+1}$.)

2.1.3. Smoothness

We assume that the initial and target spatial densities ρ_0 and ρ_1 are differentiable *within their support* (we do not require ρ_0 and ρ_1 to be zero along the boundaries of their supports; as such, they may discontinuously drop to zero across the spatial boundaries Γ_0 and Γ_1 respectively). For Ω_0 and Ω_1, the spatiotemporal density ρ is differentiable *within its support* (we do not require ρ to be zero along the boundary of its support, a necessary freedom along the flat temporal faces Γ_0 and Γ_1; otherwise, we could not impose $\rho = \rho_o$ and $\rho = \rho_1$ along these components). For Ω, we also assume that the portion of the spatiotemporal boundary $\partial\Omega$ that lies strictly within $0 < t < 1$, which we denote by Γ, is differentiable. The remaining portions of $\partial\Omega$ are provided by the embeddings of Ω_0 and Ω_1 within the hyperplanes $t = 0$ and $t = 1$ to form two flat *temporal faces* of Ω, which we denote by Γ_0 and Γ_1.

2.1.4. Piece-Wise Smooth Boundary

As such $\partial\Omega = \Gamma \cup \Gamma_0 \cup \Gamma_1$ will have the form of a compact hypersurface in $\mathbb{R}^{n+1}$ with a well-defined outward normal everywhere except along the borders of the two temporal faces where Γ_0 and Γ_1 connect to the intervening surface Γ. We may also describe the intervening spatiotemporal boundary component Γ as the *swept-out surface* generated by embedding the boundaries $\Gamma_{[t]} = \partial\Omega_{[t]}$ of the deforming spatial supports $\Omega_{[t]}$ into the corresponding temporal hyperplanes. The spatiotemporal boundary notation and decomposition is summarized as follows:

$$\partial\Omega = \underbrace{\left\{ \mathbf{X} = (0, x) \mid x \in \Omega_0 \right\}}_{\substack{\Gamma_0 \\ \text{(temporal face)}}} \cup \underbrace{\left\{ \mathbf{X} = (1, x) \mid x \in \Omega_1 \right\}}_{\substack{\Gamma_1 \\ \text{(temporal face)}}} \cup \underbrace{\left\{ \mathbf{X} = (t, x) \mid 0 \leq t \leq 1, \, x \in \overbrace{\partial\Omega_{[t]}}^{\Gamma_{[t]}} \right\}}_{\substack{\Gamma \\ \text{(swept-out surface)}}} \tag{3}$$

2.1.5. Smooth Extension

Finally, in order to treat the density ρ and its support Ω as independent entities to be optimized later on, we will also consider ρ within the support Ω to be the restriction of a positive density that smoothly extends ρ beyond its modeled support. As such, perturbations to the support boundary Γ may be considered independently of perturbations to the density values ρ. Similar treatment will be applied to upcoming vector fields to be defined over the same compact support Ω.

2.2. Local Geometry

In this section, we explore the relationship between the local geometry of the spatial support boundary $\Gamma_{[t]}$ and the swept-out spatiotemporal boundary Γ.

2.2.1. Parameterization

Let $s = (s_1, \ldots, s_{n-1})$ represent isothermal coordinates with unit speed at a point $x \in \Gamma_{[t]}$ along the boundary of the spatial support $\Omega_{[t]} \subset \mathbb{R}^n$ at time t strictly between 0 and 1. The Riemannian metric tensor of $\Gamma_{[t]}$ in these coordinates is therefore the $(n-1) \times (n-1)$ identity matrix at the point x with $n-1$ orthonormal tangent vectors $\frac{\partial \Gamma_{[t]}}{\partial s_k} \in \mathbb{R}^n$ for $k = 1, \ldots, n-1$. From (3), we know that the corresponding spatiotemporal point $\mathbf{X} = (t, x)$ belongs to the Γ portion of the spatiotemporal support boundary $\partial \Omega$, which may be locally parameterized as follows:

$$\mathbf{\Gamma}(t,s) = \left(t, \Gamma_{[t]}(s) \right) = (\underbrace{\Gamma_0}_{t}, \underbrace{\Gamma_1, \Gamma_2, \ldots, \Gamma_n}_{\Gamma_{[t]}(s)}) \tag{4}$$

2.2.2. Unit Normal

In these coordinates, we compute the following n tangent vectors to $\mathbf{\Gamma}$ in $\mathbb{R}^{n+1}$:

$$\frac{\partial \mathbf{\Gamma}}{\partial t} = \left(1, \frac{\partial \Gamma_{[t]}}{\partial t} \right) \quad \text{and} \quad \frac{\partial \mathbf{\Gamma}}{\partial s_k} = \left(0, \frac{\partial \Gamma_{[t]}}{\partial s_k} \right), \quad k = 1, \ldots, n-1 \tag{5}$$

Since the unit outward normal $N \in \mathbb{R}^n$ to the spatial boundary $\Gamma_{[t]}$ is orthogonal to the $n-1$ spatial tangent vectors $\frac{\partial \Gamma_{[t]}}{\partial s_k} \in \mathbb{R}^n$, it follows that $(\alpha, N) \in \mathbb{R}^{n+1}$ will be orthogonal to the $n-1$ spatiotemporal tangent vectors $\frac{\partial \mathbf{\Gamma}}{\partial s_k}$ expressed in (5) for any choice of scalar α. Orthogonality to the additional spatiotemporal tangent vector $\frac{\partial \mathbf{\Gamma}}{\partial t}$ expressed in (5) also requires $\alpha = -\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N$, yielding the following construction of the spatiotemporal unit outward normal $\mathbf{N} \in \mathbb{R}^{n+1}$:

$$\mathbf{N} = \begin{cases} \dfrac{\left(-\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N, N \right)}{\sqrt{1 + \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N \right)^2}}, & \mathbf{X} \in \mathbf{\Gamma} \\ +\mathbf{e}_0 = +(1, 0, \ldots, 0), & \mathbf{X} \in \mathbf{\Gamma}_1 \\ -\mathbf{e}_0 = -(1, 0, \ldots, 0), & \mathbf{X} \in \mathbf{\Gamma}_0 \end{cases} \tag{6}$$

2.2.3. Metric Tensor

Using the tangent vectors (5), we may express the Riemannian metric tensor at the point $\mathbf{X}$ in the form of the following $n \times n$ matrix:

$$\underbrace{\mathbf{I}(s_1, \ldots, s_{n-1}, t)}_{\text{I-fundamental form}} = \begin{bmatrix} \mathcal{I} & \left(\frac{\partial \Gamma_{[t]}}{\partial s} \right)^T \frac{\partial \Gamma_{[t]}}{\partial t} \\ \frac{\partial \Gamma_{[t]}}{\partial t}^T \left(\frac{\partial \Gamma_{[t]}}{\partial s} \right) & 1 + \frac{\partial \Gamma_{[t]}}{\partial t} \cdot \frac{\partial \Gamma_{[t]}}{\partial t} \end{bmatrix} \tag{7}$$

where $\frac{\partial \Gamma_{[t]}}{\partial s}$ denotes the $n \times (n-1)$ matrix, whose columns consist of the orthonormal tangent vectors to the spatial boundary $\Gamma_{[t]}$ at the point x. Using the determinant formula, $\det \begin{bmatrix} A & u \\ v^T & \alpha \end{bmatrix} = \alpha \det A - v^T \mathrm{adj} A\, u$ (for any matrix A, vector u and v, and scalar α), we may compute:

$$\det \mathbf{I} = \left(1 + \frac{\partial \Gamma_{[t]}}{\partial t} \cdot \frac{\partial \Gamma_{[t]}}{\partial t}\right) - \underbrace{\sum_{k=1}^{n-1} \left(\frac{\partial \Gamma_{[t]}}{\partial s_k} \cdot \frac{\partial \Gamma_{[t]}}{\partial t}\right)^2}_{\left\| \frac{\partial \Gamma_{[t]}}{\partial t} \right\|^2 - \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N\right)^2} = 1 + \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N\right)^2$$

2.2.4. Area Element

The relationship between the area element $dS_{[t]}$ of the spatial boundary surface $\Gamma_{[t]}$, the temporal element dt, and the area element $d\mathbf{S}$ of the swept-out spatiotemporal hypersurface $\mathbf{\Gamma}$ can be expressed via the square root of the determinant of the first fundamental form shown above:

$$d\mathbf{S} = \sqrt{1 + \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N\right)^2} \, dS_{[t]} \, dt \tag{8}$$

2.2.5. Normal Variations

Finally, using the parameterization (4) for a variation $\delta \mathbf{\Gamma}$ allows us to relate a variation of the swept-out spatiotemporal hypersurface to time parameterized variations $\delta \Gamma_{[t]}$ of the spatial support boundaries as follows: $\delta \mathbf{\Gamma} = \left(0, \delta \Gamma_{[t]}\right)$. Combining this with (6) yields the following relationship between variations of $\Gamma_{[t]}$ and variations of $\mathbf{\Gamma}$ along their respective normal directions:

$$\delta \mathbf{\Gamma} \cdot \mathbf{N} = \frac{\delta \Gamma_{[t]} \cdot N}{\sqrt{1 + \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N\right)^2}}$$

If we now further combine this with (8), we see that the normal variation of the spatiotemporal hypersurface measured against its spatiotemporal area element matches the normal variation of the corresponding spatial boundary measured by its respective area element and time element:

$$(\delta \mathbf{\Gamma} \cdot \mathbf{N}) \, d\mathbf{S} = \left(\delta \Gamma_{[t]} \cdot N\right) dS_{[t]} \, dt \tag{9}$$

3. Spatiotemporal Formulation of Optimal Mass Transport

The fluid dynamical framework of Benamou and Brenier considers two time-evolving entities, a scalar mass density $\rho(t, x)$ and a velocity field $v(t, x) \in \mathbb{R}^n$, which are coupled by the local mass-preserving continuity constraint $\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho v) = 0$.

Notice that in the standard manner regarding ρv as an ordinary 3-vector, the continuity constraint means that density and spatial momentum form a 4-vector with respect to the standard Minkowski metric in the standard physics setting [22]. We will exploit this observation and develop an equivalent formulation using a single space-time solenoidal vector field $\mathbf{W}$ with simple normal boundary conditions along the spatiotemporal hypersurface $\partial \mathbf{\Omega}$ that enable convenient numerical solutions of PDEs directly within an $n + 1$ dimensional space-time grid, with no need to treat the temporal and spatial dimensions separately or differently.

