HAL open science

A mixed finite element discretization of dynamical optimal transport

Andrea Natale, Gabriele Todeschi

▶ **To cite this version:**

Andrea Natale, Gabriele Todeschi. A mixed finite element discretization of dynamical optimal transport. 2020. hal-02501634v3

HAL Id: hal-02501634

https://hal.science/hal-02501634v3

Preprint submitted on 12 Oct 2020 (v3), last revised 22 Feb 2022 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MIXED FINITE ELEMENT DISCRETIZATION OF DYNAMICAL OPTIMAL TRANSPORT

ANDREA NATALE AND GABRIELE TODESCHI

Abstract. In this paper we introduce a new class of finite element discretizations of the quadratic optimal transport problem based on its dynamical formulation. These generalize to the finite element setting the finite difference scheme proposed by Papadakis et al. [SIAM J Imaging Sci, 7(1):212–238,2014]. We solve the discrete problem using a proximal splitting approach and we show how to modify this in the presence of regularization terms which are relevant for physical data interpolation.

1. Introduction

Optimal transport provides a convenient framework for density interpolation as a convex optimization problem. Its most remarkable feature is its sensitivity to horizontal displacement, which generally allows one to retrieve translations when interpolating between two densities. This property has motivated the application of optimal transport to many imaging problems, especially in the context of physical sciences and fluid dynamics. A typical example comes from satellite image interpolation in oceanography. In this case, one is interested in reconstructing the evolution of a quantity of interest such as Sea Surface Temperature (SST) or Sea Surface Height (SSH) between two given observations. As highlighted in [19], for this type of applications one needs to include appropriate regularization terms to avoid the appearance of unphysical phenomena such as mass concentration in the reconstructed density evolution.

In this paper we propose a finite element approach to solve the dynamical formulation of optimal transport with quadratic cost which is numerically stable on unstructured meshes (and therefore can be easily implemented on complex domains) and that can be easily modified to include different type of regularizations which are relevant for the dynamic reconstruction and interpolation of physical quantities. For some choices of finite element spaces, using the framework introduced in [22], we can prove convergence of our discrete solutions to the ones of the continuous problem.

The dynamical formulation of optimal transport inspired some of the first numerical methods for this problem. This reads as follows: given two probability measures $\rho_0, \rho_1 \in \mathcal{P}(D)$ on a compact domain $D \subset \mathbb{R}^d$, find the curve $t \in [0,1] \mapsto \rho(t,\cdot) \in \mathcal{P}(D)$ which solves

$$\inf_{\rho,v} \left\{ \int_0^1 \int_D \frac{|v(t,\cdot)|^2}{2} \mathrm{d}\rho(t,\cdot)\mathrm{d}t \,;\, \partial_t \rho + \mathrm{div}_x(\rho v) = 0, \rho(0) = \rho_0,\, \rho(1) = \rho_1 \right\} \tag{1.1}$$

where $v : [0,1] \times D \to \mathbb{R}^d$ is a time-dependent velocity field on D tangent to the boundary ∂D, and $|\cdot|$ denotes the Euclidean norm. In other words, problem (1.1) selects the curve of minimal kinetic energy with fixed endpoints ρ_0 and ρ_1.

Benamou and Brenier [5] realized that introducing the momentum $m := \rho v$, problem (1.1) can be recast into a convex optimization problem in the variables

Date: October 12, 2020.

(ρ, m), with a linear constraint, since the continuity equation becomes

$$\partial_t \rho + \mathrm{div}_x m = 0\,. \tag{1.2}$$

If we define $\sigma := (\rho, m)$, regarded as a measure on $[0,1] \times D$, this constraint is equivalent to $\mathrm{div}\,\sigma = 0$, where now div denotes the divergence operator on the space-time domain $[0,1] \times D$. Introducing the dual variable $q = (a, b)$ wher $a \in C([0,1] \times D)$ and $b \in C([0,1] \times D; \mathbb{R}^d)$, the kinetic energy minimized in (1.1) can be written in the form

$$\sup_q \left\{ \int_0^1 \int_D q \cdot \mathrm{d}\sigma\,;\, a + \frac{|b|^2}{2} \le 0 \right\}. \tag{1.3}$$

Combining this expression with (1.1) we obtain a saddle point problem in the variables (q, σ) with a nonlinear constraint on q and a linear one on σ.

The numerical method proposed in [5] involves discretizing q and σ by their values on a regular grid, and expressing the constraint on σ via a Lagrange multiplier; then the dual problem can be solved by an Augmented Lagrangian ADMM approach, optimizing separately in q and the Lagrange multiplier and then performing a gradient ascent step on σ. Disregarding the discretization in space-time, the convergence of the method has been studied in [17, 20]. The same approach was used to discretize different problems related to optimal transport (e.g., gradient flows [8], mean field games [7], unbalanced optimal transport [15]) using a finite element discretization in space-time. Importantly, in these cases the numerical method is obtained by discretizing the several steps of the augmented Lagrangian approach rather than as a discrete optimization algorithm. This implies that in general it is difficult to establish the convergence of the discrete algorithms. Moreover, for these type of methods, convergence results towards the continuous solutions with mesh refinement are only available for specific settings (e.g., the L^1-type optimal transport problems studied in [21]), but they are not available for the optimal transport problem (1.1).

Papadakis, Peyré, and Oudet proposed in [27] a staggered finite difference discretization on regular grids of the optimal transport problem (1.1), and they considered different proximal splitting algorithms to solve it. The computational bottleneck for these methods as well as for the original augmented Lagrangian approach is the solution of a Laplace equation which appears when computing the proximal operator associated with the constraint $\mathrm{div}\,\sigma = 0$. This however can be avoided by exploiting the Helmholtz decomposition of vector fields, as recently showed in [18], or adding regularization terms as in [24]. Recently, Carrillo and collaborators [11] proposed a finite difference scheme similar to that in [27] (in the context of the discretization of Wasserstein gradient flows), for which they could also prove its convergence with mesh refinement, but only upon strong regularity assumptions on the solutions of the continuous problem.

In [23] a numerical scheme was proposed using tools from finite element and finite volume methods, where one explicitly constructs a duality structure for the discrete variables. Later Lavenant [22] proved convergence of this scheme, unconditionally with respect to the time/space step size, to the solutions of the optimal transport problem, proposing a general framework for convergence of discretizations of problem (1.1) between two arbitrary probability measures. This filled a critical gap for the analysis of discrete dynamical transport models, since previously convergence results were only known in case of sufficiently smooth solutions (as in [11]) or conditional to the relative time/space step sizes (e.g., in the context of finite volume methods, combining the results in [14] and [16]).

1.1. **Contributions and structure of the paper.** In this paper we propose a mixed finite element discretization of (1.1) which generalizes to the finite element setting the scheme proposed in [27]. We derive our method by discretizing a saddle point formulation of the dynamic optimal transport problem on Hilbert spaces, where one looks for a solution $(q, \sigma) \in L^2([0,1] \times D; \mathbb{R}^{d+1})^2$. Nonetheless, we stress that the method we obtain is still well-defined when the initial and final data are arbitrary probability measures. By using $H(\mathrm{div})$-conforming spaces for the variable σ, we are able to construct discrete solutions that satisfy exactly the weak form of the continuity equation (1.2).

We solve the discrete problem using a proximal splitting algorithm [28]. Importantly, this is not only a discretization of the same algorithm applied to the continuous saddle point formulation as in previous works, but also a genuine optimization scheme applied to the finite dimensional problem.

Finally, using the framework of [22], we also show that our discrete solutions converge towards the solutions of the optimal transport problem between two arbitrary measures, and therefore even when the solution σ is only a measure (see theorem 5.4). Such a result carries over also to a slight modification of the finite difference scheme proposed in [27], which can be viewed as a particular instance of our discretization on a regular quadrilateral grid (see remark 5.6).

The paper is structured as follows. We establish the notation in section 2. In section 3 we give the precise formulation of problem (1.1) and describe the proximal splitting algorithm applied to the continuous problem in the Hilbert space setting. In section 4 we introduce and discuss the main finite element tools we use for our method. In section 5 we define our finite element discretization of problem (1.1) and state the convergence result. In section 6 we detail the steps required for solving our discrete optimal transport problem with a proximal splitting algorithm. In section 7 we describe how to introduce regularization terms in the formulation. Finally in section 8 we present some numerical results.

2. Notation

Throughout the paper we will denote by $D \subset \mathbb{R}^d$ a convex polytope and by $\Omega := [0,1] \times D$ the space-time domain. For differential operator such as ∇ or div, we use the subscript x to highlight that these are defined on D rather than Ω, but we will drop this subscript when this is clear from the context.

We use the standard notation for Sobolev spaces on D or Ω. In particular, $L^p(D; \mathbb{R}^d)$ denotes the space of functions $f : D \to \mathbb{R}^d$ whose Euclidean norm $|f|$ is in $L^p(D)$. We use a similar notation for functions taking values on a subset $K \subset \mathbb{R}^d$, or defined on Ω. We denote by $H(\mathrm{div}; D)$ the space of vector fields $f : D \to \mathbb{R}^d$ in $L^2(D; \mathbb{R}^d)$ whose divergence is in $L^2(D)$. Similarly, $H(\mathrm{div}; \Omega)$ the space of vector fields $f : \Omega \to \mathbb{R}^{d+1}$ in $L^2(\Omega; \mathbb{R}^{d+1})$ whose divergence is in $L^2(\Omega)$.

Finally, we denote by $\mathcal{M}(D)$ the set of finite signed measures on D, by $\mathcal{M}_+(D) \subset \mathcal{M}(D)$ the convex subset of positive measures; by $\mathcal{P}(D) \subset \mathcal{M}_+(D)$ the set of positive measures of total mass equal to one; and by $C(D)$ the space of continuous functions on D. We use a similar notation for the spaces of measures and continuous functions on Ω. We use $\langle \cdot, \cdot \rangle$ to denote either the duality pairing between measures and continuous functions or the L^2 inner product, on either D or Ω, according to the context.

3. Dynamical formulation of optimal transport

The dynamical optimal transport problem (1.1) can be formulated as a saddle point problem on the space of measures $\sigma := (\rho, m) \in \mathcal{M}(\Omega) \times \mathcal{M}(\Omega)^d$. This can

be written as follows

$$\inf_{\sigma\in\mathcal{C}} \mathcal{A}(\sigma), \quad \mathcal{A}(\sigma) := \sup_{q\in C(\Omega;K)} \langle q,\sigma\rangle, \tag{3.1}$$

where $\mathcal{C}$ is the set of measures $\sigma \in \mathcal{M}(\Omega)^{d+1}$ satisfying $\operatorname{div}\sigma = 0$ in distributional sense with boundary conditions

$$\sigma\cdot n_{\partial\Omega} = \mathcal{X}\,, \quad \mathcal{X} := \begin{cases} \rho_0 & \text{on } \{0\}\times D,\\ \rho_1 & \text{on } \{1\}\times D,\\ 0 & \text{otherwise,}\end{cases} \tag{3.2}$$

with $\rho_0,\rho_1 \in \mathcal{P}(D)$, and where $C(\Omega;K)$ is the space of continuous functions on Ω taking value in the convex set

$$K := \left\{(a,b)\in\mathbb{R}\times\mathbb{R}^d\,;\, a+\frac{|b|^2}{2}\le 0\right\}. \tag{3.3}$$

It will be convenient to treat time and space as separate variables. In particular we will also use the action defined by

$$A(\rho,m) := \sup_{(a,b)\in C(D;K)} \langle\rho,a\rangle + \langle m,b\rangle\,, \tag{3.4}$$

for any $(\rho,m)\in\mathcal{M}(D)^{d+1}$. Then, $A(\rho,m)$ is finite if and only if m has a density with respect to ρ and in that case $A(\rho,m) = \int_D B(\rho,m)$, where $B:\mathbb{R}\times\mathbb{R}^d\to[0,+\infty]$ is the function given by

$$B(a,b) := \begin{cases} \frac{|b|^2}{2a} & \text{if } a>0,\\ 0 & \text{if } a=0, b=0,\\ +\infty & \text{if } a=0, b\neq 0 \text{ or } a<0\,.\end{cases} \tag{3.5}$$

Due to the definition of the function (3.5), a saddle point of problem (3.1) must satisfy $\rho\ge 0$.

The value of the infimum of problem (3.1) coincides with $W_2^2(\rho_0,\rho_1)/2$, where $W_2(\cdot,\cdot)$ denotes the Wasserstein distance associated with the L^2 cost (see theorem 5.28 in [30]). Moreover the infimum itself is attained by a measure $\sigma=(\rho,m)$, where ρ is known as the Wasserstein geodesic between ρ_0 and ρ_1 (see proposition 5.32 in [30]). We refer the reader to [30] for more details on the links between the dynamical formulation and the Wasserstein distance.

