HAL
open science

A numerical solution to Monge's problem with a Finsler distance as cost

Jean-David Benamou, Guillaume Carlier, Roméo Hatchi

To cite this version:

Jean-David Benamou, Guillaume Carlier, Roméo Hatchi. A numerical solution to Monge's problem with a Finsler distance as cost. ESAIM: Mathematical Modelling and Numerical Analysis, 2018. hal-01261094

HAL Id: hal-01261094

https://hal.science/hal-01261094v1

Submitted on 23 Jan 2016

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 numerical solution to Monge's problem with a Finsler distance as cost

Jean-David Benamou [*], Guillaume Carlier [†], Roméo Hatchi [‡]

January 17, 2016

Abstract

Monge's problem with a Finsler cost is intimately related to an optimal flow problem. Discretization of this problem and its dual leads to a well-posed finite-dimensional saddle-point problem which can be solved numerically relatively easily by an augmented Lagrangian approach in the same spirit as the Benamou-Brenier method for the optimal transport problem with quadratic cost. Numerical results validate the method. We also emphasize that the algorithm only requires elementary operations and in particular never involves evaluation of the Finsler distance or of geodesics.

Keywords: Monge's problem, Finsler distance, augmented Lagrangian.

MS Classification: 65K10, 90C25, 90C46.

1 Introduction

Given a bounded domain Ω of $\mathbb{R}^d$, and two probability measures f^+ and f^- on $\overline{\Omega}$, we are interested in solving Monge's problem

$$\inf_{\pi\in\Pi(f^-,f^+)} \int_{\overline{\Omega}\times\overline{\Omega}} d_L(x,y)\mathrm{d}\pi(x,y) \tag{1.1}$$

[*]INRIA Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay, FRANCE. Jean-David.Benamou@inria.fr.

[†]CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE carlier@ceremade.dauphine.fr

[‡]CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE hatchi@ceremade.dauphine.fr.

where $\Pi(f^-, f^+)$ is the set of transport plans between f^- and f^+ i.e. the set of probability measures having f^- and f^+ as marginals and d_L is a Finsler distance. More precisely, d_L is given by

$$d_L(x,y) := \inf\Big\{ \int_0^1 L(\gamma(t), \dot\gamma(t))\ \mathrm{d}t\ :\ \gamma \in W^{1,1}([0,1], \overline{\Omega}),\ \gamma(0) = x,\ \gamma(1) = y \Big\} \tag{1.2}$$

where the Lagrangian L: $\overline{\Omega} \times \mathbb{R}^d \to \mathbb{R}_+$ is a continuous function of Finsler type i.e. for every $x \in \overline{\Omega}$, $v \mapsto L(x,v)$ is a norm and there is a constant $C > 0$ such that the following nondegeneracy condition holds:

$$\frac{|v|}{C} \le L(x,v) \le C|v|,\ \forall (x,v) \in \overline{\Omega} \times \mathbb{R}^d. \tag{1.3}$$

Of course, one difficulty is the evaluation of the cost, and we shall see how to avoid computing it. This will be done by considering suitable dual, minimal flow and saddle-point formulations for which one can easily use an augmented Lagrangian method. The use of augmented Lagrangian methods in optimal transport was pioneered in the seminal work of Benamou and Brenier [4] on the dynamic formulation of the quadratic optimal transport case. For a distance cost (Monge case), in fact there is no need to introduce an additional time-variable and the analogue of the Benamou-Brenier dynamic problem is the minimal flow problem introduced by Beckmann [3]. We refer to the recent work [5] of the first two authors for applications of these augmented Lagrangian methods to Mean-Field-Games and optimal transport and to the work of the third author [13] for applications to anisotropic congested optimal transport. To the best of our knowledge, the relevance of augmented Lagrangian methods for a general Finsler metric has remained unnoticed in the literature. For other methods to solve optimal problems with the euclidean distance as transport cost, we refer for instance to [2] where a certain regularization is considered. Our goal is to show that Monge's problem with a Finsler metric is in fact quite easy to solve directly numerically by using an augmented Lagrangian approach. Let us finally mention the work of Rubinstein and Wolansky [15] which leads to problems similar to (1.1)-(1.2) having their roots in the study of semiclassical limits for some classes of dispersive wave equations.

The paper is organized as follows. In section 2, we recall several reformulations of the Monge problem with Finsler cost (1.1): the Kantorovich dual, the minimal flow reformulation and finally a (formal) saddle-point problem for finding at the same time the Kantorovich potential and the optimal flow field. Section 3 describes the discretization of the saddle-point problem, discusses the convergence and details the steps of the augmented algorithm

ALG2 of Glowinski and Fortin in this context. Section 4 gives numerical results.

2 Reformulations

2.1 Dual and minimal flow formulations

The standard Kantorovich duality formula (see [17]) says that the infimum in Monge's problem (1.1) coincides with the value of the dual:

$$\sup\Big\{\langle u, f\rangle := \int_{\overline{\Omega}} u(x)\mathrm{d}(f^+ - f^-)(x) \ : \ u \text{ is 1-Lipschitz for } d_L\Big\}. \tag{2.1}$$

Thanks to (1.3), it is easy to see that if u is 1-Lipschitz for d_L it is actually Lipschitz hence differentiable a.e., moreover the constraint $u(x) - u(y) \leq d_L(x, y)$ can be expessed in differential form as follows. Defining the dual norm $L^*(x, .)$ of $L(x, .)$:

$$L^*(x, p) := \sup\{p \cdot v \ : \ L(x, v) \leq 1\},$$

one can express the fact that u is 1-Lipschitz for d_L by the following pointwise constraint on ∇u

$$L^*(x, \nabla u(x)) \leq 1 \text{ for a.e. } x \in \overline{\Omega} \tag{2.2}$$

i.e.

$$\sigma \cdot \nabla u(x) \leq L(x, \sigma), \ \forall \sigma \in \mathbb{R}^d.$$

Thus (2.1) can be rewritten in sup-inf form as

$$\sup_{u \in W^{1,\infty}} \inf_{\sigma \in L^1(\Omega, \mathbb{R}^d)} \langle u, f\rangle + \int_\Omega L(x, \sigma(x))\mathrm{d}x - \int_\Omega \nabla u(x) \cdot \sigma(x)\mathrm{d}x. \tag{2.3}$$

Switching the infimum and the supremum above, we obtain another dual formulation of (2.1):

$$\inf_{\sigma \in L^1(\Omega, \mathbb{R}^d)} \int_\Omega L(x, \sigma(x))\mathrm{d}x + \sup_{u \in W^{1,\infty}} \langle u, f\rangle - \int_\Omega \nabla u(x) \cdot \sigma(x)\mathrm{d}x$$

observing that the supremum with respect to u is 0 if $-\mathrm{div}(\sigma) = f = f^+ - f^-$ and $\sigma \cdot \nu = 0$ on $\partial\Omega$ in the weak sense i.e.

$$\int_{\overline{\Omega}} \nabla u(x) \cdot \sigma(x)\mathrm{d}x = \langle u, f\rangle, \ \forall u \in C^1(\overline{\Omega})$$

and $+\infty$ otherwise, we obtain the following minimal flow problem dual formulation of (2.1):

$$\inf_{\sigma\in L^1(\Omega,\mathbb{R}^d)} \Big\{ \int_\Omega L(x,\sigma(x))\mathrm{d}x \ : \ -\mathrm{div}(\sigma) = f \Big\}. \tag{2.4}$$

Minimal flow formulations for transport problems were first introduced in the 1950's by Beckmann in an economic context [3], the connection with Monge's problem was realized much later (see in particular [9] and [10]). It is obvious that (2.1) possesses solutions. Standard convex duality also implies that there is no duality gap and that

$$\sup(2.1) = \inf(1.1) = \inf(2.4). \tag{2.5}$$

It is however not clear in general that (2.4) possesses L^1 solutions. In the spatially homogeneous case where $L(x,v)$ is the euclidean norm (or more generally some smooth and uniformly convex norm), $|\sigma|$ is called the transport density and there are important and involved L^1 regularity results for the transport density under suitable assumptions on $f^\pm$ due to Feldman and McCann [10], De Pascale, Evans and Pratelli [6], De Pascale and Pratelli [7] and Santambrogio [16]. We are not aware of extensions to the Finsler case yet. Since the cost in (2.4) is convex and homogeneous of degree one, (2.4) can be relaxed to vector-valued measures which amounts to replace (2.4) by:

$$\inf_{\sigma\in\mathcal{M}(\Omega,\mathbb{R}^d)} \Big\{ \int_\Omega L\Big(x, \frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}(x)\Big)\mathrm{d}|\sigma|(x) \ : \ -\mathrm{div}(\sigma) = f \Big\}. \tag{2.6}$$

where $|\sigma|$ is the total variation measure of the vector-valued measure σ and $\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}$ is the density of σ with respect to $|\sigma|$. It is then obvious by (1.3) and Banach-Alaoglu Theorem that the relaxed problem (2.6) admits solutions. To sum up, we have the following duality and attainment relations

$$\mathrm{MK}(L,f) := \min(1.1) = \max(2.1) = \inf(2.4) = \min(2.6) \tag{2.7}$$

where we have denoted $\mathrm{MK}(L,f)$ the common value of (1.1), (2.1) and (2.4).

