COMMUN. MATH. SCI.
Vol. 13, No. 1, pp. 269–296

OPTIMAL TRANSPORT FOR PARTICLE IMAGE VELOCIMETRY*

LOUIS-PHILIPPE SAUMIER†, BOUALEM KHOUIDER‡, AND MARTIAL AGUEH§

Abstract. We present a new method for particle image velocimetry, a technique using successive laser images of particles immersed in a fluid to measure the velocity field of the fluid flow. The main idea is to recover this velocity field via the solution of the L^2-optimal transport problem associated with each pair of successive distributions of tracers. We model the tracers by a network of Gaussian-like distributions and derive rigorous bounds on the approximation error in terms of the model's parameters. To obtain the numerical solution, we employ Newton's method, combined with an efficient spectral method, to solve the Monge-Ampère equation associated with the transport problem. We present numerical experiments based on two synthetic flow fields, a plane shear and an array of vortices. Although the theoretical results are derived for the case of a single particle in dimensions one and two, the results are valid in $\mathbb{R}^d$, $d \geq 1$. Moreover, the numerical experiments demonstrate that these results hold for the case of multiple particles, provided the Monge-Ampère equation is solved on a fine enough grid.

Key words. Optimal transport, particle image velocimetry, Monge-Ampère equation, numerical method.

AMS subject classifications. 35J96, 49M15, 65M06, 76M28, 90C99.

1. Introduction

The technique of particle image velocimetry (PIV) consists in determining the velocity field of a certain fluid flow using small particles as tracers [1, 20]. The main idea is to introduce many of these passive tracers in the fluid and let them evolve. Provided they accurately follow the motion of the fluid, then one hopes to quantify the velocity field from the measurements of the distribution of the tracers at different times. The particles are typically indistinguishable, and thus individual particle tracking is not always possible, depending on the flow, the time-intervals considered and the particle concentration. In fact, when the average distance between distinct particles is much larger than the mean displacement, individual tracking becomes feasible and the technique of particle image velocimetry is referred to as particle tracking velocimetry (PTV). Usually, PIV is implemented using tracers with desirable light-reflection properties. Every time a measurement is made, the fluid (or only a section of the fluid) is illuminated with a laser (for example), and a camera records an image showing the light reflected by the particles. The experimenter therefore collects a sequence of images and uses them to approximate the velocity field. Traditionally, PIV is performed by illuminating only a very thin slice of the fluid so the image recorded is two-dimensional. However, new techniques such as Tomographic PIV make it possible to obtain three-dimensional data [9].

The standard way of extracting the velocity field from two recorded images is to use the statistical method of cross-correlation [13, 20, 27]. In a nutshell, it consists in splitting the region in which we want to recover the vector field into smaller subregions called interrogation windows. Then, the average displacement of the tracers inside

*Received: November 22, 2012; accepted (in revised form): January 2, 2014. Communicated by Paul Milewski.

†Department of Mathematics and Statistics, University of Victoria, PO Box 1700 STN CSC, Victoria, B.C., Canada, V8W 2Y2 (lsaumier@uvic.ca).

‡Department of Mathematics and Statistics, University of Victoria, PO BOX 1700 STN CSC, Victoria, B.C., Canada, V8W 2Y2 (khouider@uvic.ca).

§Department of Mathematics and Statistics, University of Victoria, PO Box 1700 STN CSC, Victoria, B.C., Canada, V8W 2Y2 (agueh@uvic.ca).

one interrogation window is found by computing the maximum of the cross-correlation function between the parts of the two images inside the window. This cross-correlation function is obtained by convolving the first image with the reflected second image. By repeating this procedure for every interrogation window, one creates a grid of average displacements from which the approximate velocity field is recovered.

The goal in this paper is to introduce a new way of estimating the velocity field from the successive images measuring the light scattered by the tracers. Our idea is to use the framework of the optimal transport (OT) problem [17, 25, 26], which consists in finding the most efficient way of redistributing material from one configuration to another. By most efficient, we mean that the optimal map T redistributing the material minimizes a given transportation cost. More precisely, if x is the initial position of a point in a domain Ω and if y is its final position in a domain Ω', then the cost of moving x to y is given by a function $c(x,y)$. Let Ω and $\Omega^{'}$ be closed and bounded subsets of $\mathbb{R}^d$. We model the material distributions by two probability densities denoted f and g. In the case where $c(x,y)$ is the square of the usual Euclidean distance, this problem, called the L^2 optimal transport problem, reads