3.1. Spatiotemporal Advection Field **U**

We begin by noting that in the combined spatiotemporal variable $\mathbf{X} = (t, x)$, the continuity constraint $\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho v) = 0$ can be written as

$$\nabla \rho \cdot \mathbf{U} + \rho \nabla \cdot \mathbf{U} = 0 \tag{10}$$

where ∇ and $\nabla \cdot$ represent the full spatiotemporal gradient and divergence operators in $\mathbb{R}^{n+1}$ and where $\mathbf{U} \in \mathbb{R}^{n+1}$ denotes the following vector field:

$$\mathbf{U}(\mathbf{X}) \doteq (1, v) = (\underbrace{U_0}_{1}, \underbrace{U_1, U_2, \ldots, U_n}_{v})$$

Note that $\mathbf{U}$ is tangent to the characteristics of this linear first-order PDE (10) in ρ and therefore defines the trajectories along which mass is transported across space-time. Since the spatiotemporal hypersurface $\partial \Omega$, more specifically, its swept-out portion $\boldsymbol{\Gamma}$, defines the boundary of the evolving support for ρ, we can conclude that these advection trajectories must flow along the hypersurface $\boldsymbol{\Gamma}$ itself, never across it (neither inward nor outward). In other words, mass cannot be transported outside of its support neither forward in time (which excludes outward-flowing characteristics) nor backward in time (which excludes inward-flowing characteristics). This leads to the boundary condition $\mathbf{U} \cdot \mathbf{N} = 0$ along $\boldsymbol{\Gamma}$, which may be written in terms of v, N, and $\frac{\partial \Gamma_{[t]}}{\partial t}$ as follows:

$$\underbrace{v \cdot N = \frac{\partial \Gamma_{[t]}}{\partial t} \cdot N}_{(\mathbf{U} \cdot \mathbf{N} = 0)} \tag{11}$$

If we plug this constraint between the support evolution $\frac{\partial \Gamma_{[t]}}{\partial t}$ and the velocity field v into (6), we obtain the following alternative expression for the outward unit normal $\mathbf{N} \in \mathbb{R}^{n+1}$ of the swept-out hypersurface $\boldsymbol{\Gamma}$ in terms of the outward normal $N \in \mathbb{R}^n$ of the support boundary $\Gamma_{[t]}$ in space at time t:

$$\mathbf{N}(\mathbf{X}) = \frac{(-v \cdot N, N)}{\sqrt{1 + (v \cdot N)^2}}, \quad \mathbf{X} \in \boldsymbol{\Gamma} \tag{12}$$

3.2. Solenoidal Vector Field **W**

While the use of the advection field $\mathbf{U}$ merges the spatial and temporal derivatives into a single derivative $\mathbb{R}^{n+1}$ operator in (11), it still keeps the density variable ρ separate. We now merge these two entities by defining another spatiotemporal vector field whose first (temporal) component represents the mass density $\rho(\mathbf{X})$, and whose remaining components represent the momentum vector $\mathrm{p}(\mathbf{X}) = \rho v$ in $\mathbb{R}^n$:

$$\mathbf{W}(\mathbf{X}) = \rho \mathbf{U} = (\rho, \mathrm{p}) = (\underbrace{W_0}_{\rho}, \underbrace{W_1, W_2, \ldots, W_n}_{\text{momentum } \mathrm{p} = \rho v}) \tag{13}$$

However, rather than considering (13) to be the definition of $\mathbf{W}$ in terms of the density ρ and momentum p, we instead consider it in reverse to be the definition of ρ and p in terms of the space-time vector field $\mathbf{W}$ subject to the continuity constraint (10), which now simplifies to a coordinate-free solenoidal condition on $\mathbf{W}$:

$$\nabla \cdot \mathbf{W} = 0 \tag{14}$$

Multiplying the boundary condition $\mathbf{U} \cdot \mathbf{N} = 0$ along $\boldsymbol{\Gamma}$ presented in (11) by ρ yields a similar vanishing flux condition for $\mathbf{W}$ across the swept-out portion of the spatiotemporal hypersurface $\partial \Omega$. For the remainder of $\partial \Omega$, we combine $\mathbf{N} = \pm \mathbf{e}_0$ from (6) with (13) to

obtain flux conditions for **W** along the temporal faces $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ as well, in terms of the known starting and target densities ρ_0 and ρ_1.

The combination of these constraints is easily summarized now in terms of **W** and its spatiotemporal domain $\mathbf{\Omega}$. Namely, we seek a *solenoidal vector field* **W** within $\mathbf{\Omega}$ with the following prescribed flux conditions along the full spatiotemporal domain boundary $\partial\mathbf{\Omega}$:

$$\mathbf{W} \cdot \mathbf{N} = \begin{cases} 0 \, , & \mathbf{X} \in \mathbf{\Gamma} \\ -\rho_0, & \mathbf{X} \in \mathbf{\Gamma}_0 \\ +\rho_1, & \mathbf{X} \in \mathbf{\Gamma}_1 \end{cases} \tag{15}$$

3.3. Extended Velocity **V**

3.3.1. Local Kinetic Energy

Before setting up the variational problem, we seek an expression for the local measure of kinetic energy $\mathrm{T}(t, x)$

$$\mathrm{T}(\mathbf{X}) = \frac{1}{2}\rho \, v \cdot v$$

in terms of the solenoidal field **W**. If we express this purely in terms of **W**, we obtain the following expression which, unfortunately, is not coordinate-free:

$$\mathrm{T}(\mathbf{X}) = \frac{1}{2}\left(\frac{\overbrace{\mathbf{W} \cdot \mathbf{W}}^{\rho^2(1+v\cdot v)}}{\underbrace{\mathbf{W} \cdot \mathbf{e}_0}_{\rho}} - \underbrace{\mathbf{W} \cdot \mathbf{e}_0}_{\rho} \right)$$

3.3.2. Extended Velocity

We may resolve this by introducing the following *extended velocity* field $\mathbf{V} \in \mathbb{R}^{n+1}$ which extends the transport velocity v from $\mathbb{R}^n$ into $\mathbb{R}^{n+1}$ by adding a temporal component equal to $-\frac{1}{2}\|v\|^2$ as follows:

$$\mathbf{V} \doteq \left(-\frac{1}{2}\|v\|^2 \, , \, v \right) = (\underbrace{V_0}_{\frac{-\|v\|^2}{2}} , \underbrace{V_1, V_2, \dots, V_n}_{v}) \tag{16}$$

Notice that, just like the advection field **U**, the extended velocity **V** depends only upon the spatial velocity v itself, and therefore contains no additional information. We may now express the local kinetic energy compactly and as being coordinate-free in terms of the solenoidal field **W** and the extended velocity field **V** by their inner product:

$$\mathrm{T} = \mathbf{W} \cdot \mathbf{V}$$

3.3.3. Generalized Momentum

Notice that if we multiply the extended velocity **V** by ρ we obtain the *generalized momentum* **P**, which we construct in terms of the kinetic energy T and the momentum p as follows:

$$\mathbf{P} = \left(-\frac{1}{2}\rho\|v\|^2 \, , \, \rho v \right) = (-\mathrm{T} \, , \, \mathrm{p}) = (\underbrace{P_0}_{-\mathrm{T}} , \underbrace{P_1, P_2, \dots, P_n}_{\mathrm{p}})$$

Both the solenoidal field **W** and the generalized momentum field **P** go to zero as the density ρ goes to zero, even in cases where the velocity v becomes undefined. While we will not be utilizing **P** in our upcoming variational formulation, it is nevertheless useful to interpret the extended velocity **V** as $\mathbf{P}/\rho$ for $\rho > 0$ just as we may interpret the advection field **U** as $\mathbf{W}/\rho$ for $\rho > 0$ without explicit reference to the velocity v. We now summarize and consolidate our notation for the four spatial–temporal vector fields in $\mathbb{R}^{n+1}$ as follows:

solenoidal field	generalized momentum
$\mathbf{W} = (\rho, \mathbf{p})$	$\mathbf{P} = (-\mathrm{T}, \mathbf{p})$

advection field	extended velocity
$\mathbf{U} = \dfrac{\mathbf{W}}{\rho} = (1, v)$	$\mathbf{V} = \dfrac{\mathbf{P}}{\rho} = \left(-\dfrac{1}{2}\|v\|^2 , v \right)$

3.4. Variational Formulation

3.4.1. Action Integral

We begin by constructing the action integral to be minimized over $\mathbf{\Omega} \subset \mathbb{R}^{n+1}$ in terms of the solenoidal field $\mathbf{W}$ as follows:

$$E(\mathbf{W}, \mathbf{\Gamma}) = \int_0^1 \int_{\Omega_{[t]}} \overbrace{\frac{1}{2}\rho\|v\|^2}^{T} dx\, dt = \int_{\mathbf{\Omega}} \mathbf{W} \cdot \mathbf{V} \, d\mathbf{X} \tag{17}$$

Note that the two unknowns are only $\mathbf{W}$ and its support $\mathbf{\Omega}$ (or equivalently, the swept-out boundary $\mathbf{\Gamma}$), even though we have expressed the action compactly also in terms of $\mathbf{V}$. We may compute $\mathbf{V}$ directly from $\mathbf{W}$

$$\mathbf{V} \doteq \left(-\frac{1}{2}\|v\|^2 , v \right) = \mathbf{U} - \frac{1}{2}(\mathbf{U} \cdot \mathbf{U} + 1)\mathbf{e}_0 = \frac{\mathbf{W}}{\mathbf{W} \cdot \mathbf{e}_0} - \frac{1}{2}\left(\frac{\mathbf{W} \cdot \mathbf{W}}{(\mathbf{W} \cdot \mathbf{e}_0)^2} + 1 \right)\mathbf{e}_0 \tag{18}$$

with the following compatible flux conditions obtained by plugging in (15):

$$\mathbf{V} \cdot \mathbf{N} = \begin{cases} -\left(1 + \frac{1}{2}\|v\|^2\right)(\mathbf{N} \cdot \mathbf{e}_0), & \mathbf{X} \in \mathbf{\Gamma} \\ +\frac{1}{2}\|v\|^2, & \mathbf{X} \in \mathbf{\Gamma}_0 \\ -\frac{1}{2}\|v\|^2, & \mathbf{X} \in \mathbf{\Gamma}_1 \end{cases}$$

We incorporate the solenoidal (mass preservation) constraint (14) through a Lagrange multiplier λ over $\mathbf{\Omega}$ and the flux constraints through additional Lagrange multipliers α_0, α_1, and α along the boundaries $\mathbf{\Gamma}_0$, $\mathbf{\Gamma}_1$, and $\mathbf{\Gamma}$, respectively:

$$E = \int_{\mathbf{\Omega}} \mathbf{W} \cdot \mathbf{V} + \underbrace{\lambda \nabla \cdot \mathbf{W}}_{\substack{\text{solenoidal} \\ \text{constraint}}} d\mathbf{X} + \underbrace{\int_{\mathbf{\Gamma}} \alpha \mathbf{W} \cdot \mathbf{N} \, d\mathbf{S} + \int_{\mathbf{\Gamma}_0} \alpha_0 \left(\overbrace{\mathbf{W} \cdot \mathbf{N}}^{-\mathbf{W} \cdot \mathbf{e_0}} + \rho_0 \right) dx + \int_{\mathbf{\Gamma}_1} \alpha_1 \left(\overbrace{\mathbf{W} \cdot \mathbf{N}}^{\mathbf{W} \cdot \mathbf{e_0}} - \rho_1 \right) dx}_{\text{flux constraints}} \tag{19}$$

3.4.2. First Variation

In Appendix A, we compute the variation of (19) which reduces to

$$\delta E = \int_{\mathbf{\Omega}} \delta\mathbf{W} \cdot (\mathbf{V} - \nabla\lambda) \, d\mathbf{X} + \int_{\mathbf{\Gamma}} \mathbf{W} \cdot (\mathbf{V} - \nabla\lambda)\delta\mathbf{\Gamma} \cdot \mathbf{N} \, d\mathbf{S} \tag{20}$$

as long as the solenoidal and flux constraints for $\mathbf{W}$ are satisfied and as long as we further impose $\delta\mathbf{W} \cdot \mathbf{e}_0 = 0$ at $t = 0$ and $t = 1$ to preserve the initial and final densities ρ_0 and ρ_1 as well as $\delta\mathbf{\Gamma} = 0$ at $t = 0$ and $t = 1$ to preserve the initial and final spatial supports Ω_0 and Ω_1.