3.1. **Hilbert space setting and proximal splitting.** Before discussing the discretization of problem (3.1), we review its reformulation on Hilbert spaces, and discuss the convergence of the proximal splitting algorithm.

Proposition 3.1 (Guittet [17]; Hug et al. [20])**.** *Suppose $\rho_0,\rho_1\in L^2(D)$. Then problem (3.1) is equivalent to*

$$\inf_{\sigma\in\mathcal{C}} \sup_{q\in L^2(\Omega;K)} \langle q,\sigma\rangle \tag{3.6}$$

where $\mathcal{C}$ is the set of functions $\sigma\in H(\operatorname{div};\Omega)$ satisfying $\operatorname{div}\sigma=0$ in weak sense with boundary conditions given by (3.2). Moreover, assuming that $\operatorname{supp}(\rho_0)\cup\operatorname{supp}(\rho_1)\subset \mathring{D}$, there exists a saddle point $(\sigma^*,q^*)\in\mathcal{C}\times L^2(\Omega;K)$ solving problem (3.6).

The equivalence of problem (3.6) to (3.1) can be easily deduced by a regularization argument on σ and then applying Lusin's theorem as in Proposition 5.18 in [30]. The proof for the existence of a saddle point problem is more delicate and can be found in [20].

In order to apply a proximal splitting algorithm to solve problem (3.6), we first write it in the form

$$\inf_{\sigma\in L^2(\Omega;\mathbb{R}^{d+1})} \sup_{q\in L^2(\Omega;\mathbb{R}^{d+1})} \langle q,\sigma\rangle + \iota_{\mathcal{C}}(\sigma) - \iota_{\mathcal{K}}(q) \tag{3.7}$$

where ι denotes the convex indicator function and

$$\mathcal{K} := L^2(\Omega;K) = \{q \in L^2(\Omega;\mathbb{R}^{d+1})\,;\, q \in K \text{ a.e.}\}. \tag{3.8}$$

Note in particular that $\mathcal{C}$ and $\mathcal{K}$ are closed convex sets of L^2.

We apply to (3.7) the primal-dual projection algorithm proposed in [28]. In particular, given $\tau_1, \tau_2 > 0$ and an admissible $(\sigma^0, q^0) \in \mathcal{C} \times \mathcal{K}$, we define the sequence $\{(\sigma^k, q^k)\}_k$ by the two-step algorithm:

$$\textbf{Step 1}: \quad \sigma^{k+1} = P_{\mathcal{C}}(\sigma^k - \tau_1 q^k) \tag{3.9a}$$

$$\textbf{Step 2}: \quad q^{k+1} = P_{\mathcal{K}}(q^k + \tau_2(2\sigma^{k+1} - \sigma^k)) \tag{3.9b}$$

where $P_{\mathcal{C}}$ and $P_{\mathcal{K}}$ are the L^2 projections on the closed convex sets $\mathcal{C}$ and $\mathcal{K}$, respectively. The projection onto $\mathcal{C}$ amounts to computing the Helmholtz decomposition of $\sigma^k - \tau_1 q^k$ and selecting the divergence-free part, whereas the projection onto $\mathcal{K}$ is a pointwise projection applied to a representative of $q^k + \tau_2(2\sigma^{k+1} - \sigma^k)$.

The proof of convergence in [28] holds also in our setting. More precisely, the following convergence theorem holds.

Theorem 3.2 (Pock et al. [28])**.** *If $\tau_1\tau_2 < 1$ then $(\sigma^k, q^k) \to (\sigma^*, q^*) \in \mathcal{C}\times\mathcal{K}$ which solves (3.6).*

Discretizing problem (3.7), and consequently the proximal splitting algorithm (3.9), with finite elements requires choosing finite-dimensional spaces for σ and q so that the steps in (3.9) are well-posed and solvable. However, satisfying these requirements is not enough to guarantee convergence of the discrete solutions to the ones of the infinite dimensional problem. Hereafter we will identify a class of finite element spaces for which the theory developed in [22] applies, which allows us to deduce convergence to the solutions of problem (3.1), i.e. even when ρ_0 and ρ_1 are arbitrary probability measures and the Hilbert space setting presented in this section is not well-defined.

4. Mixed finite element setting

4.1. **Finite element spaces on D.** We recall that D is a convex polytope in $\mathbb{R}^d$. We consider a triangulation of D which we denote $\mathcal{T}_h$, i.e. a decomposition of D in either simplicial or quadrilateral elements, where h is the maximum diameter of the elements in $\mathcal{T}_h$. We assume that there exists a constant C_{mesh} such that

$$|h|^d \le C_{mesh}|T|\,, \quad \forall\, T \in \mathcal{T}_h\,. \tag{4.1}$$

This implies that the mesh is quasiuniform, meaning that the ratio of any two element diameters is uniformly bounded by a constant depending only on C_{mesh}, and shape-regular, that is, for each element $T \in \mathcal{T}_h$, the ratio of its diameter and the diameter of the largest inscribed ball is uniformly bounded by a constant depending only on C_{mesh} (see, e.g., [2]).

For any $T \in \mathcal{T}_h$, we denote by $\mathcal{P}_k(T)$ the space of polynomials of degree up to k on T. If T is a quadrilateral element, i.e., in general, if T is obtained by an affine transformation $\phi : I^d \to T$ where I is the unit interval, then we define $\mathcal{P}_{k_1,\dots k_d}(I^d) := \mathcal{P}_{k_1}(I) \otimes \dots \otimes \mathcal{P}_{k_d}(I)$ and $\mathcal{P}_{k_1,\dots k_d}(T) := \mathcal{P}_{k_1,\dots k_d}(I^d) \circ \phi^{-1}$.

We now define the finite elment spaces Q_h and V_h which will serve to construct approximations of the denisty ρ and the momentum m, respectively. We set

$$Q_h := \{\phi \in L^2(D)\,;\, \phi|_T \in \mathcal{P}_0(T),\ \forall\, T \in \mathcal{T}_h\}, \tag{4.2}$$

$$V_h := \{v \in H(\mathrm{div}; D)\,;\, v|_T \in V_h(T),\ \forall\, T \in \mathcal{T}_h\}. \tag{4.3}$$

where $V_h(T)$ is the so-called shape function space. We distinguish two cases:

(1) for simplicial elements (triangles or tetrahedrons), we take $V_h(T)$ to be either

$$\mathcal{RT}_0(T) := \{v = v_0 + v_1\hat{x}\,;\, v_0 \in (\mathcal{P}_0(T))^d\,, v_1 \in \mathcal{P}_0(T)\} \subset (\mathcal{P}_1(T))^d,$$

where $\hat{x} = (x_1, \dots, x_d) \in (\mathcal{P}_1(T))^d$, which generates the lowest order Raviart-Thomas space; or $\mathcal{BDM}_1(T) = (\mathcal{P}_1(T))^d$, which generates the lowest order Brezzi-Douglas-Marini $H(\mathrm{div})$-conforming space;

(2) for quadrilateral elements, we set $T = \phi(I^d)$, where I is an interval and ϕ an affine transformation, and we take $V_h(T)$ to be the tensor product space which generates the lowest order Raviart-Thomas space on quadrilateral elements. This is defined as follows:

$$\mathcal{RT}_{[0]}(T) := \begin{cases} \mathcal{P}_{1,0}(T)e_1 + \mathcal{P}_{0,1}(T)e_2 & \text{if } d = 2\,, \\ \mathcal{P}_{1,0,0}(T)e_1 + \mathcal{P}_{0,1,0}(T)e_2 + \mathcal{P}_{0,0,1}(T)e_3 & \text{if } d = 3\,, \end{cases}$$

where $\{e_i\}_i$ is the basis for $\mathbb{R}^d$ aligned with the edges of T.

In practice, the space V_h is chosen among the standard lowest order $H(\mathrm{div})$-conforming spaces. In fact, the property of being piece-wise linear will be crucial in the following, namely to prove the convergence result (5.4) (see, in particular, proposition A.2 in the appendix). A graphical representation of the degrees of freedom associated with these spaces is shown in figure 1.

Importantly, with the choices mentioned above, there exist bounded projections $\Pi_{Q_h} : L^2(D) \to Q_h$ and $\Pi_{V_h} : H(\mathrm{div}; D) \to V_h$ that commute with the divergence operator [2, 10]. In other words, the following diagram commutes

$$\begin{array}{ccc} H(\mathrm{div}; D) & \xrightarrow{\ \mathrm{div}\ } & L^2(D) \\ \Big\downarrow{\scriptstyle \Pi_{V_h}} & & \Big\downarrow{\scriptstyle \Pi_{Q_h}} \\ V_h & \xrightarrow{\ \mathrm{div}\ } & Q_h \end{array} \tag{4.4}$$

As a consequence, the divergence operator is surjective onto Q_h when restricted on V_h, i.e. $\mathrm{div}\, V_h = Q_h$. Finally, we let $Q_h^+ \subset Q_h$ the convex subset of non-negative piecewise constant functions.

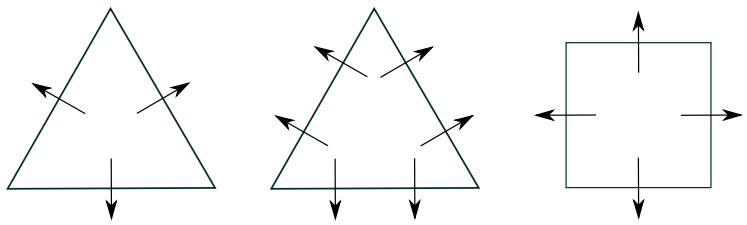
Remark 4.1. *The precise definition of the projection operator can be found in section 5.4 in* [2] *or in* [10]*. Here, we will only need the explicit definition of* Π_{Q_h}*, which is given by*

$$\Pi_{Q_h}\rho|_T = \frac{1}{|T|}\int_T \rho\,, \tag{4.5}$$

for any $T \in \mathcal{T}_h$*. Therefore,* Π_{Q_h} *is well-defined on* $\mathcal{M}(D)$ *and its resctriction on* $\mathcal{M}_+(D)$ *is surjective onto* Q_h^+*.*

4.2. **Finite element spaces on Ω.** We now introduce finite element spaces on the space-time domain $[0,1] \times D$. We first define a decomposition $\mathcal{T}_{h,\tau}$, obtained by a tensor product construction. In other words, we assume that $\mathcal{T}_{h,\tau}$ is obtained by tensor product of a triangulation $\mathcal{T}_h$ of D and a decocomposition of $[0,1]$ of maximum size τ, so that any element $S \in \mathcal{T}_{h,\tau}$ is of the form $S = [t_0, t_1] \otimes T$ where $T \in \mathcal{T}_h$.

We now define the finite element spaces $F_{h,\tau}$ and $Z_{h,\tau}$ on the space-time domain. The space $Z_{h,\tau}$ will be constructed using the standard tensor product construction based on the spaces Q_h and V_h defined on D, and continuous $\mathcal{P}_1$ and discontinuous



(A) $\mathcal{RT}_0$ (B) $\mathcal{BDM}_1$ (C) $\mathcal{RT}_{[0]}$

FIGURE 1. Degrees of freedom for different choices of shape function space $V_h(T)$

$\mathcal{P}_0$ spaces on $[0,1]$. In our discretization, the space-time vector field (ρ, m) will be an element of $Z_{h,\tau}$ whereas $F_{h,\tau}$ will be the space of discrete Lagrange multipliers associated to the continuity equation, which is just the constraint that the space-time divergence of (ρ, m) is zero.