2.2 Relations between the three problems

We now discuss in a slightly formal way the relationships between the three problems (1.1), (2.1) and (2.4). For further use, let us denote by $B(x)$ and $B^*(x)$ respectively the unit ball for $L(x,.)$ and $L^*(x,.)$:

$$B(x) := \{\sigma\in\mathbb{R}^d,\ L(x,\sigma)\le 1\},\ B^*(x) := \{q\in\mathbb{R}^d,\ :\ L^*(x,q)\le 1\}$$

and recall

$$L(x,\sigma)L^*(x,q) \geq \sigma \cdot q,\ L(x,\sigma) = \sup_{q\in B^*(x)} q\cdot\sigma,\ L^*(x,q) = \sup_{\sigma\in B(x)} q\cdot\sigma. \quad (2.8)$$

Recalling that if C is a closed convex subset of $\mathbb{R}^d$ and $z \in C$ the normal cone of C at z, $N_C(z)$ is by definition:

$$N_C(z) := \{\xi \in \mathbb{R}^d \ :\ \xi\cdot z \geq \xi \cdot y,\ \forall y \in C\},$$

it is easy to see that, if non zero vectors σ and q satisfy $L(x,\sigma)L^*(x,q) = q\cdot\sigma$ this means that

$$q \in N_{B(x)}\Big(\frac{\sigma}{L(x,\sigma)}\Big) \quad (2.9)$$

or equivalently

$$\sigma \in N_{B^*(x)}\Big(\frac{q}{L^*(x,q)}\Big). \quad (2.10)$$

In the case where $B(x)$ or $B^*(x)$ is smooth the normal cones at a point of $\partial B(x)$ or $\partial B^*(x)$ are simply the half line generated by the normal vectors (Gauss maps) and thus the previous relations give an unambiguous information on the relation between the direction of q and σ.

Any optimal plan π for (1.1) is related to any optimal potential u for (2.1) by the complementary slackness condition:

$$u(y) - u(x) = d_L(x,y),\ \forall (x,y) \in \mathrm{spt}(\pi). \quad (2.11)$$

Let then $(x,y) \in \mathrm{spt}(\pi)$ and let $t \in [0,1] \mapsto \gamma_{x,y}(t)$ be a geodesic between x and y, it is easy to deduce from (2.11) and the fact that u is 1-d_L Lipschitz that one also has for every (s,t) such that $0 \leq s \leq t \leq 1$:

$$u(\gamma_{x,y}(t)) - u(\gamma_{x,y}(s)) = d_L(\gamma_{x,y}(t), \gamma_{x,y}(s)) = (t-s)d_L(x,y). \quad (2.12)$$

In other words, u grows at the maximal rate allowed by the Lipschitz constraint on the geodesic $\gamma_{x,y}$. If u was smooth we could further write:

$$u(y) - u(x) = \int_0^1 \nabla u(\gamma_{x,y}(s)) \cdot \dot\gamma_{x,y}(s)\mathrm{d}s$$

and then

$$\int_0^1 \nabla u(\gamma_{x,y}(s)) \cdot \dot\gamma_{x,y}(s)\mathrm{d}s = d_L(x,y) = \int_0^1 L(\gamma_{x,y}(s), \dot\gamma_{x,y}(s))\mathrm{d}s.$$

but since $L^*(\gamma_{x,y}(s), \nabla u(\gamma_{x,y}(s))) \leq 1$, we pointwise have $\nabla u(\gamma_{x,y}(s)) \cdot \dot{\gamma}_{x,y}(s) \leq L(\gamma_{x,y}(s), \dot{\gamma}_{x,y}(s))$ so that

$$\nabla u(\gamma_{x,y}(s)) \cdot \dot{\gamma}_{x,y}(s) = L(\gamma_{x,y}(s), \dot{\gamma}_{x,y}(s)), \ \forall s \in [0,1], \tag{2.13}$$

which also gives

$$L^*(\gamma_{x,y}(s), \nabla u(\gamma_{x,y}(s))) = 1, \ \forall s \in [0,1]. \tag{2.14}$$

This expresses in a local way the fact that the Lipschitz constraint on u is binding on geodesics (this is again formal). Note that (2.14) gives a precise relation between $\nabla u(x)$ and the direction of geodesics passing through x: they are tangent to a vector in the normal cone $N_{B^*(x)}(\nabla u(x))$.

Now if σ solves (2.4) and u is a solution of (2.1), then complementary slackness takes the form

$$L(x, \sigma(x)) = \sigma(x) \cdot \nabla u(x) \text{ a.e.} \tag{2.15}$$

hence

$$\sigma(x) \neq 0 \Rightarrow L^*(x, \nabla u(x)) = 1, \tag{2.16}$$

which again expresses that the Lipschitz constraint is binding on the support of the transport density. The direction of optimal flows and gradients of Kantorovich potentials are therefore related by the duality relations

$$\sigma(x) \neq 0 \Rightarrow \sigma(x) \in N_{B^*(x)}(\nabla u(x)), \ \nabla u(x) \in N_{B(x)}\Big(\frac{\sigma(x)}{L(x, \sigma(x))}\Big). \tag{2.17}$$

It remains to investigate the relations between optimal plans and optimal flow fields. The following (heuristic) construction is well-known (see for instance [1]) in the euclidean setting: let π be an optimal plan i.e. a solution for (1.1). For every $(x,y) \in \mathrm{spt}(\pi)$ let $t \in [0,1] \mapsto \gamma_{x,y}(t)$ be a geodesic between x and y and define the vector-valued measure σ_π by

$$\int_{\overline{\Omega}} F \mathrm{d}\sigma_\pi := \int_{\overline{\Omega} \times \overline{\Omega}} \Big(\int_0^1 F(\gamma_{x,y}(s)) \cdot \dot{\gamma}_{x,y}(s) \mathrm{d}s \Big) \mathrm{d}\pi(x,y) \tag{2.18}$$

for every $F \in C(\overline{\Omega}, \mathbb{R}^d)$. Let $\phi \in C^1(\overline{\Omega})$, using the fact that $\pi \in \Pi(f^-, f^+)$ then gives

$$\int_{\overline{\Omega}} \nabla \phi \mathrm{d}\sigma_\pi = \int_{\overline{\Omega} \times \overline{\Omega}} (\phi(y) - \phi(x)) \mathrm{d}\pi(x,y) = \langle \phi, f \rangle$$

i.e. $-\mathrm{div}(\sigma_\pi) = f$ so that σ_π is admissible for the minimal flow problem (2.4). To see that σ_π is actually optimal, we consider a Kantorovich potential i.e. a solution u of (2.1). Thanks to (2.7), it is enough to show that

$$\int_\Omega L(x, \sigma_\pi(x))\mathrm{d}x \leq \langle u, f\rangle.$$

On the one hand, observing that:

$$\int_\Omega L(x, \sigma_\pi(x))\mathrm{d}x = \sup\Big\{ \int_\Omega F(x)\cdot\sigma_\pi(x)\mathrm{d}x \ : \ L^*(x, F(x)) \leq 1\Big\}$$

and that if $L^*(x, F(x)) \leq 1$ then

$$\begin{aligned}\int_\Omega F(x)\cdot\sigma_\pi(x)\mathrm{d}x &= \int_{\overline{\Omega}\times\overline{\Omega}} \Big(\int_0^1 F(\gamma_{x,y}(s))\cdot\dot\gamma_{x,y}(s)\mathrm{d}s\Big)\mathrm{d}\pi(x,y)\\ &\leq \int_{\overline{\Omega}\times\overline{\Omega}} \Big(\int_0^1 L(\gamma_{x,y}(s), \dot\gamma_{x,y}(s))\mathrm{d}s\Big)\mathrm{d}\pi(x,y)\end{aligned}$$

we get

$$\int_\Omega L(x, \sigma_\pi(x))\mathrm{d}x \leq \int_{\overline{\Omega}\times\overline{\Omega}} \Big(\int_0^1 L(\gamma_{x,y}(s), \dot\gamma_{x,y}(s))\mathrm{d}s\Big)\mathrm{d}\pi(x,y).$$

On the other hand, thanks to the complementary slackness condition (2.13) and $-\mathrm{div}(\sigma_\pi) = f$, we have

$$\begin{aligned}\langle u, f\rangle = \int_\Omega \nabla u\cdot\sigma_\pi &= \int_{\overline{\Omega}\times\overline{\Omega}} \Big(\int_0^1 \nabla u(\gamma_{x,y}(s))\cdot\dot\gamma_{x,y}(s)\mathrm{d}s\Big)\mathrm{d}\pi(x,y)\\ &= \int_{\overline{\Omega}\times\overline{\Omega}} \Big(\int_0^1 L(\gamma_{x,y}(s)), \dot\gamma_{x,y}(s))\mathrm{d}s\Big)\mathrm{d}\pi(x,y).\end{aligned}$$

This proves the optimality of σ_π.