$$\text{Minimize } I[T]=\frac{1}{2}\int_{\Omega}|x-T(x)|^2 f(x)\,\mathrm{d}x, \tag{1.1}$$

over all maps T transporting f to g: $T\#f=g$. The condition $T\#f=g$, called the push-forward condition, is essentially a mass conservation constraint which can be expressed by

$$\int_{\Omega'}\mathcal{X}_E(y)g(y)\,\mathrm{d}y=\int_{\Omega}\mathcal{X}_{T^{-1}(E)}(x)f(x)\,\mathrm{d}x=\int_{\Omega}\mathcal{X}_E(T(x))f(x)\,\mathrm{d}x, \tag{1.2}$$

for any Borel set E in Ω'. Here, $\mathcal{X}_E$ denotes the characteristic function of E. Equivalently, we have

$$\int_{\Omega'}\zeta(y)g(y)\,\mathrm{d}y=\int_{\Omega}\zeta(T(x))f(x)\,\mathrm{d}x$$

for any $\zeta\in\mathcal{C}(\Omega')$. Then, by the change of variable $y=T(x)$, assuming T is at least $\mathcal{C}^1$, we obtain

$$\int_{\Omega}\zeta(T(x))g(T(x))|\det(\mathrm{D}T)|\,\mathrm{d}x=\int_{\Omega}\zeta(T(x))f(x)\,\mathrm{d}x.$$

Because ζ is arbitrary, we arrive at

$$g(T(x))|\det(\mathrm{D}T)|=f(x). \tag{1.3}$$

In [5], Brenier proved that T is the gradient of a convex potential, i.e. $T=\nabla\Psi$ where Ψ is convex. If we insert this formula in (1.3), we obtain the so-called Monge-Ampère equation

$$g(\nabla\Psi(x))\det(\mathrm{D}^2\Psi)=f(x). \tag{1.4}$$

This equation is a fully nonlinear second order elliptic partial differential equation. It is well-known (see [25] for example) that the gradient of a convex function $\Psi\in\mathcal{C}^2(\Omega)$ which solves (1.4) is the optimal solution of the L^2 OT problem (1.1). There are several methods available to solve numerically the transport problem [2, 8, 12, 21, 22]. In

[21, 22], we derived an efficient numerical algorithm, based on the work of [15] and [23], for solving (1.1) through (1.4). For a comparison between this method and some of the others, we refer to [21]. We also mention that the optimal transport problem has been the basis of many image processing applications; see for example [6, 18, 19, 24]. Coming back to the PIV problem, we see that it is natural to consider $f(x)$ and $g(x)$ to be the initial and final distributions of tracers. Our aim is thus to show how the L^2 OT problem associated with such densities can be used to approximate the velocity field in the PIV problem and to analyze the different errors introduced by this procedure.

As a simple model for f and g, we select sums of arrays of Gaussian distributions with a common standard deviation σ that are centered respectively at the initial and final positions of the tracer particles. The advantage of this Gaussian approach is twofold. First, it guarantees a smooth solution for the OT problem, ensuring a nice behaviour of our numerical scheme used to solve the Monge-Ampère equation. Second, this parameter is physically motivated because the response of an imaging system to a point source, namely the point spread function, is well approximated by a Gaussian distribution. Note that it may also mimic the measurement uncertainty for the actual locations of the particles. Some models for PIV take into account possible differences in the shape of the particles caused for example by small distortions in the reflected light. In addition, in many situations the quantity of light reflected by the particles is not uniform for all particles, but rather normally distributed with mean corresponding to the center of the light beam employed to illuminate the tracers. Therefore, in such a model, the density distributions of tracers are convolved with a Gaussian kernel and the resulting distributions are the ones on which cross-correlation is being applied. For simplicity, we shall assume in this paper that all the particles have the same weight in the modelled distribution, and thus there are no distortions or variations in the light reflected. This could physically correspond to analyzing only a subsection of the illuminated region for which the light variations are negligible.

Besides the standard deviation σ, we introduce another parameter r to lift up f and g away from zero by adding the constant $1-r$. We then rescale the densities' Gaussian parts so they remain well-defined probability densities on Ω. While the constant $1-r$ can be viewed as a parameter modelling a uniformly distributed background noise in the light measurements due to the medium in which the particles evolve, the algorithm we employ to solve the optimal transport problem actually requires the densities f and g to be bounded away from zero to converge [21, 22]. This condition ensures that the approximate solution stays within the class of convex potentials as we iterate, which is required for the problem to remain well-posed. As we will see later, it is possible to take r very close to 1 and still obtain convergence of the algorithm; we just need to take a large enough grid size for the numerical resolution.