More precisely, we define

$$F_{h,\tau} \coloneqq \{\phi \in L^2(\Omega)\,;\, \phi|_S \in \mathcal{P}_0(S),\ \forall\, S \in \mathcal{T}_{h,\tau}\}, \tag{4.6}$$

$$Z_{h,\tau} \coloneqq \{v \in H(\mathrm{div};\Omega)\,;\, v|_S \in Z_{h,\tau}(S),\ \forall\, S \in \mathcal{T}_{h,\tau}\}. \tag{4.7}$$

For $S = [t_0, t_1] \otimes T$, the shape function space $Z_{h,\tau}(S)$ is built by defining a shape function space for the density, in the space-time domain, which is given by

$$Q_{h,\tau}(S) \coloneqq \mathcal{P}_1([t_0, t_1]) \otimes Q_h(T)\,, \tag{4.8}$$

(i.e. the density is piecewise linear in time), and a shape function space for the momentum, in the space-time domain, which is given by

$$V_{h,\tau}(S) \coloneqq \mathcal{P}_0([t_0, t_1]) \otimes V_h(T) \tag{4.9}$$

(i.e. the momentum is piecewise constant in time). Then, we set

$$Z_{h,\tau}(S) \coloneqq V_{h,\tau}(S) \oplus (Q_{h,\tau}(S)\,\hat{t}) \tag{4.10}$$

where $\hat{t}$ is the unit vector oriented in the time direction. The spaces $F_{h,\tau}$ and $Z_{h,\tau}$ inherit from Q_h and V_h the commuting diagram property mentioned above. In particular, there exist bounded projections $\Pi_{F_{h,\tau}} : L^2(\Omega) \to F_{h,\tau}$ and $\Pi_{Z_{h,\tau}} : H(\mathrm{div};\Omega) \to Z_{h,\tau}$ for which the following diagram commutes

$$\begin{array}{ccc} H(\mathrm{div};\Omega) & \xrightarrow{\ \mathrm{div}\ } & L^2(\Omega) \\ \Big\downarrow{\scriptstyle \Pi_{Z_{h,\tau}}} & & \Big\downarrow{\scriptstyle \Pi_{F_{h,\tau}}} \\ Z_{h,\tau} & \xrightarrow{\ \mathrm{div}\ } & F_{h,\tau} \end{array} \tag{4.11}$$

where the divergence is the one associated with the space-time domain Ω. Then, as before, the divergence operator is surjective onto $F_{h,\tau}$ when restricted on $Z_{h,\tau}$, i.e. $\mathrm{div}\, Z_{h,\tau} = F_{h,\tau}$. Note that the precise definition for the projection operators on tensor product meshes can be found in [1].

4.3. **Discrete projection on the divergence-free subspace.** Denote by $\mathcal{B}$ the kernel of the divergence operator on $H(\mathrm{div};\Omega)$. Given $\xi \in L^2(\Omega)$ we define the projection $P_{\mathcal{B}}(\xi)$ to be the divergence-free vector field σ minimizing the L^2 distance

from ξ. This can be obtained solving the following problem for $(\sigma, \phi) \in H(\mathrm{div};\Omega) \times L^2(\Omega)$

$$\begin{cases} \langle \sigma, v\rangle + \langle \phi, \mathrm{div}\, v\rangle = \langle \xi, v\rangle\,, & \forall v \in H(\mathrm{div};\Omega) \\ \langle \mathrm{div}\,\sigma, \psi\rangle = 0\,, & \forall \psi \in L^2(\Omega) \end{cases} \tag{4.12}$$

The system is well-posed thanks to the surjectivity of the divergence operator onto $L^2(\Omega)$. Let $\mathcal{B}_{h,\tau}$ be the kernel of the divergence operator restricted on $Z_{h,\tau}$. We define the projection $P_{\mathcal{B}_{h,\tau}}(\xi)$ to be the divergence-free vector field $\sigma_{h,\tau} \in Z_{h,\tau}$ minimizing the L^2 distance from ξ. This can be obtained solving the following problem for $(\sigma_{h,\tau}, \phi_{h,\tau}) \in Z_{h,\tau} \times F_{h,\tau}$

$$\begin{cases} \langle \sigma_{h,\tau}, v\rangle + \langle \phi_{h,\tau}, \mathrm{div}\, v\rangle = \langle \xi, v_h\rangle\,, & \forall v \in Z_{h,\tau} \\ \langle \mathrm{div}\,\sigma_{h,\tau}, \psi\rangle = 0\,, & \forall \psi \in F_{h,\tau} \end{cases} \tag{4.13}$$

The commuting diagram (4.11) implies well-posedness of the discrete system. In particular, it implies the following inf-sup condition: there exists a constant $\beta > 0$ independent of h and τ such that

$$\inf_{\phi \in F_{h,\tau}} \sup_{\sigma \in Z_{h,\tau}} \frac{\langle \phi, \mathrm{div}\,\sigma\rangle}{\|\sigma\|_{H(\mathrm{div})}\|\phi\|_{L^2}} > \beta\,, \tag{4.14}$$

see for example proposition 5.4.2 in [10]. Then, problem (4.13) is well-posed, i.e. it has a unique solution $(\sigma_{h,\tau}, \phi_{h,\tau})$ which verifies $\sigma_{h,\tau} \in \mathcal{B}$ and

$$\|\sigma_{h,\tau}\|_{L^2} \le C_1 \|\xi\|_{L^2}\,, \tag{4.15}$$

$$\|\phi_{h,\tau}\|_{L^2} \le C_2 \|\xi\|_{L^2}\,, \tag{4.16}$$

$$\|\sigma_{h,\tau} - \sigma\|_{L^2} + \|\phi_{h,\tau} - \phi\|_{L^2} \le C_3 \|\xi_{h,\tau} - \xi\|_{L^2}\,, \tag{4.17}$$

where $C_1, C_2, C_3 > 0$ are constants independent of h and τ, $\xi_{h,\tau}$ is the L^2 projection of ξ onto $Z_{h,\tau}$ and (σ, ϕ) is the unique solution of problem (4.12) (e.g., these results can be derived as particular cases of theorems 4.3.2, 5.2.1 and 5.2.5 in [10]).

In the following we will need to compute the discrete version of the projection on $\mathcal{C}$. In particular we define

$$\mathcal{C}_{h,\tau} := \{\sigma \in \mathcal{B}_{h,\tau}\,,\ \sigma \cdot n_{\partial\Omega} = \mathcal{X}_{h,\tau}\} \tag{4.18}$$

where, since Π_{Q_h} can be defined on $\mathcal{M}(D)$ (see equation (4.5)), we set

$$\mathcal{X}_{h,\tau} := \begin{cases} \Pi_{Q_h}\rho_0 & \text{on } \{0\} \times D, \\ \Pi_{Q_h}\rho_1 & \text{on } \{1\} \times D, \\ 0 & \text{otherwise.} \end{cases} \tag{4.19}$$

The well-posedness arguments above for the projection onto $\mathcal{B}$ and $\mathcal{B}_{h,\tau}$ hold also for the projections onto $\mathcal{C}$ and $\mathcal{C}_{h,\tau}$ up to adding boundary conditions to the spaces $H(\mathrm{div};\Omega)$ and $Z_{h,\tau}$, and replacing $L^2(\Omega)$ and $F_{h,\tau}$ by $L^2(\Omega)/\mathbb{R}$ and $F_{h,\tau}/\mathbb{R}$, respectively.

5. Discrete dynamical formulation and convergence

In this section we formulate the discrete problem and state a convergence result obtained by applying the theory developed in [22]. For this, we need to introduce a space for the discrete dual variable q. We adopt the same notation as for the spaces defined in section 4. In particular, we set for $r \in \{0, 1\}$,

$$X_h^r := \{\phi \in L^2(D)\,;\ \phi|_T \in X_h^r(T),\ \forall\, T \in \mathcal{T}_h\}. \tag{5.1}$$

The superscript r denotes the polynomial order of the shape function space $X_h^r(T)$. We distinguish two cases:

(1) for simplicial elements (triangles or tetrahedrons), we take $X_h^r(T) := \mathcal{P}_r(T)$.

(2) for quadrilateral elements, we set $T = \phi(I^d)$, where I is an interval and ϕ an affine transformation, and we take $X_h^r(T) := \mathcal{P}_r(I)^d$.

The associated space-time space is defined by

$$X_{h,\tau}^r := \{\phi \in L^2(\Omega)\,;\, \phi|_S \in X_{h,\tau}^r(S),\ \forall\, S \in \mathcal{T}_{h,\tau}\}. \tag{5.2}$$

with $X_{h,\tau}^r(S) = \mathcal{P}_0([t_0, t_1]) \otimes X_h^r(T)$. In order to simplify the notation, we will omit the superscript r when not relevant to the discussion.

Remark 5.1. *The choice $r \in \{0, 1\}$ is dictated by computational feasibility of the algorithm. In fact, for these spaces, we can compute explicitly the projection on $\mathcal{K} \cap X_{h,\tau}$ as it will be explained in the next section. On the other hand, we restrict ourselves to piecewise constant functions in time since this is crucial for the convergence of the algorithm, as shown in* [22].

The discrete action (at fixed time) is defined as follows:

$$A_h(\rho, m) := \sup_{(a,b)\in(X_h)^{d+1}} \{\langle \rho, a\rangle + \langle m, b\rangle\,;\, (a, b) \in K\ a.e.\} \tag{5.3}$$

for any $(\rho, m) \in Q_h \times V_h$. By construction, $A_h : Q_h \times V_h \to [0, +\infty]$ is a proper convex function -1-positively homogeneous in its first variable and 2-positively homogeneous in its second variable. Moreover, it is non-increasing in its first argument, i.e. $A_h(\rho_1 + \rho_2, m) \le A_h(\rho_1, m)$ for any $\rho_1, \rho_2 \in Q_h^+$ and $m \in V_h$. In fact, suppose that $A_h(\rho_1 + \rho_2, m) < +\infty$. Then there exists $(a^*, b^*) \in K$ such that $\langle \rho_1 + \rho_2, a^*\rangle + \langle m, b^*\rangle = A_h(\rho_1 + \rho_2, m)$; in particular $a^* \le 0$. Then

$$A_h(\rho_1, m) \ge A_h(\rho_1 + \rho_2, m) - \langle \rho_2, a^*\rangle \ge A_h(\rho_1 + \rho_2, m)\,, \tag{5.4}$$

and by a similar reasoning we obtain that if $A(\rho_1 + \rho_2, m) = +\infty$ then we also have $A(\rho_1, m) = +\infty$.

The space-time discretization of problem (3.1) is given by

$$\inf_{\sigma\in\mathcal{C}_{h,\tau}} \mathcal{A}_{h,\tau}(\sigma), \quad \mathcal{A}_{h,\tau}(\sigma) := \sup_{q\in(X_{h,\tau}^r)^{d+1}\cap\mathcal{K}} \langle q, \sigma\rangle\,. \tag{5.5}$$

Note that, by definition, $\mathcal{A}_{h,\tau}$ is convex and non-negative. Therefore, problem (5.5) always admits minimizers.

Suppose that the time discretization is given by a decomposition of the interval $[0, 1]$ in N elements, i.e. fixing the points $0 = t_0 < t_1 < \ldots < t_{N+1} = 1$. Given $\sigma = (\rho, m) \in Z_{h,\tau}$, we can identify the density ρ with the collection $\{\rho_i\}_{i=0}^{N+1}$ with $\rho_i \in Q_h$, and the momentum m with the collection $\{m_i\}_{i=1}^{N+1}$ with $m_i \in V_h$. Since q is piecewise constant in time, we have the following equivalent formulation

$$\mathcal{A}_{h,\tau}(\sigma) = \sum_{i=1}^{N+1} A_h\left(\frac{\rho_i + \rho_{i-1}}{2}, m_i\right) |t_i - t_{i-1}|\,. \tag{5.6}$$

Note that in order to obtain (5.6) from (5.5), we relied on the particular choice of finite element spaces for density (piecewise linear in time), momentum (piecewise constant in time) and the corresponding dual variables (piecewise constant in time).

Remark 5.2 (Continuity constraint)**.** *The choice of a $H(\mathrm{div})$-conforming finite element space for σ implies that the weak form of the continuity equation $\dot\rho + \mathrm{div}_x\, m = 0$ is satisfied exactly by any solution of the discrete saddle point problem* (5.5) *(this is also directly implied by the definition of the constraint set $\mathcal{C}_{h,\tau}$ in* (4.18)*)* .

Remark 5.3 (Positivity constraint)**.** *A disadvantage of evaluating the action on the mean density (in time) is that the positivity constraint $\rho \ge 0$ is then only enforced on $\frac{\rho_i+\rho_{i-1}}{2}$, rather than on each ρ_i separately. This fact is related to the appearance of oscillations in the discrete solutions, as shown in the numerical tests in section 8 (see also remark 5.5).*

The objects introduced until now define a finite dimensional model of optimal transport in the sense of Definition 2.5 in [22]. The framework developed therein can be used to deduce a convergence result for our scheme.

Theorem 5.4. *Let $\rho_0, \rho_1 \in \mathcal{P}(D)$ be given and $\{\mathcal{T}_{h,\tau}\}_{h,\tau\geq 0}$ a family of tensor-product decomposition of Ω such that the time discretization is uniform, i.e. $t_i - t_{i-1} = \tau$ for all $i = 1, \dots, N+1$, and the space discretization $\mathcal{T}_h$ satisfies equation (4.1). Let $\sigma_{h,\tau}$ be a minimizer of problem (5.5) associated with $\mathcal{T}_{h,\tau}$ and for $r = 1$. Then, as $h, \tau \to 0$, up to extraction of a subsequence, $\sigma_{h,\tau}$ converges weakly to $\sigma \in \mathcal{M}(\Omega)^{d+1}$ a minimizer of problem (3.1).*

The proof is essentially an extension of the one presented in [22] and is postponed to the appendix.