2.3 Lagrangian and saddle-point

Rewrite (2.1) as

$$\inf_{u,q}\Big\{ -\langle u, f\rangle + G(q) \ : \ q = \nabla u \text{ a.e.}\Big\}$$

where

$$G(q) := \int_\Omega \mathcal{G}(x, q(x))\mathrm{d}x$$

and

$$\mathcal{G}(x,q) := \begin{cases} 0, \text{ if } L^*(x,q) \leq 1 \\ +\infty, \text{otherwise} \end{cases}$$

and then rewrite (2.1)-(2.4) as the saddle point problem

$$\inf_{u,q} \sup_{\sigma} \mathcal{L}(u,q,\sigma)$$

where the Lagrangian $\mathcal{L}$ is defined by

$$\begin{aligned} \mathcal{L}(u,q,\sigma) := -\langle u, f\rangle + \int_\Omega \mathcal{G}(x,q(x))\mathrm{d}x \\ + \int_\Omega \sigma(x)\cdot(\nabla u(x) - q(x))\mathrm{d}x. \end{aligned}$$

For $r > 0$, let us also introduce the augmented Lagrangian

$$\begin{aligned} \mathcal{L}_r(u,q,\sigma) := -\langle u, f\rangle + \int_\Omega \mathcal{G}(x,q(x))\mathrm{d}x \\ + \int_\Omega \sigma(x)\cdot(\nabla u(x) - q(x))\mathrm{d}x + \frac{r}{2}\int_\Omega |\nabla u - q|^2. \end{aligned}.$$

Recall that $\mathcal{L}$ and $\mathcal{L}_r$ have the same saddle-points ([11], [12]). Note that in both $\mathcal{L}$ and $\mathcal{L}_r$ we multiply the L^∞ vector field ∇u by σ, which a priori only makes sense only if σ is L^1. Existence of saddle-points is therefore not guaranteed unless there is an L^1 solution to (2.4). However at the level of the discretized problems (see next section), there is no such regularity issue, there exists saddle-points for the discretized Lagrangian and finding such saddle-points is equivalent to solving (2.1) and (2.4) simultaneously.

3 Discretization and algorithm

3.1 Discretization

We now consider suitable approximations of our problems by finite-dimensional (convex) ones using finite elements. In these finite dimensional-approximations existence of saddle-points is not an issue anymore. More precisely, consider a family of regular triangulations $\mathcal{T}_h$ of the domain (which we now assume to be two-dimensional) indexed by the typical meshsize h (i.e. the diameter of each $T \in \mathcal{T}_h$ is less than Ch for some positive constant C), let $E_h \subset W^{1,\infty}(\Omega)$ be the corresponding finite-dimensional space of Lagrange P_1 (piecewise linear) finite elements of order 1 (a similar analysis can be done for higher order

finite-elements) whose generic elements are denoted u_h. Slightly abusing notations, we shall consider u_h both as a finite-dimensional vector and a Lipschitz, piecewise linear function defined on the whole domain, the gradient of u_h has piecewise constant components, it is still denoted ∇u_h. We further assume that the mesh is regular in the sense that the Lagrange interpolate map $I_h : W^{1,\infty}(\Omega) \to E_h$ satisfies

$$\lim_{h\to 0} \|\nabla v - \nabla(I_h(v))\|_{L^\infty} \to 0, \ \forall v \in C^1(\overline{\Omega}). \tag{3.1}$$

We also approximate the linear form f by $f_h \in (E_h)^* \simeq E_h$ (again with $\langle f_h, 1\rangle = 0$) in such a way that f_h weakly converges to f in the sense of measures as $h \to 0$.

We then consider the approximation of (2.1):

$$\sup_{u_h \in K_h} \langle f_h, u_h\rangle \tag{3.2}$$

where K_h is the convex subset of E_h consisting of all u_h's in E_h such that for every $T \in \mathcal{T}_h$ one has

$$L^*(x_T, \nabla u_h|_T) \leq 1, \tag{3.3}$$

where x_T is a given point in T (for instance its center of mass or one of its vertices). To prove that this is a consistent approximation of Kantorovich problem (2.1), it is useful to observe first that smooth functions are dense in the admissible set for (2.1):

Lemma 3.1. *Let u be a 1-Lipschitz for d_L function, then there exists a sequence u_n of $C^\infty(\mathbb{R}^d)$, 1-Lipchitz for d_L functions converging uniformly on $\overline{\Omega}$ to u. In particular this implies that*

$$\max(2.1) = \sup\Big\{\langle u, f\rangle : \ u \text{ is 1-Lipschitz for } d_L \text{ and } C^\infty(\mathbb{R}^d)\Big\}. \tag{3.4}$$

Proof. First extend u on the whole of $\mathbb{R}^d$ by setting

$$u(x) = \inf_{y\in\overline{\Omega}}\{u(y) + d_L(x,y)\}.$$

Consider a standard mollifying kernel $\rho_\varepsilon(x) := \varepsilon^{-d}\rho(\varepsilon^{-1}x)$ with $\rho \in C_c^\infty(\mathbb{R}^d)$, $\rho \geq 0$, $\rho(x) = 0$ for $|x| \geq 1$ and $\int_{\mathbb{R}^d} \rho = 1$. Then for $\varepsilon > 0$ and $\delta > 0$, define then the smooth function $u_{\varepsilon,\delta} := \frac{1}{1+\delta}\rho_\varepsilon \star u$. We then have for every

$x \in \overline{\Omega}$, using the convexity of $L^*(x,.)$, Jensen's inequality and the fact that $L^*(y, \nabla u(y)) \leq 1$ a.e.

$$\begin{aligned} L^*(x, \nabla u_{\varepsilon,\delta}(x)) &\leq \frac{1}{1+\delta} \int_{\mathbb{R}^d} \rho_\varepsilon(x-y) L^*(x, \nabla u(y)) \mathrm{d}y \\ &= \frac{1}{1+\delta} \int_{\mathbb{R}^d} \rho_\varepsilon(x-y) L^*(y, \nabla u(y)) \mathrm{d}y \\ &+ \frac{1}{1+\delta} \int_{\mathbb{R}^d} \rho_\varepsilon(x-y)(L^*(x, \nabla u(y)) - L^*(y, \nabla u(y))) \mathrm{d}y \\ &\leq \frac{1+\omega(\varepsilon)}{1+\delta} \end{aligned}$$

where

$$\omega(\varepsilon) := \sup |\{L^*(x,q) - L^*(y,q)|,\ x \in \overline{\Omega},\ |x-y| \leq \varepsilon,\ |q| \leq \|\nabla u\|_{L^\infty}\}.$$

Thus $u_{\varepsilon,\delta}$ is 1-d_L Lipschitz as soon as $\omega(\varepsilon) \leq \delta$, this clearly proves the desired result since L^* is continuous.

□

One easily deduces the following convergence result:

Proposition 3.2. *Let u_h be a solution of (3.2) normalized so as to have zero mean, then for some vanishing sequence of meshsizes $h_n \to 0$ as $n \to \infty$, u_{h_n} converges uniformly to some Kantorovich potential u i.e. some solution of (2.1).*

Proof. Thanks to (1.3), u_h is uniformly Lipschitz, since it has nonzero mean, thanks to Ascoli's theorem, for some vanishing sequence of meshsizes, it converges in $C(\overline{\Omega})$ to some Lipschitz function u. Thanks to Banach-Alaoglu's Theorem, we may also assume that ∇u_h converges weakly $*$ in L^∞ to ∇u. To check that u is 1-d_L Lipschitz, it is enough to show that for every $\sigma \in C(\overline{\Omega}, \mathbb{R}^2)$ one has

$$\int_\Omega \sigma \cdot \nabla u \leq \int_\Omega L(x, \sigma(x)) \mathrm{d}x. \tag{3.5}$$

Since $u_h \in K_h$ we have for every $T \in \mathcal{T}_h$, $\nabla u_h|_T \cdot \sigma(x_T) \leq L(x_T, \sigma(x_T))$, multiplying by the measure of T, summing over all triangles of $\mathcal{T}_h$ and letting $h \to 0$ gives (3.5). It remains to prove that u solves (2.1), which thanks to Lemma 3.1 amounts to show that $\langle f, u \rangle \geq \langle f, v \rangle$ for every smooth and 1-d_L-Lipschitz function v. Let then v be such a smooth and 1-d_L-Lipschitz function, for every $T \in \mathcal{T}_h$, we have

$$\begin{aligned} L^*(x_T, \nabla I_h(v)(x_T)) &= L^*(x_T, \nabla v(x_T)) + L^*(x_T, \nabla I_h(v)(x_T)) - L^*(x_T, \nabla v(x_T)) \\ &\leq 1 + \omega_h \end{aligned}$$

where

$$\omega_h := \sup_{T\in\mathcal{T}_h} |L^*(x_T, \nabla I_h(v)(x_T)) - L^*(x_T, \nabla v(x_T))| \leq C\|\nabla v - \nabla(I_h(v))\|_{L^\infty}$$

tends to 0 as $h \to 0$ thanks to (3.1). Then defining $v_h := (1+\omega_h)^{-1} I_h(v)$, we have $v_h \in K_h$ and v_h converges uniformly to v as $h \to 0$, passing to the limit in $\langle f_h, u_h\rangle \geq \langle f_h, v_h\rangle$, we can conclude that u is a Kantorovich potential.