Even though σ and r are first motivated by our numerical method to solve the transport problem, they are both physically meaningful for reasons we just explained. Actually, it is not unreasonable to think that in some (nice) cases, the image intensity field associated with the light captured by a camera in a PIV experiment could contain sums of Gaussian-like functions with a somewhat uniform background noise, especially because some algorithms used to solve the PIV problem [20] employ low-pass filters as pre-processing steps in order to reduce the noise coming from particles located outside of the light sheet (in the case of 2D PIV). In general, the PIV images often contain blobs of light (corresponding to the light scattered by close-by particles) which do not resemble Gaussian distributions, but because this work only aims at answering

introductory questions regarding the OT framework for PIV, we will use the simple model we introduced.

It is reasonable to think that the optimal map T transporting f onto g in the L^2 OT problem will provide a good approximation of the particles' true displacements, as long as the time interval between the two snapshots, denoted Δt, is not too large. More precisely, if Δt is small enough so that the average particle displacement is no greater than about half of the average distance between particles (in other words, if we are in the realm of PTV), then a direct computation of $(T(x)-x)/\Delta t$ should yield a good approximation of the fluid's velocity field at the particles' locations. However, in cases where Δt is larger than this, a multiresolution approach will most likely be required in order to get a good approximate flow. In fact, modern PIV methods based on cross-correlation usually employ some kind of multiresolution method, where for example several passes of the algorithm are used to "offset" the interrogation window in the direction of the flow's motion [20]. Because this paper is only a first step towards using the OT theory to tackle the PIV problem, we will be mostly concerned about analyzing the behaviour of the transport map and thus we will focus in our numerical experiments on the former case (Δt not too large) and directly use $(T(x)-x)/\Delta t$. We intend to deal with more general cases in the future.

In the regime of PTV, one possible way of recovering the field would be to first identify the meaningful particles and then solve the corresponding assignment problem. There are several algorithms available to solve the assignment problem, some based on the simplex algorithm [11, 28], the Hungarian algorithm [14, 28], or the auction algorithm [4], but in general, for densities consisting of thousands of Dirac delta distributions, this remains a challenging problem [10]. In the specific case of PIV, one can of course limit the valid assignments to close neighbours, but even then the number of possible matches can still be high. In addition, individual particles are not always easy to identify from the light density distribution; bright clusters of light can represent the light scattered by several particles. Motivated partly by these reasons, we use a continuous approach both for the model and the numerical resolution in this work. Our hope is that this will serve as a basis for future implementation of optimal transport methods handling general density distributions of tracers. In fact, it is well-known [25] that the velocity field associated to the time-dependent optimal transport map solves the system

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \qquad \rho(t=0,\cdot) = f, \\ \dfrac{\partial v}{\partial t} + v \cdot \nabla v = 0. \end{cases}$$

While we are concerned with an approximation of a vector field which is a solution to the Navier-Stokes equations, the OT method provides in essence a solution to the pressureless Euler equations at zero temperature. An interesting problem would thus be to figure out under which conditions this flow provides a good approximation to the "full" Navier-Stokes equations, but this is beyond the scope of this work.

In our model, when there is no uncertainty in measurement ($\sigma = 0$) and no noise ($r = 1$), our densities actually become sums of Dirac delta functions. In such a case, the initial positions of the tracers are moved directly to their true final positions in the transport problem. However, we observed numerically that when $\sigma > 0$ and $r < 1$, the transport map T does not map the centers of the Gaussians of f to the centers of the Gaussians of g. This introduces an additional error along with the finite differences error already present when approximating the velocity field by the

transportation map. Two questions then naturally arise when laying out this model for the images obtained with PIV. First, how do the uncertainty in the measurement of σ and the noise level $1-r$ affect the approximation of the final positions of the particles obtained by solving the OT problem? Second, how significant is this error in terms of the approximation of the actual velocity field? As we will see through numerical experiments, the answer to the second question is that it is dominated by the error in finite differences. To answer the first question, we will derive error bounds giving the rate of decay of this error with respect to σ, r, and Δt.