Remark 5.5 (Stability)**.** *The existence of bounded projections sastifying the commuting diagram (4.11) ensures stability of the projection onto $\mathcal{C}_{h,\tau}$ (see (4.15)). Such commuting projections are also crucial to estabilish the convergence result in theorem 5.4: in [22], they are used to sample the continuous solution into a discrete one satisfying the continuity equation, therefore providing an admissible candidate in the discrete problem. Nonetheless, due to the nonlinear constraint $q \in \mathcal{K}$, one cannot apply the standard linear theory in [10], for example, so the commuting diagram condition does not imply directly a stability result analogous to (4.15) for the saddle point problem (5.5) (even if we see it as a discretization of the Hilbert space formulation in proposition 3.1). The numerical experiments in section 8 suggest that the finite element pairs considered here $(Z_{h,\tau}, X^r_{h,\tau})$ are stable when $r = 1$, but osillations may occur for $r = 0$ when $V_h = \mathcal{BDM}_1$, for example, hinting at the instablity of the discretization for this case.*

Remark 5.6. *Suppose that $D = [0,1]^d$ and that $\mathcal{T}_{h,\tau}$ is a uniform quadrilateral discretization of $\Omega = [0,1]^{d+1}$. Then for $r = 0$, the discrete problem (5.5) coincides with the discretization proposed in [27]. Theorem 5.4 shows that modifying this method with $r = 1$, one can prove convergence to the solution of the continuous problem (3.1).*

6. The proximal splitting algorithm

We now describe in detail the discrete version of the proximal splitting algorithm introduced in section 3.1. Note that this is a proximal splitting algorithm applied to the discrete optimization problem rather than the discretization of the optimality conditions of the algorithm in the continuous setting.

First of all, we define

$$\mathcal{K}^r_{h,\tau} := \mathcal{K} \cap (X^r_{h,\tau})^{d+1} := \{q \in (X^r_{h,\tau})^{d+1}\,;\, q \in K \text{ a.e.}\} \tag{6.1}$$

We write the discrete problem as follows:

$$\inf_{\sigma \in L^2(\Omega;\mathbb{R}^{d+1})} \sup_{q \in L^2(\Omega;\mathbb{R}^{d+1})} \langle q, \sigma\rangle + \iota_{\mathcal{C}_{h,\tau}}(\sigma) - \iota_{\mathcal{K}_{h,\tau}}(q) \tag{6.2}$$

where $\mathcal{C}_{h,\tau}$ is defined in (4.18). Then, the proximal splitting algorithm of section 3.1 applied to problem (6.2) can be formulated as follows: given $\tau_1, \tau_2 > 0$ and an admissible $(\sigma^0, q^0) \in \mathcal{C}_{h,\tau} \times \mathcal{K}_{h,\tau}$, we define the sequence $\{(\sigma^k, q^k)\}_k$ by performing iteratively the following two steps:

$$\textbf{Step 1}: \quad \sigma^{k+1} = P_{\mathcal{C}_{h,\tau}}(\sigma^k - \tau_1 q^k) \tag{6.3a}$$

$$\textbf{Step 2}: \quad q^{k+1} = P_{\mathcal{K}_{h,\tau}}(q^k + \tau_2(2\sigma^{k+1} - \sigma^k)) \tag{6.3b}$$

The convergence result in theorem 3.2 clearly holds also in the discrete setting and gives convergence of the algorithm to a discrete saddle point $(\sigma_{h,\tau}, q_{h,\tau})$, if the condition $\tau_1\tau_2 < 1$ is satisfied. The two steps in the algorithm can be computed as follows.

Step 1. As discussed in section 4.3, the projection $P_{\mathcal{C}_{h,\tau}}$ can be computed modifying the system given by (4.13) by adding the Neumann boundary conditions associated with the function (4.19).

Step 2. Since $P_{\mathcal{K}_{h,\tau}}$ is an L^2 projection, we have that $P_{\mathcal{K}_{h,\tau}} = P_{\mathcal{K}_{h,\tau}} \circ P_{(X^r_{h,\tau})^{d+1}}$, where $P_{(X^r_{h,\tau})^{d+1}}$ denotes the L^2 projection onto $(X^r_{h,\tau})^{d+1}$. This means that we only need to be able to compute $P_{\mathcal{K}_{h,\tau}}$ when applied to an element of $X^r_{h,\tau}$. In addition, since $X^r_{h,\tau}$ is discontinuous across elements, we can compute the projection element by element, and since functions in $X^r_{h,\tau}(S)$ are constant along the time direction, we can also eliminate the time variable in the projection. More precisely, let $\xi \in (X^r_{h,\tau}(S))^{d+1}$ and $S = [t_0, t_1] \times T$, we can identify ξ with an element of $(X^r_h(T))^{d+1}$. Then,

$$\xi_{\mathcal{K}} := P_{\mathcal{K}_{h,\tau}}|_S\, \xi = \operatorname{argmin}\{\|\xi - q\|^2_{L^2(T)}\,;\, q \in (X^r_h(T))^{d+1} \cap K\}\,. \tag{6.4}$$

We distinguish two cases:

(1) if $r = 0$, the projection (6.4) is just the projection of a vector $\xi \in \mathbb{R}^{d+1}$ onto the convex set K;
(2) if $r = 1$, $\xi_{\mathcal{K}}$ is fully determined by its value on the vertices $\{v_i\}_i$ of T, and by linearity of the shape functions and convexity of the constraint, we obtain

$$\xi_{\mathcal{K}}(v_i) = \operatorname{argmin}\{|\xi(v_i) - q|^2\,;\, q \in K\}\,. \tag{6.5}$$

In both cases we only need to compute for each degree of freedom the projection of a given vector $(\bar{a}, \bar{b}) \in \mathbb{R} \times \mathbb{R}^d$ onto K. If $(\bar{a}, \bar{b}) \notin K$, such a projection is given explicitly by the vector

$$\left(-\frac{\mu^2}{2}, \mu\,\frac{\bar{b}}{|\bar{b}|}\right)$$

where $\mu \geq 0$ is the largest real root of the third order polynomial

$$x \mapsto \frac{x^3}{2} + x(\bar{a} + 1) - |\bar{b}|\,.$$

Remark 6.1. *As for the finite difference discretization studied in* [27]*, different optimization techniques could be applied to solve problem* (5.5)*. In particular, it should be noted that the ADMM approach orginally proposed by Benamou and Brenier* [5] *could also be applied. This would lead to a very similar algorithm to* (6.3)*, but it would require the introduction of an additional variable which avoids coupling of the degrees of freedom in the optimization step with respect to q. In other words, this is needed in order to be able to perform the projection on K for each degree of freedom separately. More details on this issue can be found in* [27] *for the discretization studied therein, and they hold also in the finite element setting.*

7. Regularization

The optimal transport problem does not involve any regularizing effect on the interpolation between two measures. In fact, one can even expect a loss of regularity in some cases, namely if one is interpolating between two smooth densities on a smooth but non-convex domain. Such a a loss of regularity (which is often unphysical when the density represents a physical quantity) can be avoided introducing

additional regularization terms in the formulation. In this section we describe how to do so, and how these modifications translate at the algorithmic level.

We consider the Hilbert space setting discribed in section 3.1 and we study problems in the form

$$\inf_{\sigma\in\mathcal{C}} \mathcal{A}(\sigma) + \alpha\mathcal{R}(\sigma) \tag{7.1}$$

where $\mathcal{R} : L^2(\Omega) \to \mathbb{R}$ is a convex, proper and l.s.c. functional, and $\alpha > 0$. For this type of problem, we can still apply the proximal splitting algorithm (3.9) replacing the projection onto $\mathcal{C}$ by $\mathrm{prox}_{\tau_1\mathcal{F}}$, the proximal operator of $\mathcal{F} := \iota_{\mathcal{C}} + \alpha\mathcal{R}$, defined by

$$\mathrm{prox}_{\tau_1\mathcal{F}}(\xi) = \operatorname*{argmin}_{\eta\in L^2(\Omega;\mathbb{R}^{d+1})} \frac{\|\xi-\eta\|^2}{2\tau_1} + \mathcal{F}(\eta)\,. \tag{7.2}$$

This leads to the so-called PDGH algorithm, which for $\tau_1\tau_2 < 1$ can be seen just as a proximal point method applied to a monotone operator [12], and therefore we still have convergence in the Hilbert space setting. As mentioned in [22] convergence of the discrete problem with mesh refinement is more delicate and will not be discussed here.

7.1. **Mixed L^2-Wasserstein distance.** Define for any $\sigma = (\rho, m) \in L^2(\Omega) \times L^2(\Omega;\mathbb{R}^d)$

$$\mathcal{R}(\sigma) := \begin{cases} \frac{1}{2}\|\partial_t\rho\|^2_{L^2(\Omega)} & \text{if } \partial_t\rho \in L^2(\Omega)\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{7.3}$$

With this functional, problem (7.1) yields an interpolation between the Wasserstein distance and the L^2 distance. It was originally considered in [6], where a conjugate gradient method was proposed to compute the minimizers. Let $V := H^1([0,1];L^2(D)) \times L^2([0,1];H(\mathrm{div};D))$ and let $\mathring{V}$ be the same space with homogenous boundary conditions on the fluxes. For any $\xi \in L^2(\Omega)^{d+1}$, $\sigma = \mathrm{prox}_{\tau_1\mathcal{F}}(\xi)$ is obtained by solving the following system for $(\sigma,\phi) \in V \times L^2(\Omega)/\mathbb{R}$

$$\begin{cases} \langle\sigma, v\rangle + \alpha\tau_1\langle\partial_t\rho, \partial_t v_t\rangle + \langle\phi, \mathrm{div}\, v\rangle = \langle\xi, v\rangle\,, & \forall v \in \mathring{V}\,, \\ \langle\mathrm{div}\,\sigma, \psi\rangle = 0\,, & \forall\psi \in L^2(\Omega)/\mathbb{R}\,, \\ \sigma\cdot n_{\partial\Omega} = \mathcal{X}\,, \end{cases} \tag{7.4}$$

where $v_t = v\cdot\hat{t}$ is the component of v in the time direction. Well-posedness can be obtained by standard methods for saddle point problems [10] and it translates directly into well-posedness of the discrete system obtained by replacing V with $Z_{h,\tau}$, $L^2(\Omega)$ with $F_{h,\tau}$, and $\mathcal{X}$ with $\mathcal{X}_{h,\tau}$.

7.2. **H^1 regularization.** Define for any $\sigma = (\rho, m) \in L^2(\Omega) \times L^2(\Omega;\mathbb{R}^d)$

$$\mathcal{R}(\sigma) := \begin{cases} \frac{1}{2}\|\nabla_x\rho\|^2_{L^2(\Omega)} & \text{if } \rho \in L^2([0,1];H^1(D))\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{7.5}$$

In this case we set $V := H(\mathrm{div};\Omega)$, $W := L^2([0,1];H(\mathrm{div}_x;D))$ and let $\mathring{V}$ and $\mathring{W}$ be the same spaces with homogenous boundary conditions on the fluxes. Then, for any $\xi \in L^2(\Omega)^{d+1}$, $\sigma = \mathrm{prox}_{\tau_1\mathcal{F}}(\xi)$ is obtained by solving the following system for $(\sigma,\eta,\phi) \in V \times \mathring{W} \times L^2(\Omega)/\mathbb{R}$

$$\begin{cases} \langle\sigma, v\rangle - \alpha\tau_1\langle\mathrm{div}_x\eta, v_t\rangle + \langle\phi, \mathrm{div}\, v\rangle = \langle\xi, v\rangle\,, & \forall v \in \mathring{V}\,, \\ \langle\rho, \mathrm{div}_x w\rangle + \langle\eta, w\rangle = 0\,, & \forall w \in \mathring{W}\,, \\ \langle\mathrm{div}\,\sigma, \psi\rangle = 0\,, & \forall\psi \in L^2(\Omega)/\mathbb{R}\,, \\ \sigma\cdot n_{\partial\Omega} = \mathcal{X}\,, \end{cases} \tag{7.6}$$

where $v_t = v \cdot \hat{t}$ is the component of v in the time direction. As before, well-posedness can be obtained by standard methods for saddle point problems [10].