□

3.2 Augmented Lagrangian algorithm

From now on, we drop the dependence in h in the approximation parameter and slightly abusing notations, we use the same notations as in the continuous framework, eventhough in what follows we actually consider the discretization of the augmented Lagrangian $\mathcal{L}_r$. Existence of a saddle-point is not an issue at the level of the finite-dimensional approximation and convergence of the augmented Lagrangian algorithm recalled below is well-known (see Eckstein and Bertsekas [8]). The augmented Lagrangian algorithm ALG2 splitting scheme, consists, starting from $(u^0, q^0, \sigma^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ to generate inductively a sequence (u^k, q^k, σ^k) as follows (abusing notations we still denote by ∇ the discretization of the gradient):

- **Step 1:** minimization with respect to u:

$$u^{k+1} := \text{argmin}_{u\in\mathbb{R}^n} \Big\{ -\langle u, f\rangle + \sigma^k \cdot \nabla u + \frac{r}{2}|\nabla u - q^k|^2 \Big\}, \tag{3.6}$$

- **Step 2:** minimization with respect to q:

$$q^{k+1} := \text{argmin}_{q\in\mathbb{R}^m} \Big\{ G(q) - \sigma^k \cdot q + \frac{r}{2}|\nabla u^{k+1} - q|^2 \Big\}, \tag{3.7}$$

- **Step 3:** update the multiplier by the gradient ascent formula

$$\sigma^{k+1} = \sigma^k + r(\nabla u^{k+1} - q^{k+1}). \tag{3.8}$$

Step 1 consists in solving a Laplace equation :

$$-r(\Delta u^{k+1} - \text{div}(q^k)) = f + \text{div}(\sigma^k) \text{ in } \Omega, \tag{3.9}$$

together with the Neumann boundary condition

$$r\frac{\partial u^{k+1}}{\partial \nu} = rq^k \cdot \nu - \sigma^k \cdot \nu \text{ on } \partial\Omega. \tag{3.10}$$

Step 2 is a pointwise projection problem

$$q^{k+1}(x) = p_{B^*(x)}\Big(\nabla u^{k+1} + \frac{\sigma^k}{r}\Big),$$

where $p_{B^*(x)}$ is the projection onto $B^*(x) := \{q \in \mathbb{R}^d \ : \ L^*(x,q) \leq 1\}$ the unit ball for $L^*(x,.)$.

3.3 Examples

We now give some details on how to perform the projection step 2 in practice. For the sake of simplicity we drop the dependence of L and L^* in x.

The Riemannian case

In the Riemannian case $L(v) = (Av \cdot v)^{\frac{1}{2}}$ for some symmetric positive definite matrix A. Up to diagonalizing A, there is no loss of generality in assuming that $L(v) = (\sum_{i=1}^d \lambda_i v_i^2)^{\frac{1}{2}}$ with $\lambda_i > 0$ the eigenvalues of A. The dual norm L^* is then given by $L^*(q) = (\sum_{i=1}^d \lambda_i^{-1} q_i^2)^{\frac{1}{2}}$. The projection p_{B^*} onto $B^* := \{q \in \mathbb{R}^d \ : \ L^*(q) \leq 1\}$ is almost explicit:

$$p_B^*(\overline{q}) = \begin{cases} \overline{q}, \text{ if } \overline{q} \in B^*, \\ \Big(\frac{\lambda_1 \overline{q}_1}{\lambda_1 + \alpha}, \cdots \frac{\lambda_d \overline{q}_d}{\lambda_d + \alpha}\Big) \text{ with } \alpha \text{ the positive root of (3.11) otherwise} \end{cases}$$

where the nonlinear equation to be solved by α reads

$$1 = \sum_{i=1}^d \frac{\lambda_i \overline{q}_i^2}{(\lambda_i + \alpha)^2}. \tag{3.11}$$

This single equation is monotone in α and can be efficiently solved by Newton's method.

The case where $L(x,.)$ is defined by finitely directions

The second case we have in mind is the polyhedral case where L is defined by finitely many directions. More precisely (and again this is for a fixed x), we are given a collection of unit vectors $v_1, \cdots, v_k$ which we complete by $v_{k+1} = -v_1, \cdots, v_{2k} = -v_k$ and such that 0 belongs to the interior of the (symmetric) convex polytope $\text{co}(\{v_j, \ j = 1, \cdots, 2k\})$. We are also given positive reals $(\xi_j)_{j=1,\cdots,2k}$ with $\xi_{j+k} = \xi_j$ for $j \in \{1, \cdots, k\}$ and then consider the crystalline norm

$$L(v) := \inf\Big\{\sum_{j=1}^{2k} \xi_j \alpha_j \ : \ \alpha_j \geq 0, \ \sum_{j=1}^{2k} \alpha_j v_j = v\Big\}.$$

It is immediate to see that L is the gauge of the symmetric convex polytope $B := \mathrm{co}(\{\xi_j^{-1} v_j,\ j = 1, \cdots, 2k\})$ which is then its unit ball. The dual norm L^* is then explicitly given by

$$L^*(q) := \max_{j=1,\cdots,k} \xi_j^{-1} |q \cdot v_j|, \tag{3.12}$$

its dual unit ball B^* is then defined by the inequalities $|q \cdot v_j| \leq \xi_j$ for $j = 1, \cdots, k$. In dimension two, the projection onto B^* can be easily performed as follows. First, compute the vertices and sides of B^* (note that the latter have one of the vectors v_j as normal, so these computations can be done in an automatic way) so as to be able to represent $B^* = \mathrm{co}(\{S_i,\ \ i = 1, \cdots, 2l\})$ where $S_1, \cdots, S_l$ are the successive vertices of B^* and denote by ν_i the unit exterior normal to the side $[S_i, S_{i+1}]$. Now if $\overline{q}$ is a generic vector of the plane belonging to the complement of B^* (otherwise its projection is $\overline{q}$), then $\overline{q}$ belongs either to one half strip $[S_i, S_{i+1}] + \mathbb{R}_+ \nu_i$ and in this case its projection on B^* coincides with its projection on the line $S_i + \nu_i^\perp$ or it belongs to one of the sectors $S_i + \mathbb{R}_+ \nu_{i-1} + \mathbb{R}_+ \nu_i$ and in this case the projection of $\overline{q}$ is the vertex S_i. We illustrate these considerations in Figure 1, by the following example with $k = 4$,

$$v_j = \left(\cos\left(\frac{(j-1)\pi}{k} \right), \sin\left(\frac{(j-1)\pi}{k} \right) \right), \xi_1 = 2.5, \xi_2 = 2, \xi_3 = 1.5,\ \xi_4 = 3.$$

In fact, the vector v_4 is useless since ξ_4 is very large with respect to the other ones. So the ball B^* is only defined by the inequalities $|q \cdot v_j| \leq \xi_j$ for $j = 1, \ldots, 3$. The point $\overline{q}_1$ is in the half strip $[S_6, S_1] + \mathbb{R}_+ v_1$ so that its projection is on the segment $[S_6, S_1]$. The point $\overline{q}_2$ belongs to the sector $S_6 + \mathbb{R}_+ v_6 + \mathbb{R}_+ v_1$ so that its projection is S_6.

4 Results

We use the software FreeFem++ (see [14]) to implement the ALG2 scheme described above. The Lagrangian finite elements and notations used in Subsection (3.2) are taken here. We use P_2 FE for u_h and P_1 for (q_h, σ_h) (approximation is better and convergence is faster than with P_0 and P_1). As emphasized in the previous subsection, the first step and the third one are always the same. Only the projection step 2 changes according to the geometry of the Finsler metric. For our numerical simulations Ω is a $2D$ square $(x = (x_1, x_2) \in [0,1]^2)$ and we test with different f :

$$f_1^- := e^{-40*((x_1-0.75)^2+(x_2-0.25)^2} \text{ and } f_1^+ := e^{-40*((x_1-0.25)^2+(x_2-0.65)^2)},$$
$$f_2^- := e^{-40*((x_1-0.5)^2+(x_2-0.15)^2)} \text{ and } f_2^+ := e^{-40*((x_1-0.5)^2+(x_2-0.75)^2)}.$$

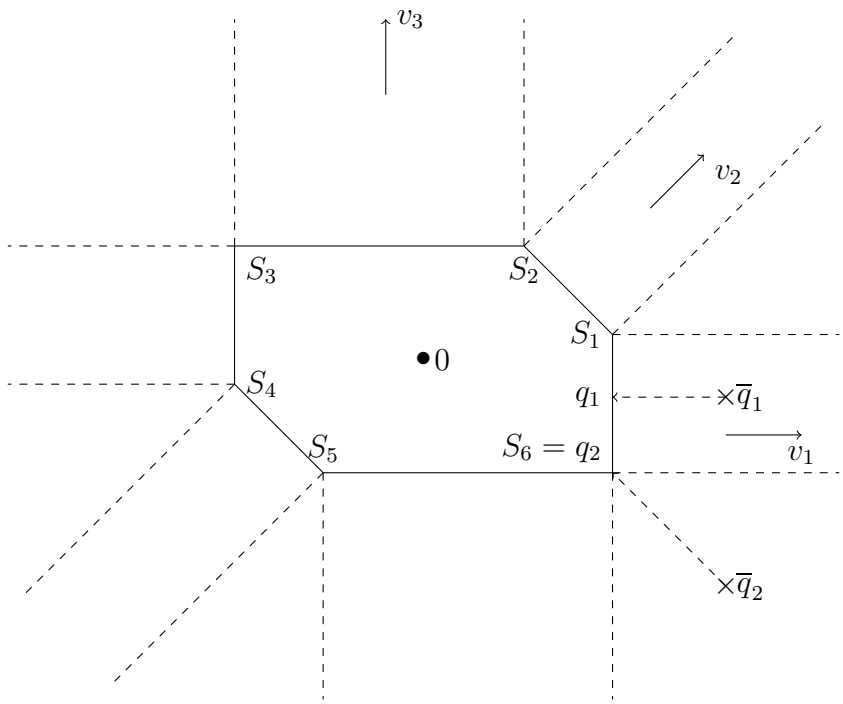

Figure 1: Projecting on the polyhedron B^*.