3.4.3. Optimality Condition for $\mathbf{W}$

By inspection of (20), optimality for $\mathbf{W}$ can only be achieved if we can solve

$$\nabla\lambda = \mathbf{V} \tag{21}$$

for λ within the interior of $\boldsymbol{\Omega}$. If we separately equate the spatial and temporal components

$$\overbrace{\left(\underbrace{\frac{\partial\lambda}{\partial X_0}}_{\frac{\partial\lambda}{\partial t}}, \underbrace{\frac{\partial\lambda}{\partial X_1}, \frac{\partial\lambda}{\partial X_2}, \dots, \frac{\partial\lambda}{\partial X_n}}_{\frac{\partial\lambda}{\partial x}}\right)}^{\nabla\lambda} = \overbrace{\left(\underbrace{V_0}_{\frac{-\|v\|^2}{2}}, \underbrace{V_1, V_2, \dots, V_n}_{v}\right)}^{\mathbf{V}}$$

then we see that (21) amounts to a more compact, coordinate-free expression of the well-known Hamilton Jacobi equation:

$$\frac{\partial\lambda}{\partial t} + \frac{1}{2}\left\|\frac{\partial\lambda}{\partial x}\right\|^2 = 0 \tag{22}$$

However, (21) will not admit a solution unless $\mathbf{V}$ is a conservative (irrotational) vector field. As such, the gradient must be related to the non-conservative (solenoidal) portion of the extended velocity field $\mathbf{V}$. Accordingly, we consider the Helmholtz decomposition, expressing $\mathbf{V}$ as the sum of two vector fields:

$$\mathbf{V} = \underbrace{\mathbf{V}^{\|}}_{\substack{\text{irrotational}\\\text{(curl-free)}}} + \underbrace{\mathbf{V}^{\perp}}_{\substack{\text{solenoidal}\\\text{(divergence-free)}}} \tag{23}$$

where $\mathbf{V}^{\perp}$ denotes the divergence-free component ($\nabla \cdot \mathbf{V}^{\perp} = 0$) and where $\mathbf{V}^{\|}$ denotes the curl-free component ($\nabla \times \mathbf{V}^{\|} = 0$) which can be written as the gradient of a scalar potential function.

We can now always choose the Lagrange multiplier λ to satisfy

$$\mathbf{V}^{\|} = \nabla\lambda \quad \text{and} \quad \mathbf{V}^{\perp} = \mathbf{V} - \nabla\lambda \tag{24}$$

which annihilates the curl-free component $\mathbf{V}^{\|}$ in (20) and yields a cleaner expression of the optimality condition purely in terms of $\mathbf{V}$ as follows:

$$\mathbf{V}^{\perp} = 0 \tag{25}$$

3.4.4. Partial Gradient with Respect to $\mathbf{W}$ (Fixed Support)

In fact, the solenoidal component $\mathbf{V}^{\perp}$ of the extended velocity represents the gradient of the action integral E with respect to the solenoidal field $\mathbf{W}$ alone, keeping the spatiotemporal support $\boldsymbol{\Omega}$ fixed:

$$\nabla_{\mathbf{W}}E = \mathbf{V}^{\perp} \tag{26}$$

This can be seen immediately by plugging $\nabla\lambda = \mathbf{V}^{\|}$ and $\delta\boldsymbol{\Gamma} = 0$ into the first variation (20) to obtain

$$\delta E\Big|_{(\delta\boldsymbol{\Gamma}=0)} = \int_{\boldsymbol{\Omega}} \delta\mathbf{W} \cdot \mathbf{V}^{\perp} \, d\mathbf{X} \tag{27}$$

Note, however, that the Helmholtz decomposition is not unique over compact domains but depends upon the boundary conditions chosen for $\mathbf{V}^{\|}$ or $\mathbf{V}^{\perp}$. Since we are treating the support of $\mathbf{W}$ as fixed, then the gradient field $\nabla_{\mathbf{W}}E$ must not contribute flux anywhere along the spatiotemporal boundary $\partial\boldsymbol{\Omega}$ (in order to respect local mass conservation). This means, by (26), that the flux of $\mathbf{V}^{\perp}$ must vanish, which makes the decomposition (23)

unique. Plugging (24) into the solenoidal and vanishing flux conditions for $\mathbf{V}^{\perp}$ yields a Poisson equation with Neumann boundary conditions for λ:

$$\Delta\lambda = \nabla \cdot \mathbf{V}, \quad \mathbf{X} \in \mathbf{\Omega} \tag{28}$$
$$\nabla\lambda \cdot \mathbf{N} = \mathbf{V} \cdot \mathbf{N}, \quad \mathbf{X} \in \partial\mathbf{\Omega}$$

Since $\mathbf{N} = \pm\mathbf{e}_0$ along the temporal faces of $\mathbf{\Omega}$, we may separate the Neumann boundary expressions along the three different portions of $\partial\mathbf{\Omega}$ as follows:

$$\nabla\lambda \cdot \mathbf{N} = \begin{cases} +\frac{1}{2}\|v\|^2, & \mathbf{X} \in \mathbf{\Gamma}_0 \\ -\frac{1}{2}\|v\|^2, & \mathbf{X} \in \mathbf{\Gamma}_1 \\ \mathbf{V} \cdot \mathbf{N}, & \mathbf{X} \in \mathbf{\Gamma} \end{cases} \tag{29}$$

3.4.5. Shape Optimality Condition for $\mathbf{\Gamma}$

By inspection of (20), shape optimality for $\mathbf{\Gamma}$ requires

$$\mathbf{W} \cdot (\mathbf{V} - \nabla\lambda) = 0, \quad \mathbf{X} \in \mathbf{\Gamma} \tag{30}$$

along the support boundary $\mathbf{\Gamma}$. Given our choice (24) for the Lagrange multiplier λ, we may rewrite the condition purely in terms of the extended velocity $\mathbf{V}$ as

$$\mathbf{W} \cdot \mathbf{V}^{\perp} = 0, \quad \mathbf{X} \in \mathbf{\Gamma} \tag{31}$$

Notice, however, that this shape optimality condition (31) is just a weaker version of the optimality condition $\mathbf{V}^{\perp} = 0$ for the solenoidal field $\mathbf{W}$, despite our independent treatment of $\mathbf{W}$ and its support boundary $\mathbf{\Gamma}$ (recall our smooth-extension assumption in Section 2.1 in which vector fields defined over the compact support $\mathbf{\Omega}$ are treated as restrictions of vector fields that smoothly extend beyond the support).

This leaves us with the conundrum that if we are able to optimize $\mathbf{W}$ for any candidate spatiotemporal support $\mathbf{\Omega}$, then not only does the partial gradient vanish with respect to $\mathbf{W}$ (and its smooth extension) but then so does the partial gradient with respect to $\mathbf{\Gamma}$ as well! Not only does this counter-intuitively imply that all smoothly changing support evolutions between $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_1$ represent critical points of the action integral (17) but it completely sabotages our ultimate shape optimization strategy, in which we hoped to propose an initial compact support evolution between $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_1$, optimize $\mathbf{W}$ over the proposed spatiotemporal support $\mathbf{\Omega}$, and then evolve the support boundaries $\mathbf{\Gamma}$ according to the shape gradient.

3.4.6. Conjectures on the Existence of Optimal Transport with Intermediate Support Constraints

A potential explanation for this conundrum is that the constrained minimization of the action integral (17) with respect to $\mathbf{W}$ over an imposed spatiotemporal support $\mathbf{\Omega}$ may actually be an ill-posed problem unless the imposed support happens to match the optimal one already. If so, then a minimizer $\mathbf{W}$ would not exist in the absence of further constraints, despite the existence of a well-defined gradient (26). This would imply that the gradient flow does not converge, not even asymptotically.

Assuming that this conjecture is correct, it is natural to wonder what would happen in the gradient descent for $\mathbf{W}$ if we impose a non-optimal support $\mathbf{\Omega}$ over space-time. Mass would likely accumulate forever toward an infinite density along portions of the constrained support boundary that lie strictly within the interior of the optimal support and would likely decay forever toward zero density within portions of the constrained support that lie outside the optimal one. Imposing lower and upper bounds on the density could potentially render the problem well posed in such scenarios.

Since the conjectures apply to any non-optimal choice of imposed spatiotemporal support, we can extrapolate them to more general conjectures in terms of the existence

of solutions to the optimal transport problem in the presence of intermediate support constraints as follows.

Conjecture 1. *A minimal action, mass-preserving flow which transports a compactly supported, uniformly continuous density ρ_0 to another compactly supported, uniformly continuous density ρ_1 does not exist (without breaking uniform continuity) when constraints are placed on the intermediate support, unless the constraints are also satisfied by the unconstrained solution.*

In the case of positive support constraints (requiring mass within regions that would otherwise fall outside the support of the unconstrained solution), then any proposed solution can likely be replaced by a lower cost solution that transports less mass into superfluous regions and therefore brings the density in such regions closer to zero. In the case of negative support constraints (prohibiting mass within regions that would otherwise fall inside the support of the unconstrained solution, perhaps due to an obstacle in the way), then any proposed solution can likely be replaced by a lower cost solution that accumulates more mass along the boundaries of the excluded regions.

Conjecture 2. *Imposing lower and upper bounds on the density may lead to the existence of solutions to the same constrained optimal transport problem, even when the intermediate support constraints are violated by the unconstrained solution.*

In the case of positive support constraints, a lower bound (if greater then 0) on the density would prevent replacing a proposed solution which already hits this lower density bound with a lower cost solution that brings the density closer to zero in regions where the lower bound is already achieved. In the case of negative support constraints, a finite upper bound would prevent replacing a proposed solution which already hits this upper density bound with a lower cost solution that further raises the density near or along support boundaries where the upper bound is already achieved.

4. Shape Optimization Strategy with Added Density Prior

Regardless of whether the first conjecture is correct, the fact that the shape gradient always vanishes if **W** is optimized (should an optimizer exist) for any arbitrary estimate of the support Ω, means that a gradient descent strategy to optimize the shape of the support boundary $\boldsymbol{\Gamma}$ cannot be employed using the variational formulation developed thus far. Motivated by the second conjecture, we will introduce a density prior (though not in the form of hard bounds as postulated in the conjecture but rather as soft bounds) that can be conveniently added to the variational formulation already developed. This will take the form of an additional potential to the Benamou–Brenier action introduced generally in Section 4.1 with an explicit, intuitive example later in Section 4.4.

(While for now we only present a potential term which depends upon the density values, we could introduce a spatiotemporal dependence as well an added potential term if we wished to soften the spatiotemporal support constraints and directly penalize mass in the proscribed regions as well as lack of mass in the prescribed regions. The density prior proposed here, instead, is intended to be paired with hard constraints on the spatiotemporal support by forcing the support boundary to keep mass out of the proscribed regions and within the prescribed regions but without infinitely piling up nor thinning out along such boundaries.)