We introduce the space $W_{h,\tau} \subset L^2([0,1]; H(\mathrm{div}_x; D))$ whose shape functions on $S = [t_0, t_1] \otimes T$ are given by

$$W_{h,\tau}(S) \coloneqq \mathcal{P}_1([t_0, t_1]) \otimes V_h(T). \tag{7.7}$$

We denote by $\mathring{W}_{h,\tau}$ the same space with the boundary conditions $\eta \cdot n_{\partial\Omega} = 0$ on $[0,1] \times \partial D$. Denote by $\nabla_x^h : L^2(\Omega) \to \mathring{W}_{h,\tau}$ the adjoint of $-\mathrm{div}_x$ defined by

$$\langle \nabla_x^h \phi, \eta \rangle = -\langle \phi, \mathrm{div}_x \eta \rangle \,, \quad \forall\, (\phi, \eta) \in L^2(\Omega) \times \mathring{W}_{h,\tau} \tag{7.8}$$

We define a discrete version of (7.5) as follows:

$$\mathcal{R}_{h,\tau}(\sigma) \coloneqq \frac{1}{2} \|\nabla_x^h \rho\|^2_{L^2(\Omega)} \tag{7.9}$$

Let $\mathcal{F}_{h,\tau} \coloneqq \iota_{\mathcal{C}_{h,\tau}} + \alpha \mathcal{R}_{h,\tau}$. Then for any $\xi \in L^2(\Omega)^{d+1}$, $\sigma = \mathrm{prox}_{\tau_1 \mathcal{F}_{h,\tau}}(\xi)$ is obtained by solving the following system for $(\sigma, \eta, \phi) \in \mathring{V}_{h,\tau} \times \mathring{W}_{h,\tau} \times F_{h,\tau}/\mathbb{R}$:

$$\begin{cases} \langle \sigma, v \rangle - \alpha \tau_1 \langle \mathrm{div}_x \eta, v_t \rangle + \langle \phi, \mathrm{div}\, v \rangle = \langle \xi, v \rangle \,, & \forall v \in \mathring{V}_{h,\tau} \,, \\ \langle \rho, \mathrm{div}_x w \rangle + \langle \eta, w \rangle = 0 \,, & \forall w \in \mathring{W}_{h,\tau} \,, \\ \langle \mathrm{div}\, \sigma, \psi \rangle = 0 \,, & \forall \psi \in F_{h,\tau}/\mathbb{R} \,, \\ \sigma \cdot n_{\partial\Omega} = \mathcal{X}_{h,\tau} \,. \end{cases} \tag{7.10}$$

8. Numerical results

In this section we describe two numerical tests that demonstrate the behaviour of the proposed discretization both qualitatively and in terms of convergence of the algorithm. For both tests the time discretization is fixed with a time step $\tau \coloneqq |t_{i+1} - t_i| = 1/30$, but we will use different meshes and finite element spaces for the discretization in space. The results shown hereafter have been obtained using the finite element software Firedrake [29] (see [26, 9], for the tensor product constructions) and the linear solver for the Poisson equation is based on PETSc [3, 4]. The code to perform the tests in this section can be found in the repository https://github.com/andnatale/dynamic-ot.git.

8.1. Gaussian interpolation.

We set $D = [0,1]^2$ and

$$\rho_0 = \exp\left(-\frac{|x - x_0|^2}{2s^2}\right), \quad \rho_1 = \exp\left(-\frac{|x - x_1|^2}{2s^2}\right), \tag{8.1}$$

where $s = 0.1$, $x_0 = (0.35, 0.35)$ and $x_1 = (0.65, 65)$. The optimal transport interpolation between the density in equation (8.1) when the domain is $\mathbb{R}^2$ is simply given by

$$\rho_t = \exp\left(-\frac{|x - (1-t)x_0 - t x_1|^2}{2s^2}\right). \tag{8.2}$$

This is still a good approximation on our bounded domain and it allows us to assess qualitatively the different discretizations we have proposed. In figure 2, it is shown the interpolation at time $t = 0.5$ for different choices of spaces V_h and X_h. In general, for the same space V_h, the discretizations corresponding to X_h^1 are slightly more diffusive than those corresponding to X_h^0, but they also lead to smaller oscillations on the positivity. On the other hand for the couple of spaces $V_h = \mathcal{BDM}_1$, X_h^0, we observe some oscillations appearing in the interpolation (see also remark 5.3). In figure (3), the different schemes are compared in terms of convergence of the proximal splitting algorithm. The cases corresponding to X_h^1 appear to converge faster than X_h^0 (when used jointly with the spaces $\mathcal{BDM}_1$ and $\mathcal{RT}_{[0]}$).

8.2. **Non-convex domain.** We now consider a non-convex domain polygonal D represented in figure 4. Note that even if the case of a non-convex domain is beyond the domain of applicability of the convergence results presented in this paper, our scheme is still well-defined for this case. The boundary conditions are still given by (8.1), but with $s = 0.1$, $x_0 = (0.5, 0.1)$ and $x_1 = (0.5, 0.9)$. The mesh $\mathcal{T}_h$ and the boundary conditions are illustrated in figure 4. In this case, the exact density interpolation is not absolutely continuous, since mass concentrates on the segment connecting the two non-convex corners of the domain. Note that we have therefore refined the mesh along the diagonal where we expect the mass to concentrate.

In figure 5, 6 and 7 we show the density evolution up to time $t = 0.5$ (the other half of the time evolution being symmetric in space given the boundary conditions and the domain shape) for the non-regularized case, the H^1 regularization and the L^2 regularization, respectively. For both regularizations the density profile appears to be smoothened, but only the H^1 regularization avoids concetration at the the corners.

The proximal operator of the projection on the continuity equation is more expensive computationally for the H^1 regularization than for the other two cases, since we have to solve a larger mixed system at each iteration. However, for both regularizations, the proximal splitting algorithm itself converges much faster than the non-regularized case, as it can be seen in figure 8.

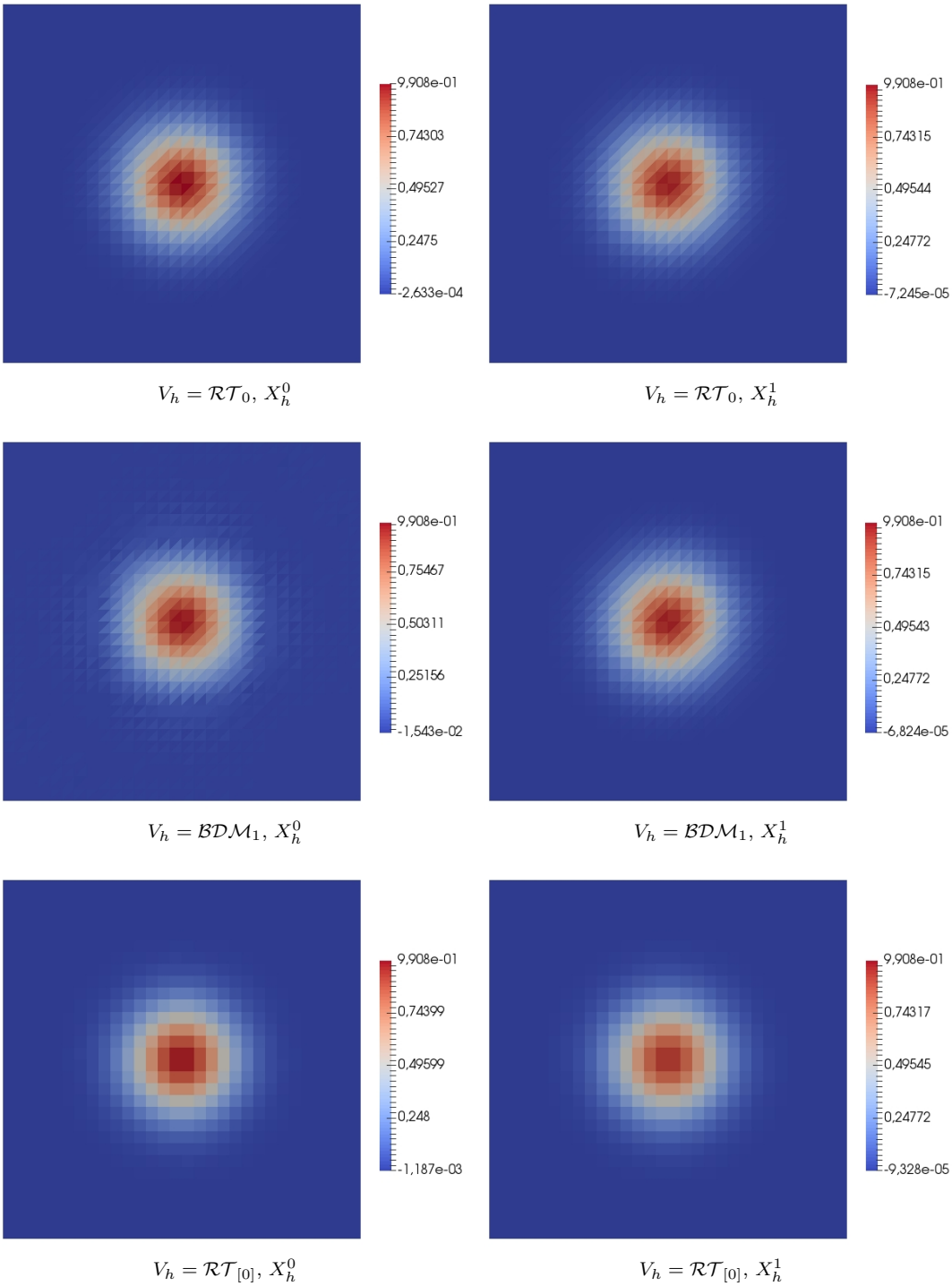


FIGURE 2. Comparison between Optimal Transport interpolation of two Gaussians for different spaces.

Appendix A. Proof of theorem 5.4

Applying theorem 2.13 in [22], in order to prove theorem 5.4 it is sufficient to check that the conditions listed in definition 2.9 of [22] are verified. Such conditions translated to our finite element settings are listed in proposition A.2 below.

From now on, we assume $r = 1$, and X_h stands for X_h^1. We also denote by $\mathcal{I}$ is the standard nodal interpolant onto X_h, defined element by element.

First of all, we introduce some notation and list some technical results [13]. Denote by P_{X_h} and P_{V_h} the L^2 projections onto X_h and V_h, respectively. Then,

$$\|P_{X_h}\varphi\|_{L^p} \leq C\|\varphi\|_{L^p}\,, \quad \forall \varphi \in L^p\,,\, 1 \leq p \leq \infty\,, \tag{A.1}$$

and moreover $\forall\, T \in \mathcal{T}_h$

$$\|\varphi - P_{X_h}\varphi\|_{L^p(T)} \leq Ch_T\|\nabla\varphi\|_{L^p(T)}\,, \quad \forall \varphi \in W^{1,p}(T)\,,\, 1 \leq p \leq \infty\,, \tag{A.2}$$

where, with an abuse of notation, we have used P_{X_h} to denote the L^2 projection onto $X_h(T)$. These imply the following lemma.

Lemma A.1. *Given the regularity assumption in* (4.1) *on* $\mathcal{T}_h$*, we have*

$$\|\mathcal{I}|P_{V_h}b|^2\|_{L^\infty} \leq C\|b\|^2_{L^\infty} \tag{A.3}$$

for any $b \in L^\infty(D)$*, and*

$$\|\mathcal{I}|P_{V_h}b|^2 - |b|^2\|_{L^\infty} \leq Ch|b|_{W^{1,\infty}}\|b\|_{L^\infty} \tag{A.4}$$

for any $b \in W^{1,\infty}(D)$*.*

Proof. For the first inequality, using standard inverse inequalities, we have

$$\begin{aligned}
\|\mathcal{I}|P_{V_h}b|^2\|_{L^\infty} &\leq \||P_{V_h}b|^2\|_{L^\infty} \\
&\leq Ch^{-d}\||P_{V_h}b|^2\|_{L^1} \\
&= Ch^{-d}\|P_{V_h}b\|^2_{L^2} \\
&\leq Ch^{-d}\|P_{(X_h)^d}b\|^2_{L^2} \\
&\leq C\|P_{(X_h)^d}b\|^2_{L^\infty} \\
&\leq C\|b\|^2_{L^\infty}
\end{aligned}$$

For the second inequality , we observe that

$$\|\mathcal{I}|P_{V_h}b|^2 - |b|^2\|_{L^\infty} \leq \|\mathcal{I}|P_{V_h}b|^2 - \mathcal{I}|b|^2\|_{L^\infty} + \|\mathcal{I}|b|^2 - |b|^2\|_{L^\infty} \tag{A.5}$$

The second term of the right-hand side is easy to control. For the first term, we have

$$\begin{aligned}
\|\mathcal{I}|P_{V_h}b|^2 - \mathcal{I}|b|^2\|_{L^\infty} &\leq \||P_{V_h}b|^2 - |b|^2\|_{L^\infty} \\
&\leq \||P_{V_h}b|^2 - |P_{(X_h)^d}b|^2\|_{L^\infty} + \||b|^2 - |P_{(X_h)^d}b|^2\|_{L^\infty}
\end{aligned}$$

Again, the second term is easy to control. For the first tem, using the same reasoning as above,

$$\begin{aligned}
\||P_{V_h}b|^2 - |P_{(X_h)^d}b|^2\|_{L^\infty} &\leq Ch^{-d}\||P_{V_h}b|^2 - |P_{(X_h)^d}b|^2\|_{L^1} \\
&\leq Ch^{-d}\sum_{i=1}^d \|(P_{V_h}b)_i^2 - (P_{X_h}b_i)^2\|_{L^1} \\
&\leq Ch^{-d}\sum_{i=1}^d \|(P_{V_h}b)_i - P_{X_h}b_i\|_{L^1}\|b\|_{L^\infty} \\
&\leq Ch^{-\frac{d}{2}}\|P_{V_h}b - P_{X_h}b\|_{L^2}\|b\|_{L^\infty} \\
&\leq Ch\|\nabla P_{X_h}b\|_{L^\infty}\|b\|_{L^\infty} \leq Ch\|\nabla b\|_{L^\infty}\|b\|_{L^\infty}
\end{aligned}$$

□

We will need also some approximation properties for the projection operators Π_{Q_h} and Π_{V_h}. Specifically, we will assume that these verify for any $\phi \in H^1(D)$ and $\eta \in H^1(D)^d$

$$\|\Pi_{Q_h}\phi - \phi\|_{L^2(D)} \leq Ch\|\phi\|_{H^1(D)}\,, \quad \|\Pi_{V_h}\eta - \eta\|_{L^2(D)^d} \leq Ch\|\eta\|_{H^1(D)^d}\,. \tag{A.6}$$

Given the mesh regularity assumption (4.1), this is a standard property for Π_{Q_h} as defined in equation (4.5), whereas for Π_{V_h} (since this has not been defined explicitly) we refer to the construction discussed in [2] (see, in particular, theorem 5.6) based on a Clément interpolant, for example.