In the third case, we take f_3^- a constant density and f_3^+ is the sum of three concentrated Gaussians

$$f_3^+(x_1, x_2) = e^{-400*((x_1-0.25)^2+(x_2-0.75)^2)} + e^{-400*((x_1-0.35)^2+(x_2-0.15)^2)} + e^{-400*((x_1-0.85)^2+(x_2-0.7)^2)}.$$

In the following two subsections, in each figure, there are two images. The top one represents σ and the bottom one corresponds to the level lines of u.

4.1 Riemannian case

Here $L(x, v) = (\lambda_1(x)v_1^2 + \lambda_2(x)v_2^2)^{\frac{1}{2}}$, we take one λ_i constant and the other one non constant, that is equal to the inverse of

$$g(x_1, x_2) = 1.5 - \exp(-100 * ((x_1 - 0.5)^2 + (x_2 - 0.5)^2)).$$

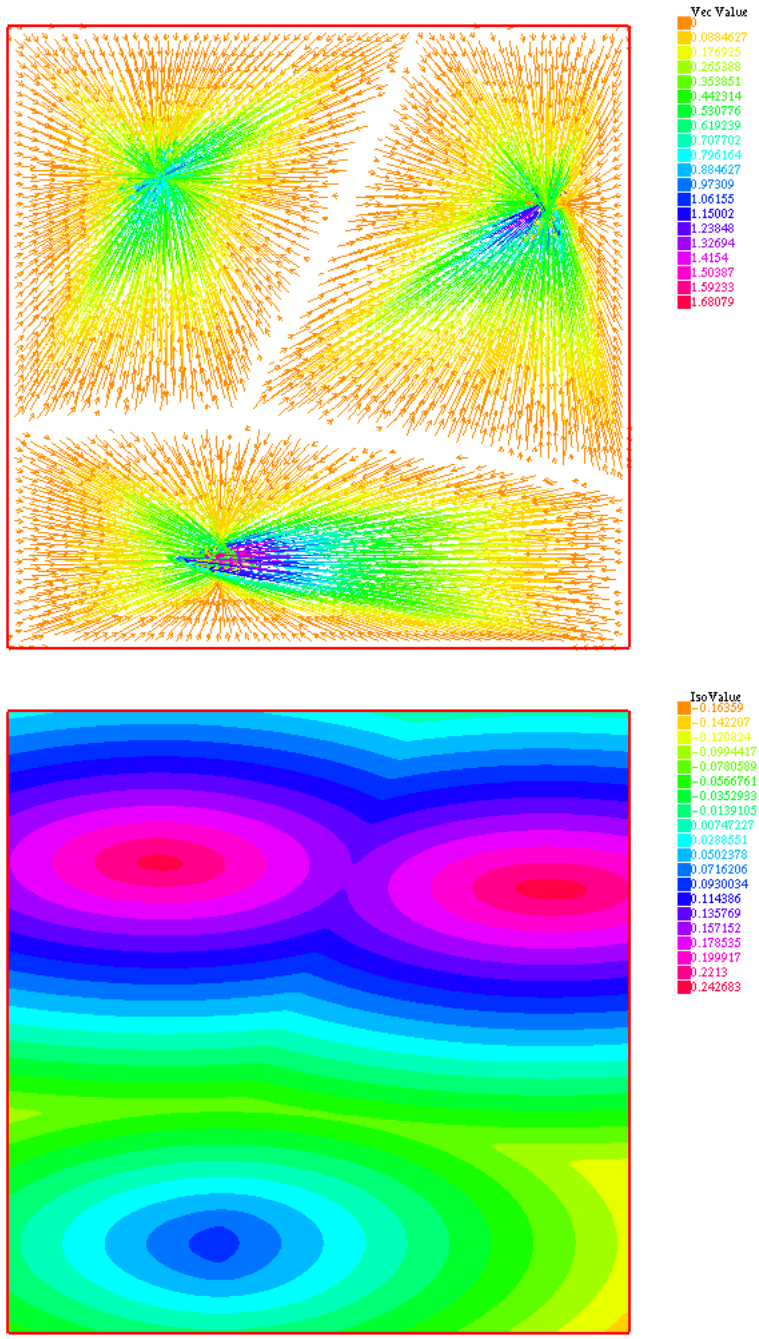

Figure 2: Test case 1: Riemannian metric with $\lambda_1 = 0.1$, $\lambda_2 = 1/g$ and $f = f_3$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

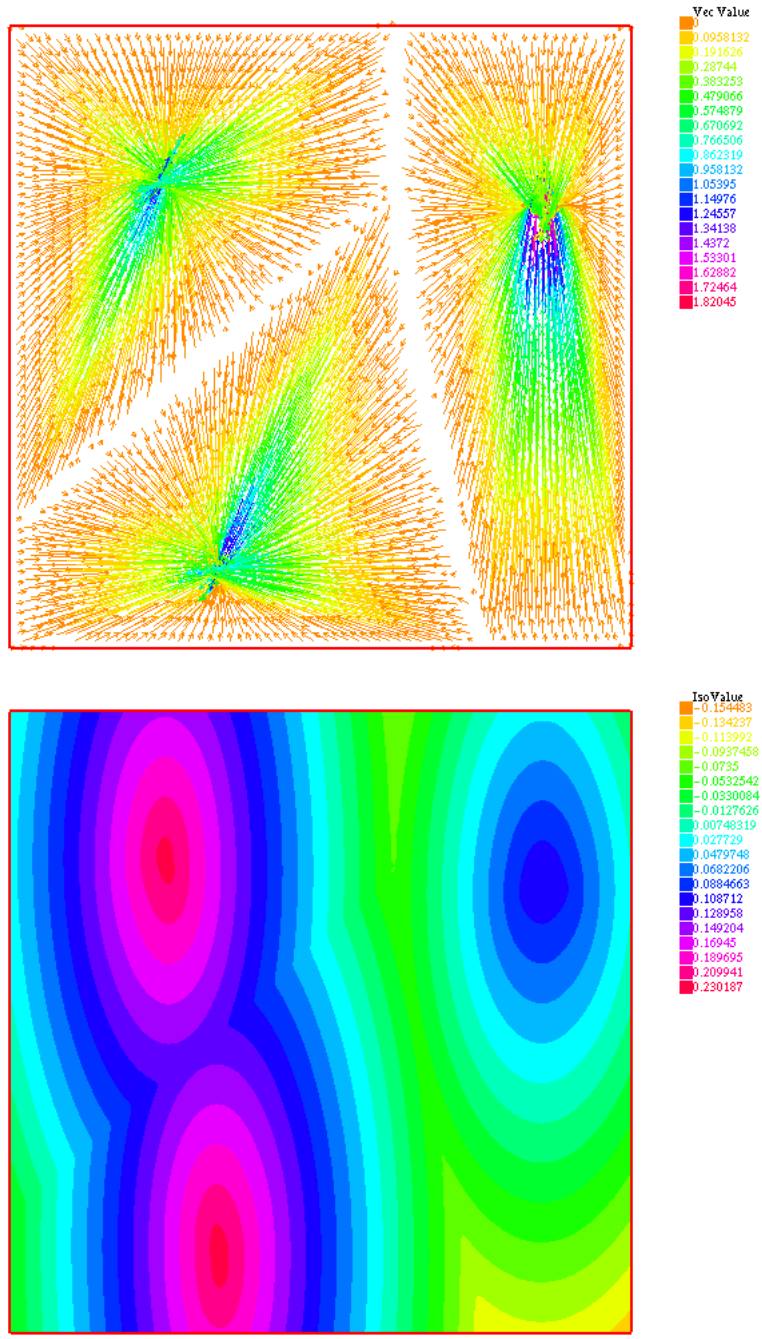

Figure 3: Test case 2: Riemannian metric with $\lambda_1 = 1/g$, $\lambda_2 = 0.1$ and $f = f_3$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

4.2 Polyhedral case

We tested the following polyhedral examples. In Figure 6 we take $k = 2$ and v_1 perpendicular to v_2, the dual unit ball B^* then is a rectangle. In all other examples, $k = 15$ and the angle between two consecutive directions is π/k. The form of B^* then depends on the chosen ξ_j's. If the ξ_j's are (almost) equal, B^* is a polyhedron with thirty edges. It is in particular the case for Figure 5, Figure 8, Figure 9 and Figure 10. In the last examples, we have $\xi_j = \cos\left(\frac{2(j-1)\pi}{k}\right) + 1.5$ and the ball B^* then only has 12 edges.