The added potential term will allow us to approach the optimal mass transport problem between compact supports as a shape optimization problem, where an initial space-time support estimate $\boldsymbol{\Omega}$ is constructed via a smooth spatiotemporal hypersurface $\boldsymbol{\Gamma}$ which interpolates between the support boundaries $\partial\Omega_0$ and $\partial\Omega_1$ of the initial and final densities, where the imposed-support optimal transport problem is solved along this estimate by optimizing the solenoidal field **W** along this support estimate, and where the optimized **W** is used to determine the total shape gradient to optimally perturb the support

estimate using a gradient descent step for the hypersurface $\boldsymbol{\Gamma}$. The process is then repeated until this shape evolution process converges. The ingredients for this process are outlined as follows.

4.1. Density Prior

Consider a density prior on $\mathbf{W}$ in the form of a convex penalty function $h(\rho) \geq 0$ applied to the density values $\rho = \mathbf{W} \cdot \mathbf{e}_0$. We could, for example, set $h = 0$ (i.e., no penalty) for values of ρ within a desired range with a progressively increasing penalty ($h > 0$) for values of ρ that move further away from the desired range. Adding such a soft prior to the action integral, together with the Lagrange multipliers in (19), yields a new energy functional to be minimized as follows:

$$\tilde{E} = E + \underbrace{\int_{\boldsymbol{\Omega}} h(\overbrace{\mathbf{W} \cdot \mathbf{e}_0}^{\rho}) \, d\mathbf{X}}_{\text{density prior}} \tag{32}$$

The first variation of the new combined energy functional $\tilde{E}$ becomes

$$\begin{aligned}
\delta\tilde{E} &= \delta E + \int_{\boldsymbol{\Omega}} \dot{h}(\mathbf{W} \cdot \mathbf{e}_0) \, \mathbf{e}_0 \cdot \delta\mathbf{W} \, d\mathbf{X} + \int_{\boldsymbol{\Gamma}} h(\mathbf{W} \cdot \mathbf{e}_0) \, \delta\boldsymbol{\Gamma} \cdot \mathbf{N} \, d\mathbf{S} \\
&= \int_{\boldsymbol{\Omega}} \delta\mathbf{W} \cdot \Big(\underbrace{\mathbf{V} + \dot{h} \, \mathbf{e}_0}_{\tilde{\mathbf{v}}} - \nabla\lambda \Big) \, d\mathbf{X} + \int_{\boldsymbol{\Gamma}} (\mathbf{W} \cdot (\mathbf{V} - \nabla\lambda) + h) \delta\boldsymbol{\Gamma} \cdot \mathbf{N} \, d\mathbf{S}
\end{aligned} \tag{33}$$

assuming again, as in (20), that the solenoidal and flux constraints are satisfied for $\mathbf{W}$, and that both $\delta\mathbf{W} \cdot \mathbf{e}_0 = 0$ and $\delta\boldsymbol{\Gamma} = 0$ are imposed along the temporal boundaries $t = 0$ and $t = 1$ in order to preserve the initial and final densities and spatial supports.

4.2. Optimality Conditions and Partial Gradients

The optimality condition for $\mathbf{W}$ now becomes

$$\underbrace{\mathbf{V} + \dot{h} \, \mathbf{e}_0}_{\tilde{\mathbf{v}}} - \nabla\lambda = 0 \tag{34}$$

while the optimality condition for $\boldsymbol{\Gamma}$ now becomes

$$\mathbf{W} \cdot (\mathbf{V} - \nabla\lambda) + h = 0 \tag{35}$$

and is not automatically satisfied by the condition for $\mathbf{W}$ anymore. As such, the conundrum we faced before the introduction of the density penalty function h is now resolved.

Choosing the Lagrange multiplier such that

$$\tilde{\mathbf{V}}^{\|} = \nabla\lambda \quad \text{and} \quad \tilde{\mathbf{V}}^{\perp} = \tilde{\mathbf{V}} - \nabla\lambda$$

where $\tilde{\mathbf{V}}^{\|}$ and $\tilde{\mathbf{V}}^{\perp}$ denote the curl-free and divergence-free components in the Helmholtz decomposition of

$$\tilde{\mathbf{V}} = \mathbf{V} + \dot{h} \, \mathbf{e}_0 \tag{36}$$

(assuming vanishing flux conditions for $\tilde{\mathbf{V}}^{\perp}$) yields the following partial gradient with respect to $\mathbf{W}$

$$\nabla_{\mathbf{W}} \tilde{E} = \tilde{\mathbf{V}}^{\perp} = \tilde{\mathbf{V}} - \nabla\lambda \tag{37}$$

and the following partial gradient with respect to $\boldsymbol{\Gamma}$

$$\nabla_{\boldsymbol{\Gamma}} \tilde{E} = \Big(\mathbf{W} \cdot (\mathbf{V} - \nabla\lambda) + h \Big) \mathbf{N} = \Big(\mathbf{W} \cdot \tilde{\mathbf{V}}^{\perp} + h - \rho\dot{h} \Big) \mathbf{N} \tag{38}$$

Both of these partial gradient expressions require that we first solve the following Poisson equation for the Lagrange multiplier λ (in order to compute $\tilde{\mathbf{V}}^\perp$) with Neumann boundary conditions:

$$\Delta\lambda = \nabla \cdot \tilde{\mathbf{V}}, \quad \mathbf{X} \in \boldsymbol{\Omega} \tag{39}$$
$$\nabla\lambda \cdot \mathbf{N} = \tilde{V} \cdot \mathbf{N}, \quad \mathbf{X} \in \partial\boldsymbol{\Omega}$$

4.3. Optimization in **W** (Fixed Support)

4.3.1. Computing an Initial Solenoidal Field $\mathbf{W}_{\text{init}}$

If we combine the solenoidal constraint and boundary flux conditions for **W** with the additional constraint that the initial field $\mathbf{W}_{\text{init}}$ be conservative as well, then we may plug $\mathbf{W} = \nabla\Phi$ into (14) and (15), for some scalar spatiotemporal function Φ to obtain the Laplace equation with Neumann boundary conditions (non-homogeneous along the two temporal faces):

$$\Delta\Phi = 0, \qquad \mathbf{X} \in \boldsymbol{\Omega} \tag{40}$$

$$\nabla\Phi \cdot \mathbf{N} = \begin{cases} 0 \;, & \mathbf{X} \in \boldsymbol{\Gamma} \\ -\rho_0, & \mathbf{X} \in \boldsymbol{\Gamma}_0 \\ +\rho_1, & \mathbf{X} \in \boldsymbol{\Gamma}_1 \end{cases}$$

A solution will exist as long as $0 = \int_{d\boldsymbol{\Omega}} \nabla\Phi \cdot \mathbf{N}\, d\mathbf{S} = \int_{d\boldsymbol{\Omega}} \mathbf{W} \cdot \mathbf{N}\, d\mathbf{S}$, which in this case is equivalent to our balanced assumption $\int_{\Omega_0} \rho_0\, dx = \int_1 \rho_1\, dx$ for the initial and target densities. The solution will be unique up to an additive constant, which will then disappear after applying the gradient to obtain the following initial vector field:

$$\mathbf{W}_{\text{init}} = \nabla\Phi \tag{41}$$

4.3.2. Iterated Gradient Descent for the Solenoidal Field **W**

Referring to our k'th iterated estimate of the optimal **W** over our current support estimate as $\mathbf{W}_k$ (beginning with $\mathbf{W}_0 = \mathbf{W}_{\text{init}}$), a descent step can be taken by constructing $\tilde{\mathbf{V}}_k^\perp = \tilde{\mathbf{V}}_k - \nabla\lambda$ after solving the Poisson Equation (39) for λ over the existing space-time support $\boldsymbol{\Omega}$ and then applying a gradient update step

$$\mathbf{W}_{k+1} \to \mathbf{W}_k - \gamma_k \tilde{\mathbf{V}}_k^\perp \tag{42}$$

We may use Newton's method to determine the optimal step factor γ_k by solving

$$0 = \frac{dE_{k+1}}{d\gamma_k} = \int_{\boldsymbol{\Omega}} \tilde{\mathbf{V}}_{k+1}(\gamma_k) \cdot \tilde{\mathbf{V}}_k^\perp\, d\mathbf{X}$$

for each gradient step. We repeat this process until $\tilde{\mathbf{V}}_{k+1}^\perp \approx 0$

4.4. Total Shape Gradient

We obtain the total shape gradient $\nabla\tilde{E}$ with respect to the current support boundary $\boldsymbol{\Gamma}$ by plugging the optimized value of **W** over that support (now causing the first variation (33) to depend only upon the shape perturbation $\delta\boldsymbol{\Gamma}$) into the partial gradient $\nabla_{\boldsymbol{\Gamma}}\tilde{E}$. Since $\tilde{\mathbf{V}}^\perp = 0$ when **W** has been optimized for $\boldsymbol{\Gamma}$, the partial gradient $\nabla_{\boldsymbol{\Gamma}}\tilde{E}$ expression (38) simplifies into the following expression for the total shape gradient:

$$\nabla\tilde{E} = \left(h - \rho\dot{h}\right)\mathbf{N} \tag{43}$$

An interesting choice for the convex density penalty function h might be, for example,

$$h(p) = \begin{cases} \beta_{lo} \frac{(\rho - \rho_{lo})^2}{\rho}, & \rho < \rho_{lo} \\ 0, & \rho_{lo} \leq \rho \leq \rho_{hi} \\ \beta_{hi} (\rho - \rho_{hi})^2, & \rho > \rho_{hi} \end{cases}$$

where $\rho_{lo} = \min\big(\min_{x \in \Omega_0} \rho_0(x), \min_{x \in \Omega_1} \rho_1(x)\big)$ and $\rho_{hi} = \max\big(\max_{x \in \Omega_0} \rho_0(x), \max_{x \in \Omega_1} \rho_1(x)\big)$ denote the lower and upper bounds for a preferred (non-penalized) density range matching the combined ranges of the initial and final densities (to introduce a soft form of maximum principle and minimum principle), and where $\beta_{lo}, \beta_{hi} > 0$ represent weighting factors to tune the penalty strength outside of the preferred range. Here, h is chosen to be continuous in its first derivative (though not in its second derivative, which still remains non-negative for convexity)

$$\dot{h}(p) = \begin{cases} \beta_{lo} \left(\frac{\rho^2 - \rho_{lo}^2}{\rho^2} \right), & \rho < \rho_{lo} \\ 0, & \rho_{lo} \leq \rho \leq \rho_{hi} \\ 2\beta_{hi}(\rho - \rho_{hi}), & \rho > \rho_{hi} \end{cases} \qquad \ddot{h}(p) = \begin{cases} \frac{2\beta_{lo}\rho_{lo}^2}{\rho^3}, & \rho < \rho_{lo} \\ 0, & \rho_{lo} \leq \rho \leq \rho_{hi} \\ 2\beta_{hi}, & \rho > \rho_{hi} \end{cases}$$

as well as tend to infinity when $\rho \to \infty$ and when $\rho \to 0$. Plugging this into the total shape gradient expression (43) would yield

$$\nabla \tilde{E} = \left(h - \rho \dot{h} \right) \mathbf{N} = \begin{cases} 2\beta_{lo}\rho_{lo} \left(\frac{\rho_{lo} - \rho}{\rho} \right), & \rho < \rho_{lo} \\ 0, & \rho_{lo} \leq \rho \leq \rho_{hi} \\ \beta_{hi} \left(\rho_{hi}^2 - \rho^2 \right), & \rho > \rho_{hi} \end{cases}$$

which, if applied in the opposite (descent) direction, yields positive (outward) pressure along portions of the current support boundary $\boldsymbol{\Gamma}$ where the optimized density ρ exceeds the desired upper bound ρ_{hi} while yielding negative (inward) pressure along portions of $\boldsymbol{\Gamma}$ where the optimized density falls below the desired lower bound ρ_{lo}.