Proposition A.2 below contains the properties needed for convergence: it can be seen as a specific instance of Definition 2.9 of [22]. Note that a few of the properties listed therein are omitted here because they are either unnecessary or true by construction in our setting. Note also that the sampling operators used in [22] are replaced here with the bounded projections Π_{Q_h} and Π_{V_h}, where Π_{Q_h} can be naturally extended to $\mathcal{M}(D)$ (see equation (4.5)) and Π_{V_h} is considered to be defined on a domain different than $\mathcal{M}(D)^d$. Moreover the reconstruction operators are simply the injection operators from Q_h and V_h to $\mathcal{M}(D)$ and $\mathcal{M}(D)^d$, respectively. Finally, we define for any $(\rho, b) \in \mathcal{M}(D) \times C(D;\mathbb{R}^d)$

$$A^*(\rho, b) := \int_D \frac{|b|^2}{2}\rho\,, \tag{A.7}$$

so that if $(\rho, m) \in \mathcal{M}_+(D) \times \mathcal{M}(D)^d$ then

$$A(\rho, m) = \sup_{b\in C(D;\mathbb{R}^d)} \langle m, b\rangle - A^*(\rho, b); \tag{A.8}$$

and for any $(\rho, b) \in Q_h \times V_h$,

$$A_h^*(\rho, b) := \sup_{m\in V_h} \langle m, b\rangle - A_h(\rho, m)\,. \tag{A.9}$$

Proposition A.2. *The following properties hold:*

(1) *For any $\rho \in \mathcal{M}_+(D)$, $\Pi_{Q_h}\rho \to \rho$ as $h \to 0$ weakly in $\mathcal{M}(D)$.*
(2) *Let $B \subset (C^1(D))^d$ a bounded subset. Then there exists a constant ε_h tending to 0 as $h \to 0$ such that for any $b \in B$ and $\rho \in Q_h$*

$$A_h^*(\rho, P_{V_h}b) \leq A^*(\rho, b) + \epsilon_h\|\rho\| \tag{A.10}$$

where P_{V_h} denotes the L^2 projection onto V_h. Moreover there exists a constant $C \geq 1$ such that for any $b \in C(D)^d$, there holds

$$A_h^*(\rho, P_{V_h}b) \leq \frac{C}{2}\|\rho\|\|b\|_{L^\infty}^2 \tag{A.11}$$

(3) *Let $B \subset C^0(D)\cap H^1(D)$ a bounded subset such that for all $\rho \in B$ there holds $\rho > C > 0$, and let $B' \subset (C^0(D) \cap H^1(D))^d$ a bounded subset. There exists a constant ε_h tending to 0 as $h \to 0$ such that, given $(\rho, m) \in \mathcal{M}(D)^{d+1}$ such that ρ has density in B and m in B', then*

$$A_h(\Pi_{Q_h}\rho, \Pi_{V_h}m) \leq A(\rho, m) + \varepsilon_h. \tag{A.12}$$

(4) *There exists ε_h tending to 0 as $h \to 0$ and a continuous function ω satisfying $\omega(0) = 0$ such that: for any $x, y \in D$ and $h > 0$ there exists $\rho \in Q_h^+$ and $m_1, m_2 \in V_h$ satisfying*

$$\begin{cases} \operatorname{div} m_1 = \rho - \Pi_{Q_h}(\delta_x) \\ \operatorname{div} m_2 = \rho - \Pi_{Q_h}(\delta_y) \end{cases} \quad \text{and} \quad A_h(\rho, m_i) \leq \omega(|x-y|)+\varepsilon_h\,, \forall i \in \{1, 2\}. \tag{A.13}$$

Remark A.3. *In [22] point (3) of proposition A.2 is stated with B and B' bounded subsets of $C^1(D)$ and $C^1(D)^d$, respectively. The condition we require here is stronger, but it is needed since we considered a convex polytope domain rather than a domain with a smooth boundary as in [22]. As a matter of fact, in [22] one applies the condition (3) on a regularized measure $(\tilde{\rho}, \tilde{m}) \in \mathcal{M}(D)^{d+1}$ obtained by convolution with the heat kernel and by solving an appropriate elliptic problem (see proposition 3.2 in [22]). For a convex polytope domain this procedure yields a couple $(\tilde{\rho}, \tilde{m})$ with densities which are not C^∞ given the singularities of the boundary. By classical elliptic regularity estimates on non-smooth domains (e.g., [31] and [25]), the regularity we require in condition (3) is however sufficient for the proof in [22] to apply without changes.*

Proof. The first point is immediate from the definition of Π_{Q_h} in equation (4.5). For (2), we observe that

$$A_h(\rho, m) = \sup_{b \in X_h} \langle m, b\rangle - \frac{1}{2}\langle \rho, \mathcal{I}|b|^2\rangle \tag{A.14}$$

where we recall that $\mathcal{I}$ is the standard element-wise nodal interpolant onto X_h. In fact, for any $b \in (X_h)^d$, we have $b^2 \leq \mathcal{I}|b|^2$, and therefore when $\rho \geq 0$ we can "saturate" the constraint setting $a = -\mathcal{I}|b|^2/2$. On the other hand if $\rho < 0$ on some element both sides of the equality are $+\infty$. For $(\rho, b, m) \in Q_h \times V_h \times V_h$ define

$$A^*_{\mathcal{I},h}(\rho, b) := \frac{1}{2}\langle \rho, \mathcal{I}|b|^2\rangle, \quad \bar{A}_{\mathcal{I},h}(\rho, m) := \sup_{b \in V_h} \langle m, b\rangle - A^*_{\mathcal{I},h}(\rho, b). \tag{A.15}$$

Then, since when $\rho < 0$ on some element $A^*_h(\rho, b) = -\infty$,

$$A_h(\rho, m) \geq \bar{A}_{\mathcal{I},h}(\rho, m), \quad A^*_h(\rho, b) \leq \bar{A}^*_{\mathcal{I},h}(\rho, b) \leq A^*_{\mathcal{I},h}(\rho, b), \tag{A.16}$$

and we can prove (2) for $A^*_{\mathcal{I},h}$. In particular, we have

$$A^*_{\mathcal{I},h}(\rho, P_{V_h} b) \leq A^*(\rho, b) + \frac{1}{2}\|\mathcal{I}|P_{V_h} b|^2 - |b|^2\|_{L^\infty}\|\rho\|, \tag{A.17}$$

and we obtain the result applying lemma A.1. Using again lemma A.1, we easily obtain the second bound as well.

For point (3), observe first that $A_h(\Pi_{Q_h}\rho, \Pi_{V_h} m) \leq A(\Pi_{Q_h}\rho, \Pi_{V_h} m)$ by definition. Then given the assumption on ρ and m we can simply write

$$\begin{aligned} A_h(\Pi_{Q_h}\rho, \Pi_{V_h} m) - A(\rho, m) &\leq \int_D \frac{|\Pi_{V_h} m|^2}{2\Pi_{Q_h}\rho} - \frac{|m|^2}{2\rho} \\ &\leq \frac{1}{2}\int_D \Big|\frac{|\Pi_{V_h} m|^2 - |m|^2}{\Pi_{Q_h}\rho}\Big| + \Big|\frac{|m|^2}{\Pi_{Q_h}\rho} - \frac{|m|^2}{\rho}\Big| \\ &\leq C(\|\Pi_{Q_h}\rho - \rho\|_{L^2} + \||\Pi_{V_h} m|^2 - |m|^2\|_{L^1}) \end{aligned} \tag{A.18}$$

where the constant C depends on the uniform lower bound on ρ and on the L^∞ norm of $|m|$. We conclude using Cauchy–Schwarz inequality on the second term and then equation (A.6).

For the last point, we will establish a connection between our scheme and the one proposed by Gladbach, Kopfer and Maas [16] and then use propoperty (A.13) for this scheme which was proved in [22]. We will consider the case of a simplicial mesh, the quadrilateral case being completely analogous upon substituting $\mathcal{RT}_0$ with $\mathcal{RT}_{[0]}$ in the discussion below.

First, we introduce some notation. For each $T \in \mathcal{T}_h$, let $\mathcal{T}_{h,T}$ be the set of neighbouring elements $L \in \mathcal{T}_h$ such that $f_{T,L} := \overline{T} \cap \overline{L} \neq \emptyset$, which we assume to be oriented. Define by $\mathcal{F}_h$ the set of $(d-1)$-dimensional facets in the triangulation. Let $T, L \in \mathcal{T}_h$ be neighbouring elements, we denote by $\varphi_{T,L} \in \mathcal{RT}_0$ the canonical

basis function associated with the oriented facet $f_{T,L}$. Then, any $m \in \mathcal{RT}_0$ can be written as

$$m = \sum_{f_{T,L} \in \mathcal{F}_h} m_{T,L} \varphi_{T,L} \tag{A.19}$$

where $m_{T,L}$ is the flux of m on the oriented facet $f_{T,L}$. In other words we can identify functions in $(\rho, m) \in Q_h \times V_h$ with their finite volume representation $\{\rho_T, m_{T,L}\}_{T,L}$. Then, we can interpret the action for the finite volume scheme [16], which we denote by $A_h^{FV}(\rho, m)$, as a function on $Q_h \times V_h$. This is given by the following expression

$$A_h^{FV}(\rho, m) := \sum_{f_{T,L} \in \mathcal{F}_h} \frac{m_{T,L}^2}{2\theta(\rho_T, \rho_L)} |f_{T,L}||x_T - x_L| \tag{A.20}$$

where $\theta : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is an appropriate function (see [16]) which we take to be the harmonic mean.

Now, in order to construct $\rho \in Q_h^+$ and $m_1, m_2 \in V_h$ satisfying (A.13), we use the same construction as in [22] for the finite volume scheme, and interpolate this to the spaces $\mathcal{RT}_0 \subseteq V_h$ and Q_h^+ to obtain ρ, m_1 and m_2 satisfying

$$\begin{cases} \operatorname{div} m_1 = \rho - \Pi_{Q_h}(\delta_x) \\ \operatorname{div} m_2 = \rho - \Pi_{Q_h}(\delta_y) \end{cases} \tag{A.21}$$

In particular the support of ρ, m_1 and m_2 is a chain of neighbouring elements $T_1, \dots, T_N$. To prove the bound on the action, we observe that $A_h(\rho, m_i) \leq A(\rho, m_i)$. Then, we only need to bound $A(\rho, m_i)$ by the action of the finite-volume scheme $A_h^{FV}(\rho, m_i)$, since A_h^{FV} satisfies the desired inequality thanks to the regularity assumption (4.1) on the mesh [22].

By the regularity assumption on the triangulation, we can assume

$$\int_{T \cup L} |\varphi_{T,L}|^2 \, \mathrm{d}x \leq C |f_{T,L}||x_T - x_L| \tag{A.22}$$

uniformly. Then, by explicit calculations we obtain $A(\rho, m_i) \leq C A_h^{FV}(\rho, m_i)$ and we are done.

□

Acknowledgements

The work of A. Natale was supported by the European Research Council (ERC project NORIA). G. Todeschi acknowledges that this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754362.