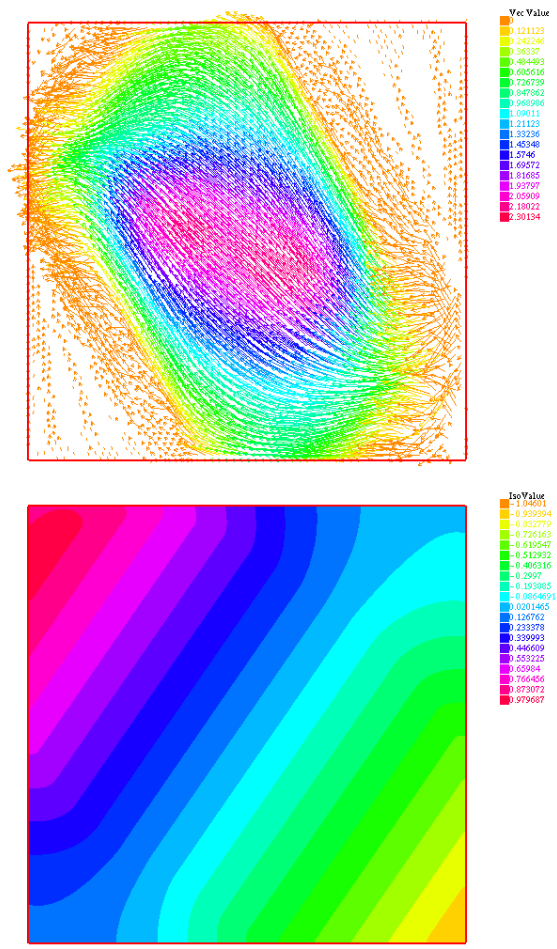

Figure 4: Test case 3: polyhedral metric, $k = 15$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k}\right), \sin\left(\frac{(j-1)\pi}{k}\right)\right)$ and $\xi_j = \cos\left(\frac{2(j-1)\pi}{k}\right) + 1.5$ for $j = 1, \ldots, k$ with $f = f_1$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

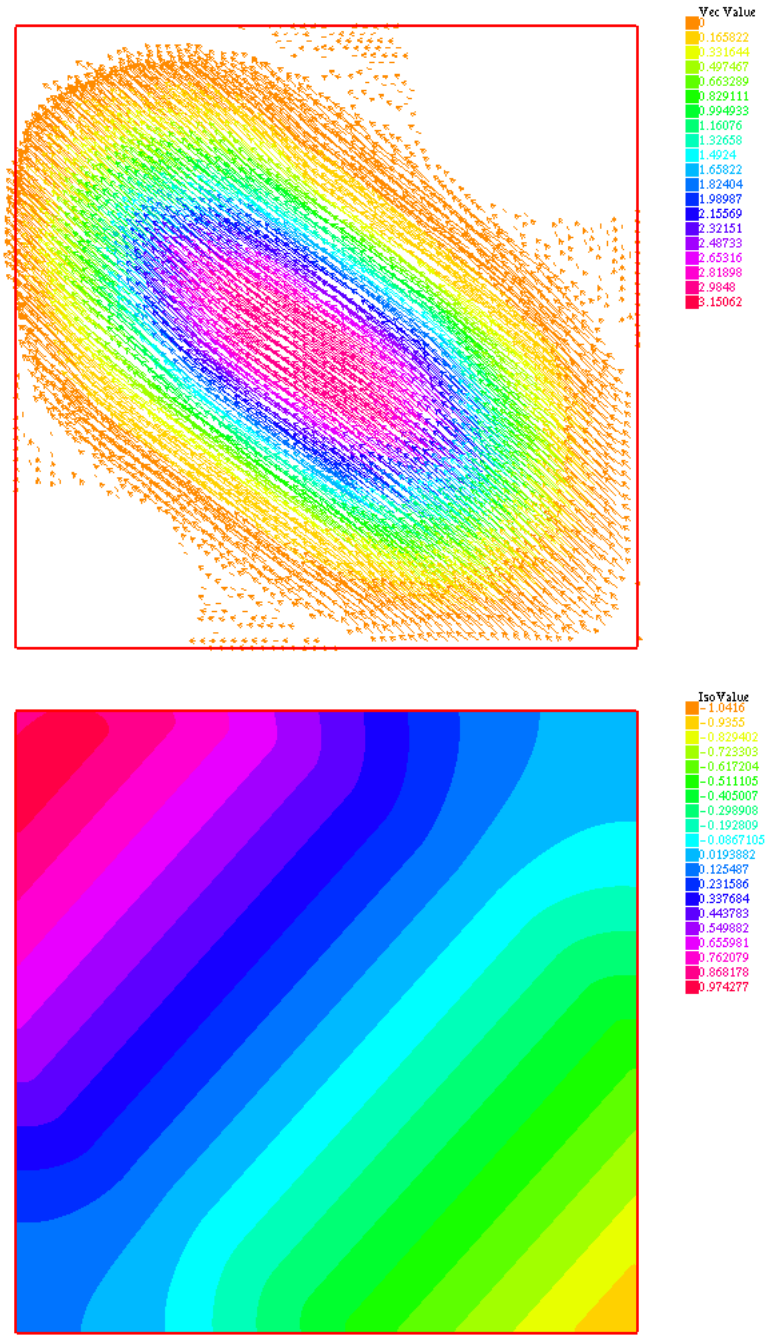

Figure 5: Test case 4: polyhedral metric, $k = 15$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k}\right), \sin\left(\frac{(j-1)\pi}{k}\right)\right)$ and $\xi_j = 1.5$ for $j = 1, \ldots, k$ with $f = f_1$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

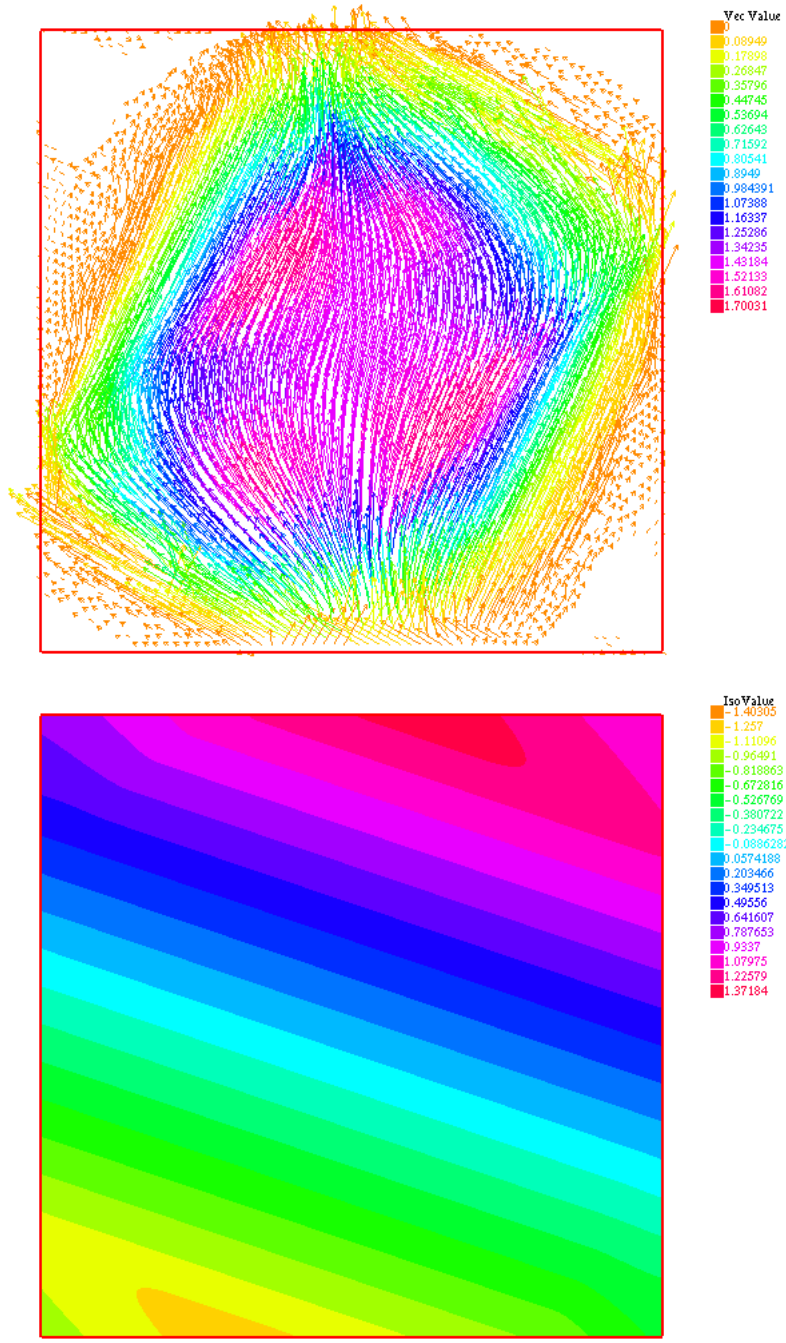

Figure 6: Test case 5: polyhedral metric, $k = 2$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k} + \frac{\pi}{3}\right), \sin\left(\frac{(j-1)\pi}{k} + \frac{\pi}{3}\right)\right)$ and $\xi_j = \cos\left(\frac{2(j-1)\pi}{k}\right) + 1.5$ for $j = 1, \ldots, k$ with $f = f_2$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