5. Results

We conclude with two experimental results which illustrate the benefits of this variational approach, stemming in particular from its separate yet coupled optimization of the compact spatiotemporal support and the density within. While the mathematical formulation of the approach is fully outlined here, the numerical implementation strategies are still under investigation and, as such, the following results are intended to be preliminary indications of what we may expect after more sophisticated numerical strategies have been further explored and developed.

5.1. Interpolation between Two Different Non-Convex Supports

In this first example, we tackle the problem of interpolating between cardiac images captured at two different moments within the heartbeat cycle shown on the left and the right in Figure 1 Notice that both cell boundaries represent non-convex shapes with several small concavities. We also see structures of interest inside the ventricle (papillary muscle cross sections), which not only move and deform along with the rest of the image but which also change in their topological appearance. At a coarse scale, the boundary shapes appear to be similar, but detailed inspection reveals that finer scale protrusions and intrusions around the boundary differ, especially the concavity along the upper right edge of the left image which disappears in the right image. Nevertheless, both boundaries exhibit a matching simple topology, which we would like to preserve when morphing one into the other. Numerically, this can be challenging without an explicit model of the support boundary, making it difficult to guarantee that small-scale protrusions do not break off during transport to yield transitional topological changes.
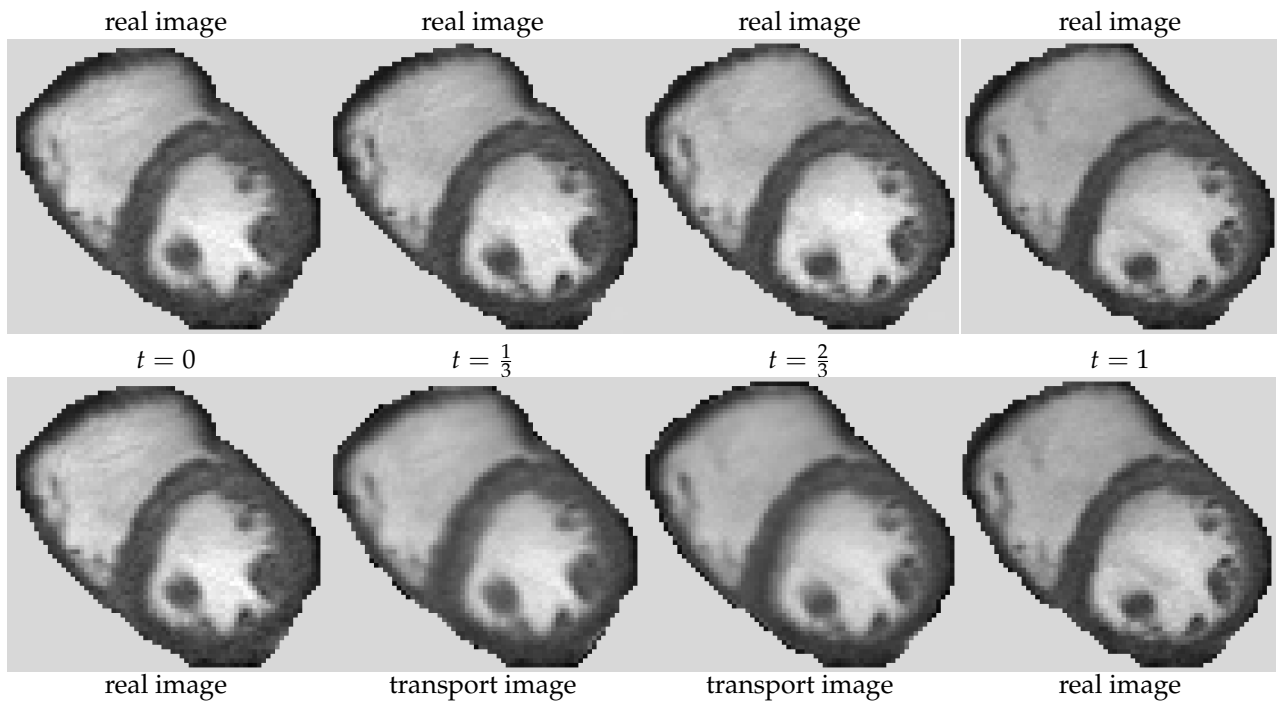
Figure 1. Density evolution between two non-convex, differing compact supports. Left and right input images ($t = 0$ and $t = 1$) show the starting and ending densities and supports (gray background does not represent any density value and has no effect on the computations), while the middle images (bottom) represent transported densities and supports computed at equally spaced intermediate times (with the actual measured cardiac images at corresponding times shown above for comparison).

Using these two cardiac images as the starting and ending densities at time $t = 0$ and $t = 1$, respectively, we solve the compact optimal transport problem with the variational approach outlined in this paper, using a 3D space-time grid for the solenoidal vector field $\mathbf{W}$ with 64 temporal slices, each of size 128×128 (same resolution as the two input images). A matching level set grid Ψ is used to represent the spatiotemporal support as the set $\Psi < 0$. We see the density (temporal component of $\mathbf{W}$) in Figure 1 at equally spaced intervals along the computed transport. We can see this more explicitly by visualizing the entire swept-out hypersurface $\mathbf{\Gamma}$ (the portion of the spatiotemporal support boundary strictly between 0 and 1) in Figure 2. This is very easily rendered as the zero level set of Ψ and clearly reveals a smooth homotopy connecting the two end curves, one exposed within the top rendering and the other within the bottom).
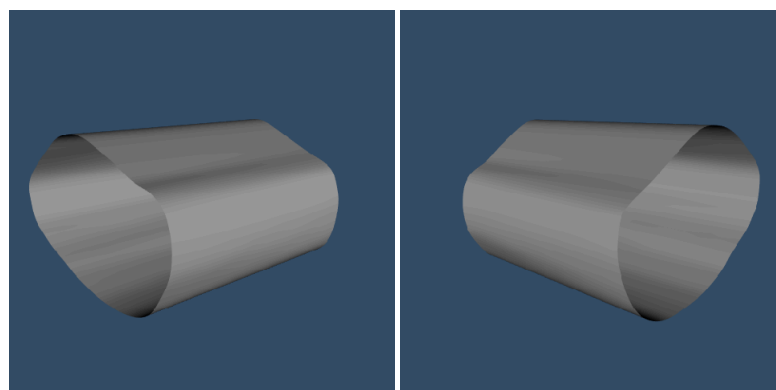

Figure 2. Two rendered viewpoints of the hypersurface $\mathbf{\Gamma}$ which constrains the dynamic support between $t = 0$ to $t = 1$.

5.2. Optimal Transport with Support Constraints

This next experiment illustrates through an intuitive toy example an important and useful extension to the class of optimal transport problems which, to the best of our knowledge, is not accommodated by prior methods but is easily handled within the methodology outlined here. Namely, how might one compute the optimal transport between densities subject to constraints imposed on where (and possibly even when) mass is allowed or not allowed to move along the way?

5.2.1. Known Densities with Intermediate Support Constraints

We can easily motivate the utility of such constraints if we go back to the classic problem posed by Gaspard Monge. In formulating the problem to optimally move a pile of dirt from one place to another, no constraints were imposed on the path taken by each portion of moved dirt. While the unconstrained optimal solution may yield a realistic and realizable transport strategy in several practical circumstances, this cannot always be guaranteed. For example, suppose the task is to move a pile of dirt across a limited number of bridges to the other side of a river. The unconstrained solution could easily yield an impractical transport strategy which involves crossing open portions of the river. A related larger-scale problem might involve planning the motion of massive numbers of land troops distributed over a set of territories to another set of territories, taking into account geographical barriers such as mountains and bodies of water as well as political barriers, which would render certain intermediate territories out of bounds.

Even when the topology of the desired transport is known, optimization subject to geometric constraints can still be non-trivial. For example, transporting a single pile of dirt across a single bridge which is narrower than the base of the pile is already an interesting problem. The unconstrained optimal transport will likely want to move some portion of dirt outside the confines of the bridge. Should all of that excess dirt simply be re-routed and accumulated along the closest edge of the bridge, or should some of it be moved more centrally inside the bridge, which would make the trajectory deviate even further from the unconstrained optimum but attenuate an otherwise massive spike in density along the edge? Clearly there is a trade-off that is not easily intuited directly from the unconstrained solution.

These types of constraints are easily imposed using the variational framework presented in this paper due to the explicit representation and separate-yet-coupled evolution of the support boundary. As long as we choose an initial spatiotemporal support that satisfies the provided set of spatiotemporal constraints, the calculated gradient descent evolution of the resulting spatiotemporal hypersurface can simply be set to zero locally wherever its application would otherwise violate the constraints.

5.2.2. Unknown Final Density But with Known Support

Another extension of problems that are easily accommodated by this coupled variational approach include scenarios where the support of the target density is given but the target density itself is unknown (and therefore part of the optimization). Such a problem can easily be transformed into the problem of a known final density with intermediate constraints on the support by treating the desired final density as the halfway point ($t = \frac{1}{2}$) in transporting the initial density at $t = 0$ back to itself again at $t = 1$. In this way, the desired final support becomes a constraint on the intermediate support instead. Optimization of this reconfigured problem will yield both a forward copy (from $t = 0$ to $t = \frac{1}{2}$) and a backward copy (from $t = \frac{1}{2}$ to $t = 1$) of the optimal transport for the original problem as well as the optimal target density itself at $t = \frac{1}{2}$. As such, from an implementation standpoint, this class of problems can be handled the same way as the class of problems just described above.

A practical application for this form of constrained optimal transport would be to compute the most efficient evacuation strategy to clearing mass out of a given subregion while keeping it within a larger encompassing region that already contains pre-existing

mass. In this case, we know the initial density and support, and we know the final support, simply the initial support minus the subregion to be evacuated, but we do not know or otherwise want to constrain the resulting new density within the now reduced support. We therefore seek the least costly way (according to the Wasserstein metric) to redistribute the mass originally contained within the subregion into its surrounding, already occupied neighborhood. Such a problem may arise, for example, when seeking to clear extensive zones of all materials and/or personnel while keeping them within the confines of larger zones whose occupants are free to be internally relocated if needed.

We illustrate precisely this scenario with an intuitive toy problem in Figure 3. We start with uniformly distributed mass at $t = 0$ with density 3.0 across a disk representing the global confines. In turn, we define the target distribution at $t = 1$ to be the same but impose a hole in the intermediate support at $t = \frac{1}{2}$ within the center of the disk. Solving the constrained optimal transport problem between these matching uniform distributions causes the initially filled hole to be evacuated from $t = 0$ to $t = \frac{1}{2}$ as illustrated from left-to-right and top-to-bottom in Figure 3, and then to be refilled from $t = \frac{1}{2}$ to $t = 1$ as also illustrated in Figure 3 when read in reverse. The optimal redistribution of mass is attained at $t = \frac{1}{2}$ and is displayed in grayscale at the end of the figure.