References

[1] Douglas N Arnold, Daniele Boffi, and Francesca Bonizzoni. Finite element differential forms on curvilinear cubic meshes and their approximation properties. *Numer. Math.*, 2014. arXiv:1204.2595.

[2] Douglas N Arnold, Richard S Falk, and Ragnar Winther. Finite element exterior calculus, homological techniques, and applications. *Acta numerica*, 15:1–155, 2006.

[3] Satish Balay, Shrirang Abhyankar, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Lisandro Dalcin, Victor Eijkhout, William D. Gropp, Dmitry Karpeyev, Dinesh Kaushik, Matthew G. Knepley, Dave A. May, Lois Curfman McInnes, Richard Tran Mills, Todd Munson, Karl Rupp, Patrick Sanan, Barry F. Smith, Stefano Zampini, Hong Zhang, and

Hong Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.11, Argonne National Laboratory, 2019.

[4] Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhäuser Press, 1997.

[5] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[6] Jean-David Benamou and Yann Brenier. Mixed l 2-wasserstein optimal mapping between prescribed density functions. *Journal of Optimization Theory and Applications*, 111(2):255–271, 2001.

[7] Jean-David Benamou and Guillaume Carlier. Augmented lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167(1):1–26, 2015.

[8] Jean-David Benamou, Guillaume Carlier, and Maxime Laborde. An augmented lagrangian approach to wasserstein gradient flows and applications. *ESAIM: Proceedings and Surveys*, 54:1–17, 2016.

[9] Gheorghe-Teodor Bercea, Andrew T. T. McRae, David A. Ham, Lawrence Mitchell, Florian Rathgeber, Luigi Nardi, Fabio Luporini, and Paul H. J. Kelly. A structure-exploiting numbering algorithm for finite elements on extruded meshes, and its performance evaluation in firedrake. *Geoscientific Model Development*, 9(10):3803–3815, 2016.

[10] Daniele Boffi, Franco Brezzi, Michel Fortin, et al. *Mixed finite element methods and applications*, volume 44. Springer, 2013.

[11] Jose A Carrillo, Katy Craig, Li Wang, and Chaozhen Wei. Primal dual methods for wasserstein gradient flows. *arXiv preprint arXiv:1901.08081*, 2019.

[12] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[13] Jim Douglas, Todd Dupont, and Lars Wahlbin. The stability in L^q of the L^2-projection into finite element function spaces. *Numerische Mathematik*, 23(3):193–197, 1974.

[14] Matthias Erbar, Martin Rumpf, Bernhard Schmitzer, and Stefan Simon. Computation of optimal transport on discrete metric measure spaces. *Numerische Mathematik*, 144(1):157–200, 2020.

[15] Thomas Gallouët, Maxime Laborde, and Leonard Monsaingeon. An unbalanced optimal transport splitting scheme for general advection-reaction-diffusion problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:8, 2019.

[16] Peter Gladbach, Eva Kopfer, and Jan Maas. Scaling limits of discrete optimal transport. *arXiv preprint arXiv:1809.01092*, 2018.

[17] Kevin Guittet. On the time-continuous mass transport problem and its approximation by augmented lagrangian techniques. *SIAM Journal on Numerical Analysis*, 41(1):382–399, 2003.

[18] Morgane Henry, Emmanuel Maitre, and Valérie Perrier. Primal-dual formulation of the dynamic optimal transport using helmholtz-hodge decomposition. 2019.

[19] Romain Hug, Emmanuel Maitre, and Nicolas Papadakis. Multi-physics optimal transportation and image interpolation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1671–1692, 2015.

[20] Romain Hug, Emmanuel Maitre, and Nicolas Papadakis. On the convergence of augmented lagrangian method for optimal transport between nonnegative densities. 2017.

[21] Noureddine Igbida and Van Thanh Nguyen. Augmented Lagrangian Method for Optimal Partial Transportation. *IMA Journal of Numerical Analysis*, 38(1):156–183, 03 2017.

[22] Hugo Lavenant. Unconditional convergence for discretizations of dynamical optimal transport. *arXiv preprint arXiv:1909.08790*, 2019.

[23] Hugo Lavenant, Sebastian Claici, Edward Chien, and Justin Solomon. Dynamical optimal transport on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 37(6):1–16, 2018.

[24] Wuchen Li, Penghang Yin, and Stanley Osher. Computations of optimal transport distance with fisher information regularization. *Journal of Scientific Computing*, 75(3):1581–1595, 2018.

[25] Gary M Lieberman. Oblique derivative problems in Lipschitz domains. II. Discontinuous boundary data. *J. reine angew. Math*, 389:1–21, 1988.

[26] Andrew T. T. McRae, Gheorghe-Teodor Bercea, Lawrence Mitchell, David A. Ham, and Colin J. Cotter. Automated generation and symbolic manipulation of tensor product finite elements. *SIAM Journal on Scientific Computing*, 38(5):S25–S47, 2016.

[27] Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[28] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. An algorithm for minimizing the mumford-shah functional. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1133–1140. IEEE, 2009.

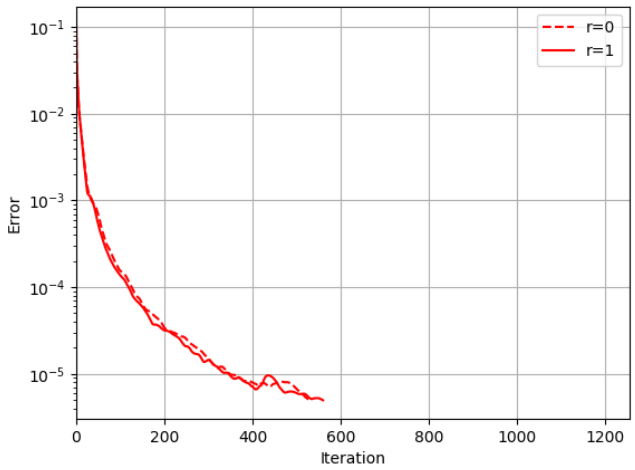
[29] Florian Rathgeber, David A. Ham, Lawrence Mitchell, Michael Lange, Fabio Luporini, Andrew T. T. McRae, Gheorghe-Teodor Bercea, Graham R. Markall, and Paul H. J. Kelly. Firedrake: automating the finite element method by composing abstractions. *ACM Trans. Math. Softw.*, 43(3):24:1–24:27, 2016.

[30] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, pages 99–102, 2015.

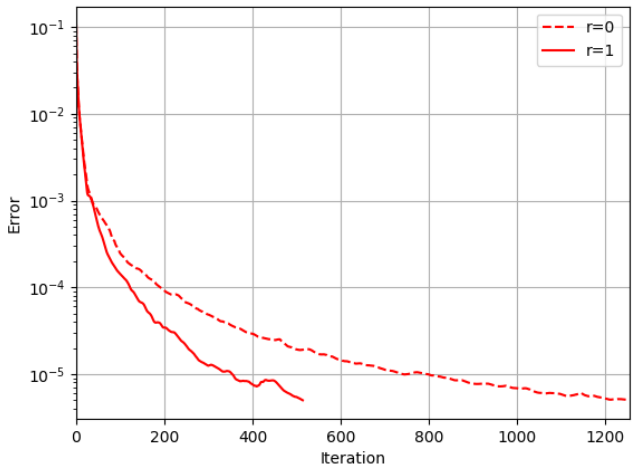
[31] Guido Stampacchia. Problemi al contorno ellittici, con dati discontinui, dotati di soluzioni hölderiane. *Annali di Matematica pura ed applicata*, 51(1):1–37, 1960.

Andrea Natale (andrea.natale@u-psud.fr) Laboratoire de Mathematiques d'Orsay, Université Paris-Sud

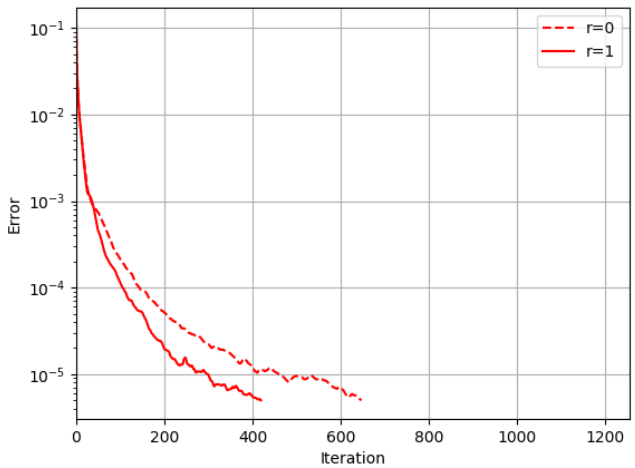
Gabriele Todeschi (gabriele.todeschi@inria.fr): Inria, Project team Mokaplan, Université Paris-Dauphine, PSL Research University, UMR CNRS 7534-Ceremade



$V_h = \mathcal{RT}_0$



$V_h = \mathcal{BDM}_1$



$V_h = \mathcal{RT}_{[0]}$

FIGURE 3. Convergence of the proximal splitting algorithm measured by $\|\sigma_{n+1} - \sigma_n\|_{L^2(\Omega)}$ for the optimal transport problem between two Gaussians for different spaces V_h and X_h^r.

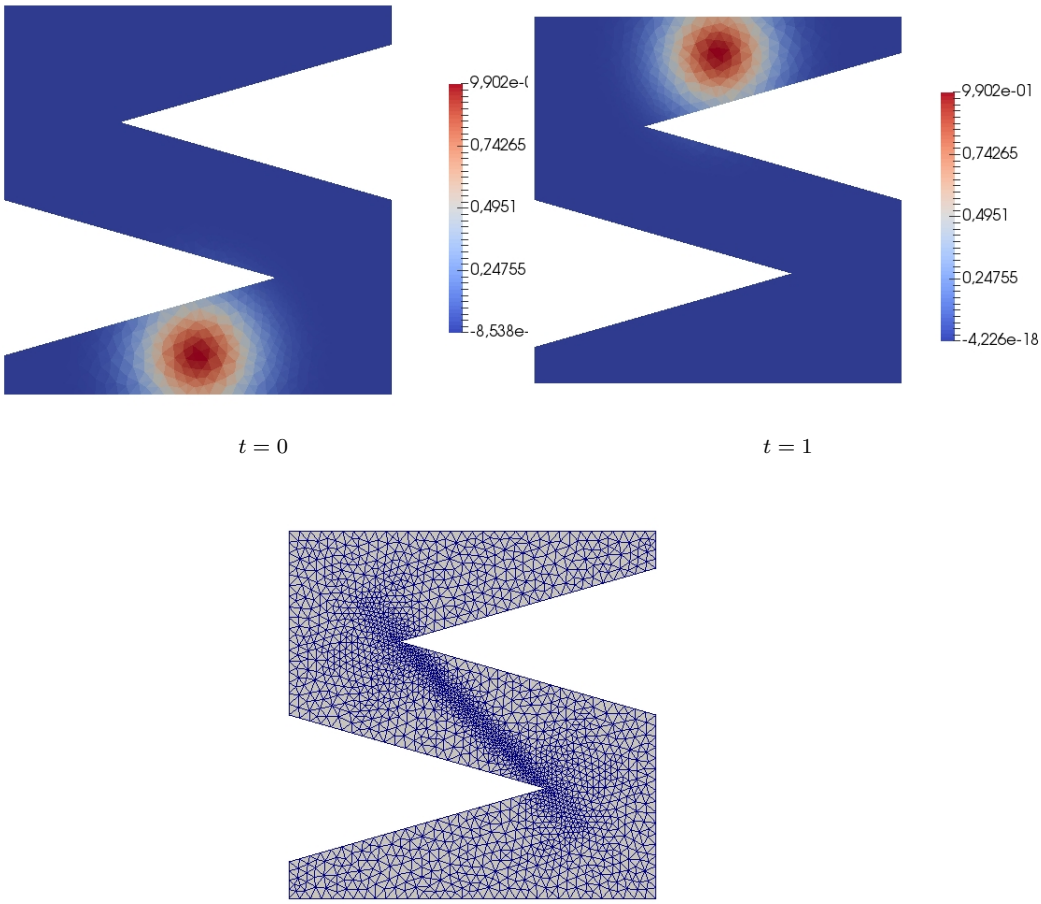


FIGURE 4. Initial and final density (top) and mesh (bottom) for the non-convex domain test.