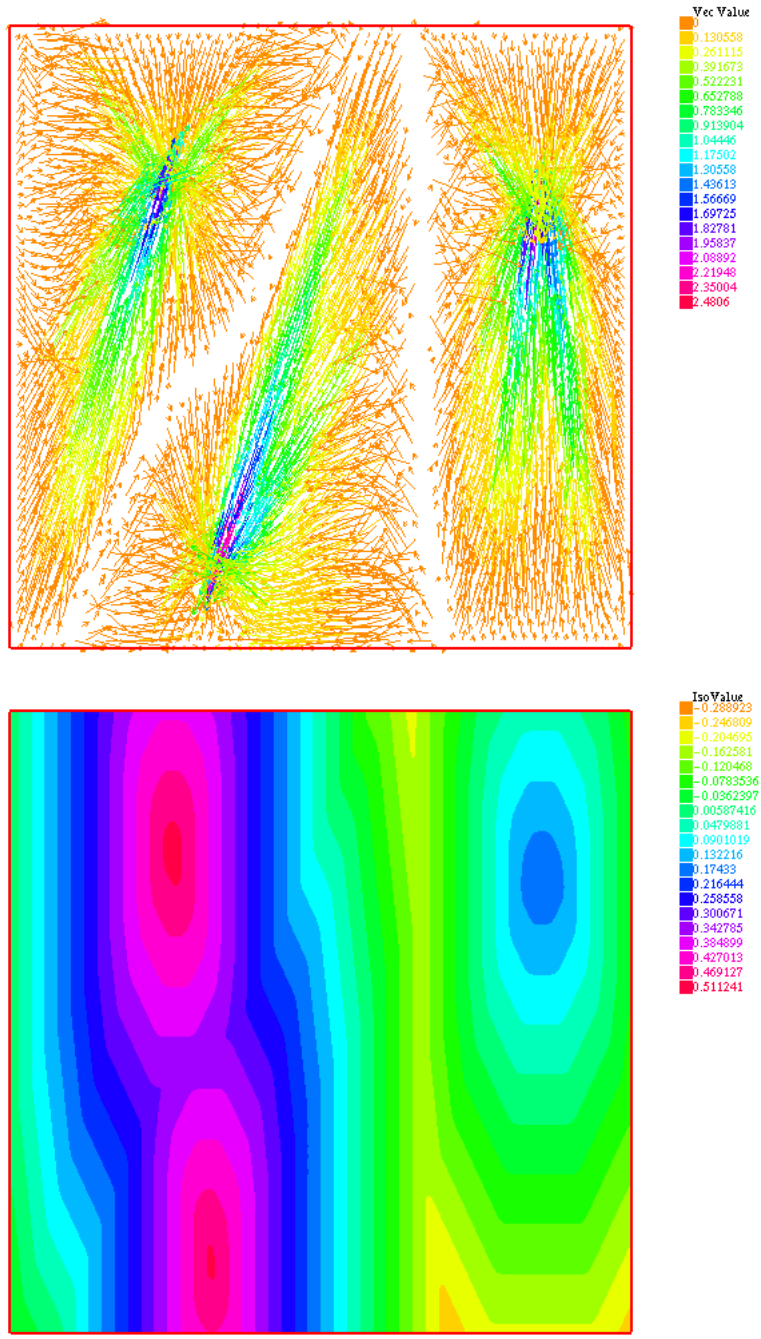

Figure 7: Test case 6: polyhedral metric, $k = 15$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k}\right), \sin\left(\frac{(j-1)\pi}{k}\right)\right)$ and $\xi_j = \cos\left(\frac{2(j-1)\pi}{k}\right) + 1.5$ for $j = 1, \ldots, k$ with $f = f_3$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

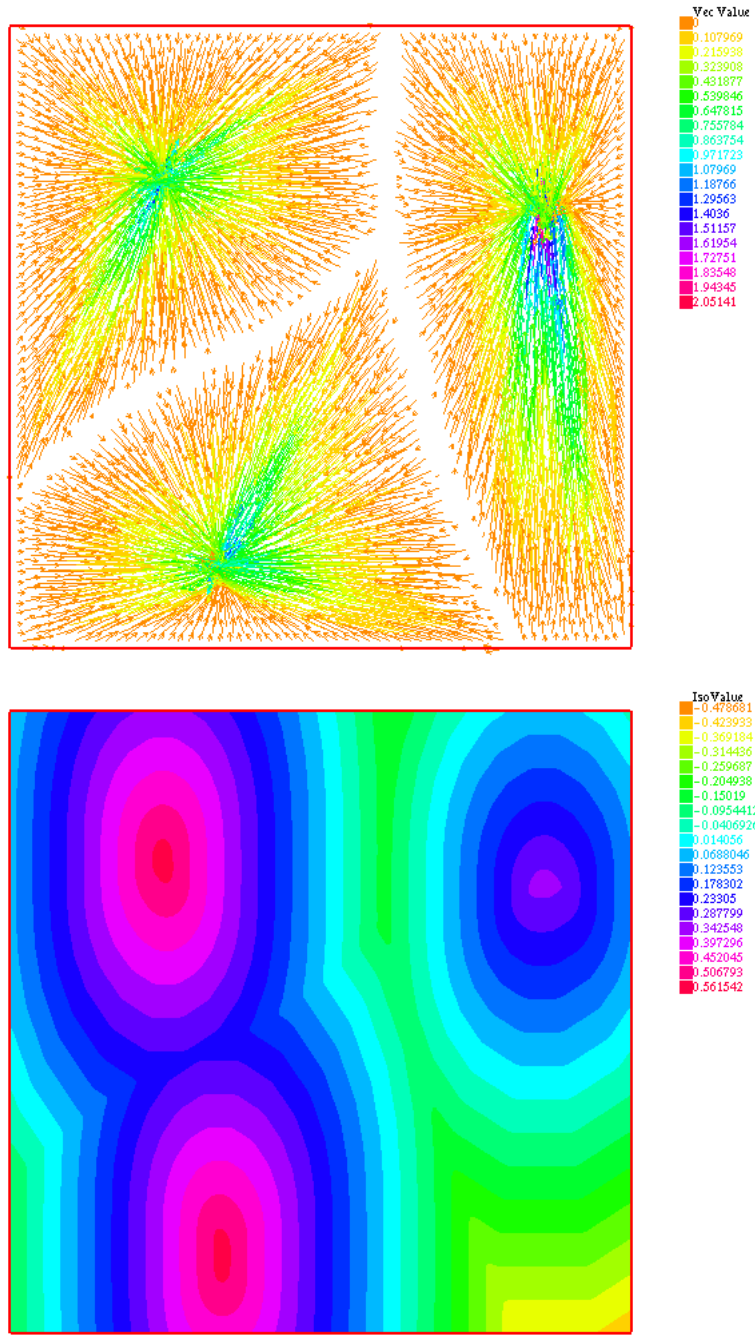

Figure 8: Test case 7: polyhedral metric, $k = 15$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k}\right), \sin\left(\frac{(j-1)\pi}{k}\right)\right)$ and $\xi_j = \frac{1}{2}\cos\left(\frac{2(j-1)\pi}{k}\right) + 1.5$ for $j = 1, \ldots, k$ with $f = f_3$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

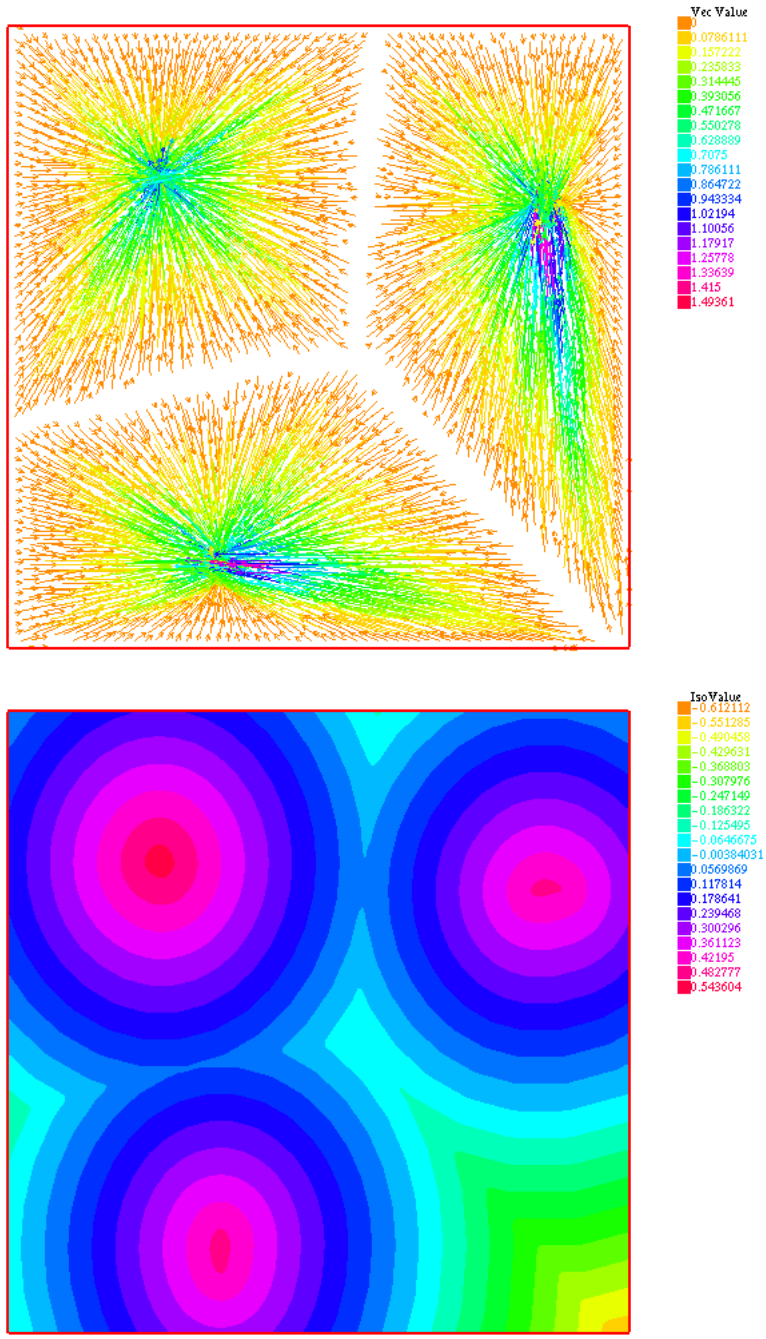

Figure 9: Test case 8: polyhedral metric, $k = 15$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k}\right), \sin\left(\frac{(j-1)\pi}{k}\right)\right)$ and $\xi_j = \frac{1}{10}\cos\left(\frac{2(j-1)\pi}{k}\right) + 1.5$ for $j = 1, \ldots, k$ with $f = f_3$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