Even in this simplest illustrative example with constant density and maximal symmetry, it is by no means intuitively obvious how far away mass should be displaced from the hole boundary compared to how much it should be allowed to accumulate along the boundary. In fact, the density would become infinite if the evacuated mass were to remain strictly along the boundary. To obtain a better sense of where displaced mass accumulates upon pre-existing mass (displaced mass includes mass evacuated from the hole as well as mass moved away to make room), we show the net density change in Figure 4 (left side), which attains a maximum rise of 2.3 all around the boundary of the evacuated hole and gradually rolls off further outward.
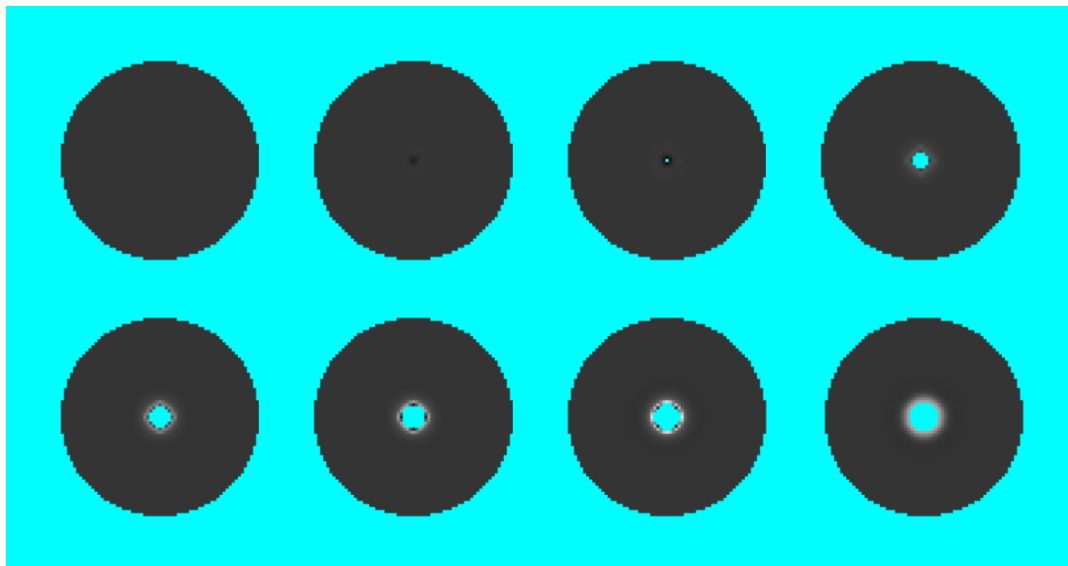

Figure 3. Example of optimal transport with spatiotemporal support constraints. Mass is evacuated from the center of a disk with uniform initial density of 3.0 (appears as dark-gray/almost-black in the top-left frame), while being constrained to remain inside the disk. The optimal evacuation strategy is computed by solving the optimal transport problem between the initial uniform density on the disk back to itself (normally the trivial transport of zero) but imposing a hole in the spatiotemporal support at $t = \frac{1}{2}$. Mass therefore evacuates the hole from $t = 0$ to $t = \frac{1}{2}$ (shown above) and refills the hole from $t = \frac{1}{2}$ to $t = 1$ (reverse of above).
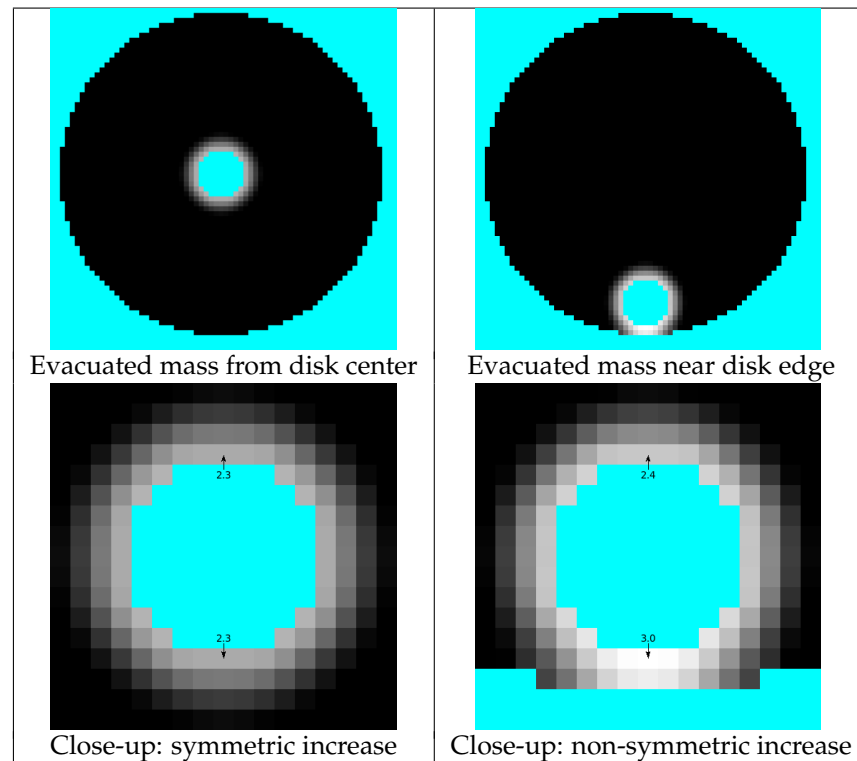
Figure 4. Optimized mass evacuation from two different locations. Both cases begin with uniform mass density (3.0) over the disk then evacuate a hole. Accumulated density results wherever mass is relocated (see previous figure). Final net accumulation is shown here as the rise in density. When the hole is perfectly centered (**left**), evacuated mass is redistributed symmetrically with a peak density rise of 2.3 (77%). Non-symmetric redistribution produces a peak rise of 3.0 when the hole is created near the boundary (**right**).

We can make this toy problem even more interesting by evacuating mass near the boundary of the disk rather than its center. Looking at the resulting net density increase in Figure 4 (right side), we can make several observations. First, as expected, the redistribution is no longer symmetric. The symmetry is broken two different ways. First, the rise in density is much higher (3.0, a full 100% jump) along the bottom border of the hole compared to the top border. This is unsurprising since there is not as much room to move away from the hole, and so the same amount of evacuated mass distributed over a thinner local neighborhood must necessarily result in a larger accumulated density.

However, we also notice that the density jump near the top of the hole (2.4), while lower than the bottom, still exceeds the symmetric jump (2.3) when the hole was centered inside the disk. This differential grows as we travel along the border of the hole toward the bottom. This means that some of the mass within the lower half of the hole, which was all evacuated downward from the centered hole, actually was evacuated upward from the hole near the boundary of the confining domain. As such, at least two interesting interplays are relevant in this optimization. First, as in the symmetric case, how far away should mass be displaced away from borders versus accumulated along borders? Second, when displacement distances are limited, what is the right trade-off between higher accumulation at nearby borders versus more costly redirection toward farther borders of the region to be evacuated? These considerations are both naturally handled by this coupled variational framework.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Detailed First Variation and Gradient Calculation

We first note that a variation of $\mathbf{W}$ yields a variation of $\mathbf{V}$ which is orthogonal to $\mathbf{W}$ itself. This may be demonstrated directly using the expression (18) as follows:

$$\delta\mathbf{V} = \left(\frac{\delta\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)} - \frac{\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)^2}(\delta\mathbf{W}\cdot\mathbf{e}_0) \right) - \left(\frac{\delta\mathbf{W}\cdot\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)^2} - \frac{\mathbf{W}\cdot\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)^3}(\delta\mathbf{W}\cdot\mathbf{e}_0) \right)\mathbf{e}_0$$

$$\mathbf{W}\cdot\delta\mathbf{V} = \underbrace{\left(\frac{\delta\mathbf{W}\cdot\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)} - \frac{\mathbf{W}\cdot\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)^2}(\delta\mathbf{W}\cdot\mathbf{e}_0) \right) - \left(\frac{\delta\mathbf{W}\cdot\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)} - \frac{\mathbf{W}\cdot\mathbf{W}}{(\mathbf{W}\cdot\mathbf{e}_0)^2}(\delta\mathbf{W}\cdot\mathbf{e}_0) \right)}_{\text{terms cancel}} = 0$$

Using Lagrange multipliers to incorporate the mass preservation constraints, we write the energy

$$E(\mathbf{W},\mathbf{\Gamma},\lambda,\alpha) = \int_{\mathbf{\Omega}}\mathbf{W}\cdot\mathbf{V} + \lambda\,\nabla\cdot\mathbf{W}\,d\mathbf{X} + \int_{\mathbf{\Gamma}}\alpha\mathbf{W}\cdot\mathbf{N}\,d\mathbf{S} + \int_{\mathbf{\Gamma}_0}\alpha_0\left(\overbrace{\mathbf{W}\cdot\mathbf{N}}^{-\mathbf{W}\mathbf{e}_0} + \rho_0 \right)dx + \int_{\mathbf{\Gamma}_1}\alpha_1\left(\overbrace{\mathbf{W}\cdot\mathbf{N}}^{\mathbf{W}\mathbf{e}_0} - \rho_1 \right)dx$$

and note that perturbations $\delta\Gamma_{[0]}$ and $\delta\Gamma_{[1]}$ along the swept-out hypersurface boundaries at $t=0$ and $t=1$ are coupled to perturbations along the temporal face boundaries $\delta(\partial\mathbf{\Gamma}_0)$ and $\delta(\partial\mathbf{\Gamma}_1)$ as follows.

$$\delta(\partial\mathbf{\Gamma}_0) = (0,\delta\Gamma_{[0]})$$
$$\delta(\partial\mathbf{\Gamma}_1) = (0,\delta\Gamma_{[1]})$$