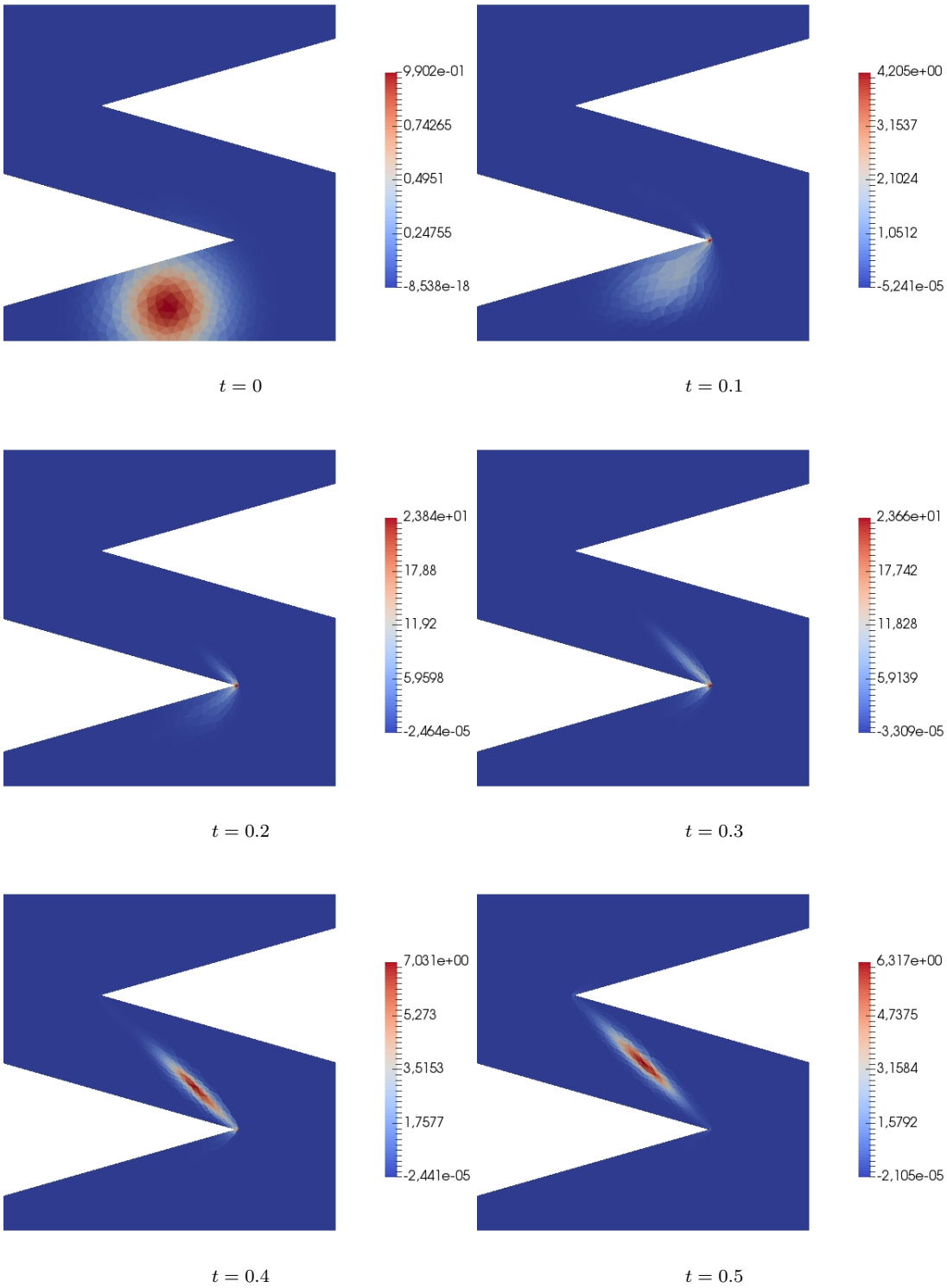


FIGURE 5. Density evolution for the non-convex domain test without regularization, $V_h = \mathcal{RT}_0$, X_h^1.

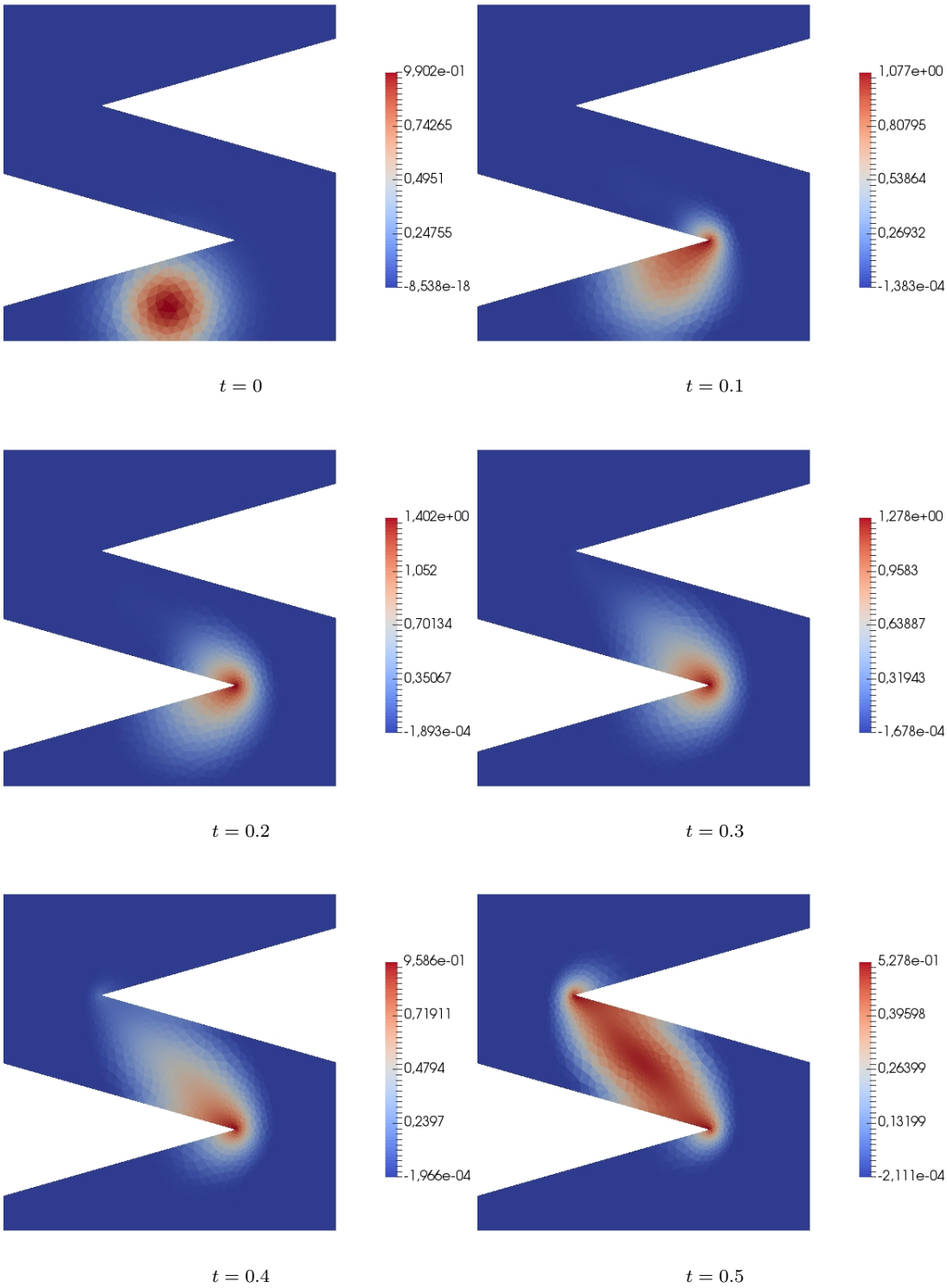


FIGURE 6. Density evolution for the non-convex domain test with the H^1 regularization $\alpha = 0.002$, $V_h = \mathcal{RT}_0$, X_h^1.

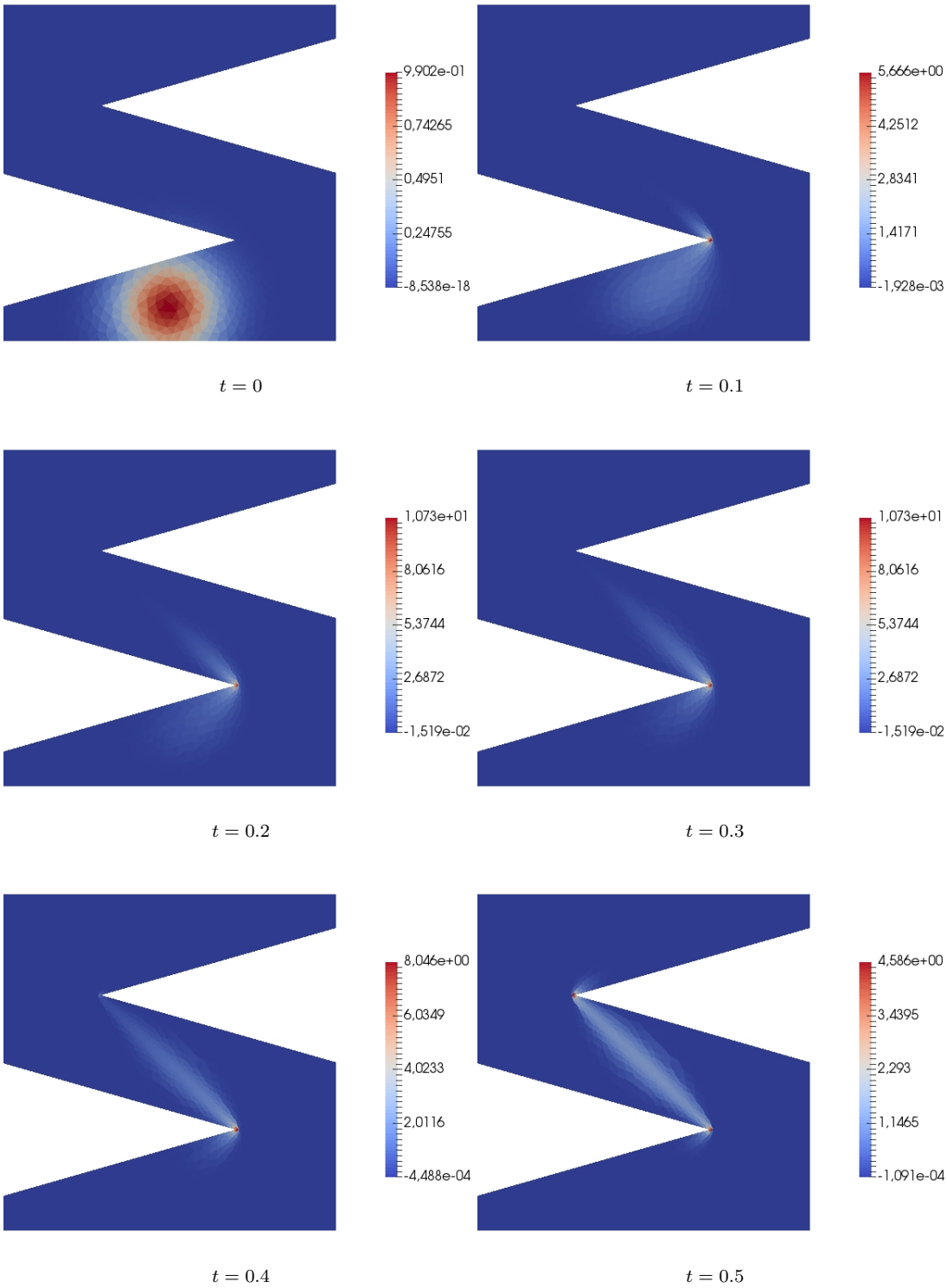


FIGURE 7. Density evolution for the non-convex domain test with the L^2 regularization $\alpha = 0.002$, $V_h = \mathcal{RT}_0$, X_h^1.

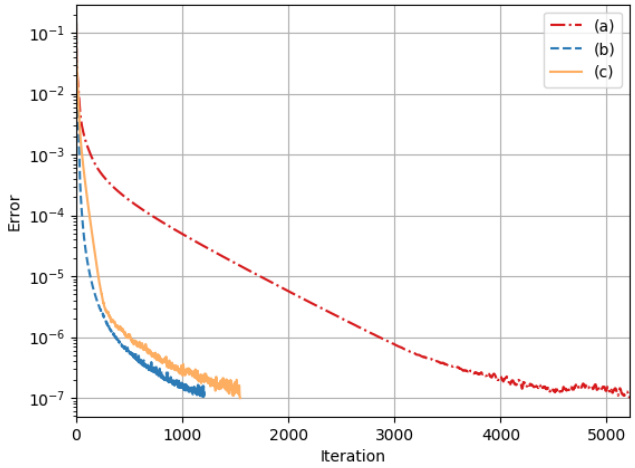


FIGURE 8. Convergence of the proximal splitting algorithm measured by $\|\sigma_{n+1} - \sigma_n\|_{L^2(\Omega)}$ for non-convex domain test without regularization (a); with the H^1 regularization and $\alpha = 0.002$ (b); with the L^2 regularization and $\alpha = 0.002$ (c).