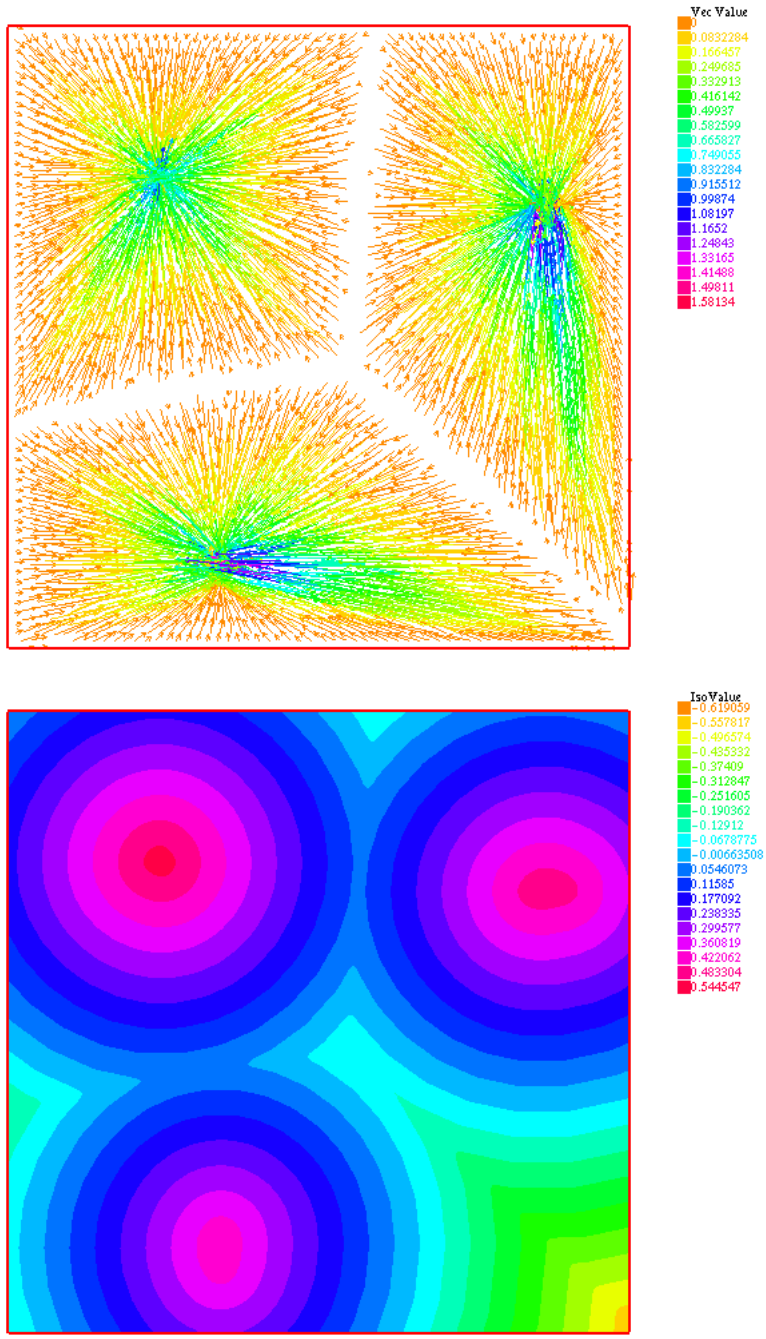

Figure 10: Test case 9: polyhedral metric, $k = 15$, $v_j = \left(\cos\left(\frac{(j-1)\pi}{k}\right), \sin\left(\frac{(j-1)\pi}{k}\right)\right)$ and $\xi_j = 1$ for $j = 1, \ldots, k$ with $f = f_3$. Top: vector fied σ, bottom: level sets of the Kantorovich potential u.

4.3 Error Criteria

To analyze the convergence of our simulations, we have considered three criteria corresponding to the optimality conditions:

$$-\mathrm{div}(\sigma) = f, \text{ in } \Omega \;\; \sigma \cdot \nu = 0 \text{ on } \partial\Omega, \tag{4.1}$$

as well as duality relation

$$L(x, \sigma) = \sigma \cdot \nabla u \tag{4.2}$$

which can be equivalently rewritten in a dual way as

$$\sigma(x) \neq 0 \Rightarrow L^*(x, \nabla u(x)) = 1. \tag{4.3}$$

We use a triangulation of the unit square with $n = 1/h$ element on each side and the following convergence criteria:

1. DIV.Error $= \left(\int_{\Omega_h} (\mathrm{div}\sigma_h^k + f)^2\right)^{1/2}$ is the L^2 error on the divergence constraint.
2. BND.Error $= \left(\int_{\partial\Omega_h} (\sigma_h^k \cdot \nu)^2\right)^{1/2}$ is the $L^2(\partial\Omega_h)$ error on the Neumann boundary condition.
3. DUAL.Error $= \left(\int_{\Omega_h} |L(\cdot, \sigma_h^k(\cdot)) - \nabla u_h^k \cdot \sigma_h^k|\right)$ for the Riemannian case. DUAL.Error $= \left(\int_{\Omega_h} |L^*(\cdot, \nabla u_h^k(\cdot)) - 1|\chi_{\{|\sigma_h^k| > \varepsilon\}}\right)$ for the polyhedral case with $\varepsilon = 10^{-2}$.

The first two criteria represent the optimality conditions for the minimization of the Lagrangian with respect to u and the third one is for the maximization with respect to σ. We do not take exactly the same criteria for both examples. Indeed, in the Riemannian case, $L(\cdot, \sigma)$ is simple to compute whereas in the polyhedral case, it is tedious in general. On the other hand, $L^*(\cdot, \nabla u(\cdot))$ has an explicit form given by (3.12). We show below the results of our numerical simulations after 400 iterations for each case.

Test case	DIV.Error	BND.Error	DUAL.Error	Time execution (seconds)
1	3.0940e-05	4.9502e-04	1.6274e-06	287
2	3.2576e-05	4.0942e-04	2.1978e-06	285
3	9.3806e-05	7.9803e-04	8.1512e-04	435
4	6.1646e-06	2.5572e-04	2.7813e-03	658
5	1.9829e-05	2.2784e-03	4.8522e-04	310
6	1.1407e-04	8.5331e-04	1.8588e-03	446
7	1.0402e-04	8.5816e-04	1.2846e-03	660
8	9.9358e-05	4.9236e-04	1.2181e-03	654
9	8.3469e-05	5.0099e-04	1.1265e-03	656

Table 1: Convergence of the finite element discretization for all test cases.

Acknowledgements: G.C. and R. H. gratefully acknowledge the hospitality of the Mathematics and Statistics Department at UVIC (Victoria, Canada), and the support from the CNRS, from the ANR, through the project ISOTACE (ANR-12- MONU-0013) and from INRIA through the *action exploratoire* MOKAPLAN. R. H. is grateful to the support from the Fondation Ledoux.

References

[1] Luigi Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Math.*, pages 1–52. Springer, Berlin, 2003.

[2] John W. Barrett and Leonid Prigozhin. A mixed formulation of the Monge-Kantorovich equations. *M2AN Math. Model. Numer. Anal.*, 41(6):1041–1060, 2007.

[3] Martin Beckmann. A continuous model of transportation. *Econometrica*, 20:643–660, 1952.

[4] 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.

[5] 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*, pages 1–26, 2015.

[6] L. De Pascale, L. C. Evans, and A. Pratelli. Integral estimates for transport densities. *Bull. London Math. Soc.*, 36(3):383–395, 2004.

[7] Luigi De Pascale and Aldo Pratelli. Sharp summability for Monge transport density via interpolation. *ESAIM Control Optim. Calc. Var.*, 10(4):549–552 (electronic), 2004.

[8] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3, Ser. A):293–318, 1992.

[9] L. C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137(653):viii+66, 1999.

[10] Mikhail Feldman and Robert J. McCann. Uniqueness and transport density in Monge's mass transportation problem. *Calc. Var. Partial Differential Equations*, 15(1):81–113, 2002.

[11] Michel Fortin and Roland Glowinski. *Augmented Lagrangian methods*, volume 15 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.

[12] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element methods. *Computers and Mathematics with applications*, 2:17–40, 1976.

[13] Roméo Hatchi. Wardrop equilibria : long-term variant, degenerate anisotropic PDEs and numerical approximations. Technical report, 2015.

[14] Frédéric Hecht. New development in freefem++. *Journal of Numerical Mathematics*, 20(3-4):251–266, 2012.

[15] Jacob Rubinstein and Gershon Wolansky. A weighted least action principle for dispersive waves. *Ann. Physics*, 316(2):271–284, 2005.

[16] Filippo Santambrogio. Absolute continuity and summability of transport densities: simpler proofs and new estimates. *Calc. Var. Partial Differential Equations*, 36(3):343–354, 2009.

[17] Cédric Villani. *Topics in optimal transportation*, volume 58. AMS Bookstore, 2003.