We now compute its first variation

$$\delta E = \int_{\mathbf{\Omega}} \delta \mathbf{W} \cdot \mathbf{V} + \underbrace{\mathbf{W} \cdot \delta \mathbf{V}}_{0} + \lambda \nabla \cdot \delta \mathbf{W} + \delta \lambda \, \nabla \cdot \mathbf{W} \, d\mathbf{X} + \int_{\mathbf{\Gamma}} (\mathbf{W} \cdot \mathbf{V} + \lambda \nabla \cdot \mathbf{W}) \delta \mathbf{\Gamma} \cdot \mathbf{N} \, d\mathbf{S}$$

$$+ \int_{\mathbf{\Gamma}} \delta\alpha \, \mathbf{W} \cdot \mathbf{N} + \alpha \, \delta \mathbf{W} \cdot \mathbf{N} \, d\mathbf{S} + \int_{\mathbf{\Gamma}} \underbrace{(\nabla_S \alpha \cdot \mathbf{W} + \alpha \nabla \cdot \mathbf{W})}_{\nabla_S(\alpha \mathbf{W})} \delta \mathbf{\Gamma} \cdot \mathbf{N} \, d\mathbf{S}$$

$$+ \int_{\Gamma_{[0]}} \alpha \mathbf{W} \cdot \left(\overbrace{\left(\delta \Gamma_{[0]} \cdot N \right) \underbrace{(0, N) \times (0, T)}_{\mathbf{e}_0}}^{\delta(\partial \mathbf{\Gamma}_0) \times (0,T) = (0, \delta \Gamma_{[0]}) \times (0,T)} \right) dS_{[0]} + \int_{\Gamma_{[1]}} \alpha \mathbf{W} \cdot \left(\overbrace{\left(\delta \Gamma_{[1]} \cdot N \right) \underbrace{(0, N) \times (0, T)}_{-\mathbf{e}_0}}^{\delta(\partial \mathbf{\Gamma}_1) \times (0,T) = (0, \delta \Gamma_{[1]}) \times (0,T)} \right) dS_{[1]}$$

$$- \int_{\mathbf{\Gamma}_0} \delta\alpha_0 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_0) + \alpha_0 \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx - \int_{\Gamma_{[0]}} \alpha_0 (\mathbf{W}(0, x) \cdot \mathbf{e}_0 - \rho_0) \, \delta \Gamma_{[0]} \cdot N \, dS_{[0]}$$

$$+ \int_{\mathbf{\Gamma}_1} \delta\alpha_1 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_1) + \alpha_1 \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx + \int_{\Gamma_{[1]}} \alpha_1 (\mathbf{W}(1, x) \cdot \mathbf{e}_0 - \rho_1) \, \delta \Gamma_{[1]} \cdot N \, dS_{[1]}$$

$$= \int_{\mathbf{\Omega}} \delta \mathbf{W} \cdot (\mathbf{V} - \nabla \lambda) + \delta \lambda \, \nabla \cdot \mathbf{W} \, d\mathbf{X} + \int_{\partial \mathbf{\Omega}} \lambda \, \delta \mathbf{W} \cdot \mathbf{N} \, d\mathbf{S}$$

$$+ \int_{\mathbf{\Gamma}} \delta\alpha \, \mathbf{W} \cdot \mathbf{N} + \alpha \, \delta \mathbf{W} \cdot \mathbf{N} + (\mathbf{W} \cdot (\mathbf{V} + \nabla_S \alpha) + (\lambda + \alpha) \nabla \cdot \mathbf{W}) \delta \mathbf{\Gamma} \cdot \mathbf{N} \, d\mathbf{S}$$

$$- \int_{\mathbf{\Gamma}_0} \delta\alpha_0 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_0) + \alpha_0 \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx - \int_{\Gamma_{[0]}} (\alpha_0 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_0) - \alpha \mathbf{W} \cdot \mathbf{e}_0) \, \delta \Gamma_{[0]} \cdot N \, dS_{[0]}$$

$$+ \int_{\mathbf{\Gamma}_1} \delta\alpha_1 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_1) + \alpha_1 \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx + \int_{\Gamma_{[1]}} (\alpha_1 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_1) - \alpha \mathbf{W} \cdot \mathbf{e}_0) \, \delta \Gamma_{[1]} \cdot N \, dS_{[1]}$$

$$= \int_{\mathbf{\Omega}} \delta \mathbf{W} \cdot (\mathbf{V} - \nabla \lambda) + \delta \lambda \, \nabla \cdot \mathbf{W} \, d\mathbf{X}$$

$$+ \int_{\mathbf{\Gamma}} \delta\alpha \, \mathbf{W} \cdot \mathbf{N} + (\lambda + \alpha) \, \delta \mathbf{W} \cdot \mathbf{N} + (\mathbf{W} \cdot (\mathbf{V} + \nabla_S \alpha) + (\lambda + \alpha) \nabla \cdot \mathbf{W}) \delta \mathbf{\Gamma} \cdot \mathbf{N} \, d\mathbf{S}$$

$$- \int_{\mathbf{\Gamma}_0} \delta\alpha_0 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_0) + (\lambda + \alpha_0) \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx + \int_{\mathbf{\Gamma}_1} \delta\alpha_1 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_1) + (\lambda + \alpha_1) \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx$$

$$- \int_{\Gamma_{[0]}} (\alpha_0 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_0) - \alpha \mathbf{W} \cdot \mathbf{e}_0) \delta \Gamma_{[0]} \cdot N \, dS_{[0]}$$

$$+ \int_{\Gamma_{[1]}} (\alpha_1 (\mathbf{W} \cdot \mathbf{e}_0 - \rho_1) - \alpha \mathbf{W} \cdot \mathbf{e}_0) \delta \Gamma_{[1]} \cdot N \, dS_{[1]}$$

Maintaining the mass conservation and initial/final density constraints eliminates the dependence on $\delta\lambda$, $\delta\alpha$, $\delta\alpha_0$, and $\delta\alpha_1$, yielding the simpler expression

$$\delta E = \int_{\mathbf{\Omega}} \delta \mathbf{W} \cdot (\mathbf{V} - \nabla \lambda) \, d\mathbf{X} + \int_{\mathbf{\Gamma}} (\lambda + \alpha) \, \delta \mathbf{W} \cdot \mathbf{N} + \mathbf{W} \cdot (\mathbf{V} + \nabla_S \alpha) \delta \mathbf{\Gamma} \cdot \mathbf{N} \, d\mathbf{S}$$

$$- \int_{\mathbf{\Gamma}_0} (\lambda + \alpha_0) \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx + \int_{\mathbf{\Gamma}_1} (\lambda + \alpha_1) \, \delta \mathbf{W} \cdot \mathbf{e}_0 \, dx$$

$$+ \int_{\Gamma_{[0]}} \alpha \underbrace{\mathbf{W} \cdot \mathbf{e}_0}_{\rho_0} \underbrace{\sqrt{1 + \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N \right)^2} \delta \mathbf{\Gamma} \cdot \mathbf{N}}_{\delta \Gamma_{[0]} \cdot N} \, dS_{[0]} - \int_{\Gamma_{[1]}} \alpha \underbrace{\mathbf{W} \cdot \mathbf{e}_0}_{\rho_1} \underbrace{\sqrt{1 + \left(\frac{\partial \Gamma_{[t]}}{\partial t} \cdot N \right)^2} \delta \mathbf{\Gamma} \cdot \mathbf{N}}_{\delta \Gamma_{[1]} \cdot N} \, dS_{[1]}$$

We now choose $\alpha = -\lambda$ such that the joint sensitivity of the boundary term with respect to $\delta \mathbf{W} \cdot \mathbf{N}$ and $\delta \mathbf{\Gamma} \cdot \mathbf{N}$ is expressed purely in terms of $\delta \mathbf{\Gamma} \cdot \mathbf{N}$. This is possible since the vanishing flux condition for $\mathbf{W} \cdot \mathbf{N} = 0$ along $\mathbf{\Gamma}$ introduces an explicit coupling (A2)

between the normal perturbations $\delta\mathbf{W} \cdot \mathbf{N}$ and $\delta\boldsymbol{\Gamma} \cdot \mathbf{N}$, which is derived in Appendix B. Further noting that $\mathbf{W} \cdot \nabla_S \lambda = \mathbf{W} \cdot \nabla\lambda$ by the same constraint $\mathbf{W} \cdot \mathbf{N} = 0$, we obtain

$$
\begin{aligned}
\delta E = &\int_{\boldsymbol{\Omega}} \delta\mathbf{W} \cdot (\mathbf{V} - \nabla\lambda) \, d\mathbf{X} + \int_{\boldsymbol{\Gamma}} \mathbf{W} \cdot (\mathbf{V} - \nabla\lambda)\delta\boldsymbol{\Gamma} \cdot \mathbf{N} \, dS \\
&- \int_{\boldsymbol{\Gamma}_0} (\lambda + \alpha_0) \, \delta\mathbf{W} \cdot \mathbf{e}_0 \, dx + \int_{\boldsymbol{\Gamma}_1} (\lambda + \alpha_1) \, \delta\mathbf{W} \cdot \mathbf{e}_0 \, dx \\
&- \int_{\boldsymbol{\Gamma}_{[0]}} \lambda \, \rho_0 \underbrace{\sqrt{1 + \left(\frac{\partial\boldsymbol{\Gamma}_{[t]}}{\partial t} \cdot N\right)^2} \delta\boldsymbol{\Gamma} \cdot \mathbf{N} \, dS_{[0]}}_{\delta\Gamma_{[0]} \cdot N} + \int_{\boldsymbol{\Gamma}_{[1]}} \lambda \, \rho_1 \underbrace{\sqrt{1 + \left(\frac{\partial\boldsymbol{\Gamma}_{[t]}}{\partial t} \cdot N\right)^2} \delta\boldsymbol{\Gamma} \cdot \mathbf{N} \, dS_{[1]}}_{\delta\Gamma_{[1]} \cdot N}
\end{aligned}
$$

Finally the middle row drops out if we impose $\delta\mathbf{W} \cdot \mathbf{e}_0 = 0$ along $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ to maintain the initial and final densities ρ_0 and ρ_1. Note, however, that while we have treated strictly positive density functions and their spatial supports as independent entities, we have, on the other hand, only modeled the variational problem here with constraints on $\mathbf{W}$ to maintain the given density values ρ_0 and ρ_1 at $t = 0$ and $t = 1$ but have not introduced corresponding geometrical constraints on $\boldsymbol{\Gamma}$ to maintain their given supports. As a consequence, we obtained the integrals in the bottom row along the shared spatial boundary $\boldsymbol{\Gamma}_{[0]} = \partial\boldsymbol{\Omega}_{[0]} = \partial\boldsymbol{\Omega}_0$, where $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$ come together at $t = 0$, as well as along the shared spatial boundary $\boldsymbol{\Gamma}_{[1]} = \partial\boldsymbol{\Omega}_{[1]} = \partial\boldsymbol{\Omega}_1$, where $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_1$ come together at $t = 1$. Just as the middle row drops out by imposing $\delta\mathbf{W} \cdot \mathbf{e}_0 = 0$ at $t = 0$ and $t = 1$, the bottom row will also drop out by imposing $\delta\boldsymbol{\Gamma} = 0$ at $t = 0$ and $t = 1$ as well.

Appendix B. Coupled Boundary and Flux Perturbations

To maintain the vanishing flux constraint $\mathbf{W} \cdot \mathbf{N} = 0$ along the hypersurface $\boldsymbol{\Gamma}$ portion of the support boundary $\partial\boldsymbol{\Omega}$, the normal perturbation $\delta\boldsymbol{\Gamma} \cdot \mathbf{N}$ of the boundary itself and the normal component of the solenoidal field perturbation $\delta\mathbf{W} \cdot \mathbf{N}$ cannot be applied independently but are coupled. This should not be surprising because the field $\mathbf{W}$ defines the transport which, by virtue of determining the density evolution, also determines the evolution of its support. To determine the resulting coupling between a support perturbation $\delta\boldsymbol{\Gamma} \cdot \mathbf{N}$ and the matching flux perturbation $\delta\mathbf{W} \cdot \mathbf{N}$, we differentiate the following constraint along the swept-out hypersurface:

$$
\mathbf{W}\big(\boldsymbol{\Gamma}(s)\big) \cdot \mathbf{N}(s) = 0 \tag{A1}
$$

where

$$
s = (s_1, \ldots, s_n)
$$

denotes isothermal coordinates aligned with the principal directions $\mathbf{T}_1, \ldots, \mathbf{T}_n$ (unit tangent vectors) of the hypersurface. We choose this parameterization for the convenient property that geodesic torsions vanish along principal directions, and therefore

$$
\frac{\partial\mathbf{T}_j}{\partial s_i} = \begin{cases} \kappa_i\mathbf{N}, & i = j \\ 0, & i \neq j \end{cases}
$$

where κ_i denotes the principle curvature. We now expand

$$
\delta(\mathbf{W} \cdot \mathbf{N}) = \left(\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\delta\boldsymbol{\Gamma} + \delta\mathbf{W}\right) \cdot \mathbf{N} + \mathbf{W} \cdot \delta\mathbf{N} = 0
$$

to obtain

$$\delta\mathbf{W}\cdot\mathbf{N} = -\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\delta\mathbf{\Gamma} - \mathbf{W}\cdot\delta\mathbf{N}$$

$$= \underbrace{-\left(\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{N}\right)\delta\mathbf{\Gamma}\cdot\mathbf{N} - \sum_{i=1}^n\left(\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i}_{\text{orthogonal components of } -\mathbf{N^T}\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\delta\mathbf{\Gamma}} - \mathbf{W}\cdot\underbrace{\left(\sum_{i=1}^n -\mathbf{T}_i\left(\frac{\partial(\delta\mathbf{\Gamma})}{\partial s_i}\cdot\mathbf{N}\right)\right)}_{\delta\mathbf{N}}$$

$$= \left(\overbrace{\nabla\cdot\mathbf{W}}^{0} - \mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{N}\right)\delta\mathbf{\Gamma}\cdot\mathbf{N} - \sum_{i=1}^n\left(\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i + \sum_{i=1}^n\mathbf{W}\cdot\mathbf{T}_i\left(\frac{\partial(\delta\mathbf{\Gamma})}{\partial s_i}\cdot\mathbf{N}\right)$$

$$= \left(\sum_{i=1}^n\mathbf{T}_i^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\delta\mathbf{\Gamma}\cdot\mathbf{N} - \sum_{i=1}^n\left(\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i + \sum_{i=1}^n\left(\frac{\partial(\delta\mathbf{\Gamma})}{\partial s_i}\cdot\mathbf{N}\right)(\mathbf{W}\cdot\mathbf{T}_i)$$

$$= \left(\sum_{i=1}^n\mathbf{T}_i^T\overbrace{\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i}^{\frac{\partial\mathbf{W}}{\partial s_i}} + \overbrace{\mathbf{W}\cdot\frac{\partial\mathbf{T}_i}{\partial s_i}}^{\kappa_i\mathbf{W}\cdot\mathbf{N}=0}\right)\delta\mathbf{\Gamma}\cdot\mathbf{N} - \sum_{i=1}^n\left(\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i + \sum_{i=1}^n\left(\frac{\partial(\delta\mathbf{\Gamma})}{\partial s_i}\cdot\mathbf{N}\right)(\mathbf{W}\cdot\mathbf{T}_i)$$

$$= \sum_{i=1}^n\frac{\partial}{\partial s_i}(\mathbf{W}\cdot\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{N} - \left(\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i + \left(\frac{\partial(\delta\mathbf{\Gamma})}{\partial s_i}\cdot\mathbf{N}\right)(\mathbf{W}\cdot\mathbf{T}_i)$$

Differentiating the vanishing flux condition (A1) along each principal direction yields

$$\frac{\partial}{\partial s_i}(\mathbf{W}\cdot\mathbf{N}) = \left(\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i\right)\cdot\mathbf{N} - \mathbf{W}\cdot\kappa_i\mathbf{T} = 0$$

allowing us to substitute

$$\mathbf{N}^T\frac{\partial\mathbf{W}}{\partial\mathbf{X}}\mathbf{T}_i = \mathbf{W}\cdot\kappa_i\mathbf{T}_i$$

into our previous expression and to continue as follows

$$\delta\mathbf{W}\cdot\mathbf{N} = \sum_{i=1}^n\frac{\partial}{\partial s_i}(\mathbf{W}\cdot\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{N} - (\mathbf{W}\cdot\kappa_i\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i + \left(\frac{\partial(\delta\mathbf{\Gamma})}{\partial s_i}\cdot\mathbf{N}\right)(\mathbf{W}\cdot\mathbf{T}_i)$$

$$= \sum_{i=1}^n\frac{\partial}{\partial s_i}(\mathbf{W}\cdot\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{N} - (\mathbf{W}\cdot\kappa_i\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i + \frac{\partial\left(\sum_{j=1}^n(\delta\mathbf{\Gamma}\cdot\mathbf{T}_j)\mathbf{T}_j + (\delta\mathbf{\Gamma}\cdot\mathbf{N})\mathbf{N}\right)}{\partial s_i}\cdot\mathbf{N}(\mathbf{W}\cdot\mathbf{T}_i)$$

$$= \sum_{i=1}^n\frac{\partial}{\partial s_i}(\mathbf{W}\cdot\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{N} - \kappa_i(\mathbf{W}\cdot\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{T}_i$$

$$+ \left(\sum_{j\neq i}(\delta\mathbf{\Gamma}\cdot\mathbf{T}_j)\underbrace{\frac{\partial\mathbf{T}_j}{\partial s_i}}_{0} + (\delta\mathbf{\Gamma}\cdot\mathbf{T}_i)\underbrace{\frac{\partial\mathbf{T}_i}{\partial s_i}}_{\kappa_i} + \frac{\partial}{\partial s_i}(\delta\mathbf{\Gamma}\cdot\mathbf{N})\right)(\mathbf{W}\cdot\mathbf{T}_i)$$

$$= \sum_{i=1}^n\frac{\partial}{\partial s_i}(\mathbf{W}\cdot\mathbf{T}_i)\delta\mathbf{\Gamma}\cdot\mathbf{N} + \frac{\partial}{\partial s_i}(\delta\mathbf{\Gamma}\cdot\mathbf{N})(\mathbf{W}\cdot\mathbf{T}_i)$$

yielding the final expression for the coupling between $\delta\mathbf{W}\cdot\mathbf{N}$ and $\delta\mathbf{\Gamma}\cdot\mathbf{N}$:

$$\delta\mathbf{W}\cdot\mathbf{N} = \sum_{i=1}^n\frac{\partial}{\partial s_i}\left((\mathbf{W}\cdot\mathbf{T}_i)(\delta\mathbf{\Gamma}\cdot\mathbf{N})\right) \tag{A2}$$

It is generally difficult to invert this expression to express the boundary perturbation $\delta\boldsymbol{\Gamma} \cdot \mathbf{N}$ as a function of the flux perturbation $\delta\mathbf{W} \cdot \mathbf{N}$. However, in the special case of zero flux perturbation $\delta\mathbf{W} \cdot \mathbf{N} = 0$, we obtain

$$\sum_{i=1}^{n} \frac{\partial}{\partial s_i}\Big((\mathbf{W} \cdot \mathbf{T}_i)(\delta\boldsymbol{\Gamma} \cdot \mathbf{N}) \Big) = 0$$

which, combined with the constraint that $\delta\boldsymbol{\Gamma} \cdot \mathbf{N} = 0$ along the temporal boundaries of $\boldsymbol{\Gamma}$ at $t = 0$ and at $t = 1$ only admits the solution $\delta\boldsymbol{\Gamma} \cdot \mathbf{N} = 0$ along the entirety of the hypersurface $\boldsymbol{\Gamma}$. As such, a vanishing flux perturbation implies a vanishing perturbation of the support boundary. The converse is also trivially demonstrated directly through Equation (A1).

References

1. Monge, G. *Mémoire sur la Théorie des Déblais et des Remblais*; Histoire de l'Académie Royale des Sciences de Paris; Imprimerie Royale: Paris, France, 1781.
2. Kantorovich, L.V. On a problem of monge. *CR (Dokl.) Acad. Sci. URSS (NS)* **1948**, *3*, 225–226. [CrossRef]
3. Gangbo, W.; McCann, R. The geometry of optimal transportation. *Acta Math.* **1996**, *177*, 113–161. [CrossRef]
4. Villani, C. *Topics in Optimal Transportation*; American Mathematical Soc.: Ann Arbor, MI, USA, 2003.
5. Villani, C. *Optimal Transport: Old and New*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; Volume 338.
6. Evans, L.C.; Gangbo, W. Differential equations methods for the Monge-Kantorovich mass transfer problem. In *Memoirs of the American Mathematical Society*; American Mathematical Soc.: Ann Arbor, MI, USA, 1999; Volume 137, no. 653.
7. Rachev, S.T.; Rüschendorf, L. *Mass Transportation Problems: Volume I: Theory*; Probability and Its Applications; Springer: Berlin/Heidelberg, Germany, 1998.
8. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
9. Carlen, E.A.; Maas, J. An analog of the 2-Wasserstein metric in non-commutative probability under which the Fermionic Fokker–Planck equation is gradient flow for the entropy. *Commun. Math. Phys.* **2014**, *331*, 887–926. [CrossRef]
10. Haker, S.; Zhu, L.; Tannenbaum, A.; Angenent, S. Optimal mass transport for registration and warping. *Int. J. Comput. Vis.* **2004**, *60*, 225–240. [CrossRef]
11. Mittnenzweig, M.; Mielke, A. An entropic gradient structure for lindblad equations and generic for quantum systems coupled to macroscopic models. *arXiv*, **2016**, arXiv:1609.05765.
12. Rachev, S.T.; Rüschendorf, L. *Mass Transportation Problems: Volumes I and II*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998.
13. Mathews, J.C.; Nadeem, S.; Pouryahya, M.; Belkhatir, Z.; Deasy, J.O.; Levine, A.J.; Tannenbaum, A.R. Functional network analysis reveals an immune tolerance mechanism in cancer. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 16339–16345. [CrossRef] [PubMed]
14. Benamou, J.-D.; Brenier, Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **2000**, *84*, 375–393. [CrossRef]
15. Otto, F. The geometry of dissipative evolution equations: the porous medium equation. *Commun. Partial. Differ. Equ.* **2001**, *26*, 101–174. [CrossRef]
16. Zhu, L.; Yang, Y.; Haker, S.; Tannenbaum, A. An image morphing technique based on optimal mass preserving mapping. *IEEE Trans. Image Process.* **2007**, *16*, 1481–1495. [CrossRef] [PubMed]
17. Haker, S.; Tannenbaum, A.; Kikinis, R. Mass preserving mappings and image registration. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001: 4th International Conference Utrecht, The Netherlands, 14–17 October 2001; Proceedings 4*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 120–127.
18. Chen, Y.; Georgiou, T.; Pavon, M. On the Relation Between Optimal Transport and Schrödinger Bridges: A Stochastic Control Viewpoint. *J. Optim. Theory Appl.* **2016**, *169*, 671–691. [CrossRef]
19. Osher, S.; Fedkiw, R. *Level Set Methods and Dynamic Implicit Surfaces*; Springer: Berlin/Heidelberg, Germany, 2003.
20. Sethian, J. *Level Set Methods Fast Marching Methods*; Cambridge University Press: Cambridge, UK, 1999.
21. Sapiro, G. *Geometric Partial Equations and Image Analysis*; Cambridge University Press: Cambridge, UK, 2001.
22. Susskind, L.; Friedman, A. *Special Relativity and Classical Field Theory: The Theoretical Minimum*; Hachette Book Group: New York, NY, USA, 2017.