Despite some possible difficulties which remain to be further explored, we claim that the optimal transport theory for PIV has many major advantages. Indeed, it is a global approach, i.e. it is not necessary to split the region into interrogation windows to compute the velocity field; the solution is computed everywhere at once. Several algorithms are now available to compute the solution of the optimal transport problem, and some, like the one we employ here, are especially efficient. Moreover, even though we use Gaussian distributions to model individual tracers (which makes it easier to mathematically analyze the behavior of OT applied to PIV), the flexibility of OT would easily allow us to take virtually any distribution of tracers. As previously mentioned, one could imagine the resulting images looking more like random clusters of pixels than nice, mathematically appealing density distributions. In addition, OT can be employed for solving two-dimensional problems just as well as three-dimensional problems. We will further discuss these claims later.

The rest of the paper is organized as follows. In Section 2, we present the details of the synthetic PIV data model and discuss the numerical method used to solve the OT problem. In Section 3, we consider the case of one particle in dimension one to derive and validate the main results of this paper: we show that the error due to the OT approximation converges linearly in $(1-r)/r$ and σ. Next, in Section 4, we generalize the results of Section 3 to dimension two. To do so, we need to impose some rather intuitive assumptions, and we justify them with numerical simulations. In Section 5, we consider the case of multiple particles immersed in two different synthetic flows: a simple plane shear flow and an array of vortices. In order to investigate the behaviour of the transport map (and thus the behaviour of our approximate vector field) in our model, we compute the approximation error for different values of the parameters. Our numerical results demonstrate that the error bounds established in sections 3 and 4 remain valid for the case of many particles. Finally, a discussion and some concluding remarks are given in Section 6.

2. Model and algorithm

2.1. Model setup. Let $\Omega=[0,1]^2$ be the domain in which we want to recover a vector field $\mathbf{v}(\mathbf{x},t)$ representing the velocity of a given two-dimensional fluid, where $\mathbf{x}=(x_1,x_2)$. Note that from this section onwards, we will use the boldface notation to denote vectors. Consider $(M-1)^2$ tracers (point particles) immersed in the fluid that are represented by Gaussian distributions modified in the following way. We consider the global distribution defined on the domain Ω by

$$f(x_1,x_2)=r\sum_{i=1}^{M-1}\sum_{j=1}^{M-1}\frac{1}{k_{\boldsymbol{\lambda}_{ij}}}\exp\Big(-18M^2s^2\big[(x_1-\lambda_{1_{ij}})^2+(x_2-\lambda_{2_{ij}})^2\big]\Big)+(1-r),$$

$$\text{where}\quad k_{\boldsymbol{\lambda}_{ij}}=(M-1)^2\int_{[0,1]^2}\exp\Big(-18M^2s^2\big[(x_1-\lambda_{1_{ij}})^2+(x_2-\lambda_{2_{ij}})^2\big]\Big)\mathrm{d}\mathbf{x}$$

and $r \in [0,1]$, $s > 0$. Here, $\boldsymbol{\lambda}_{ij} = (\lambda_{1_{ij}}, \lambda_{2_{ij}})$ represents the initial position of particle ij and the $k_{\boldsymbol{\lambda}_{ij}}$ are normalization constants. In addition, to avoid any major overlap of the Gaussians, we took the standard deviation to be $\sigma = 1/(6Ms)$ where s is a parameter introduced to vary the sharpness of the spikes.

If we denote by

$$\boldsymbol{\Phi}(x_1, x_2, t) \equiv (\Phi_1(x_1, x_2, t), \Phi_2(x_1, x_2, t))$$

the position at time t of a particle evolving in the velocity field $\mathbf{v}$, given that it started at position (x_1, x_2), then this trajectory function solves

$$\begin{cases} \dfrac{\partial \boldsymbol{\Phi}}{\partial t} = \mathbf{v}(\boldsymbol{\Phi}(\mathbf{x}, t), t), \\ \boldsymbol{\Phi}(\mathbf{x}, 0) = \mathbf{x}. \end{cases}$$

We let the $(M-1)^2$ tracers evolve in the field for a small time Δt and assume that the flow does not alter the shape of the Gaussians, but only translates their centers. This makes physical sense because we are only concerned with the movement of the tracers. The change in shape of the Gaussians will be important to incorporate if we are considering clusters rather than single particles (see Section 6). In the sequel, we denote by $\boldsymbol{\mu}_{ij} \equiv (\mu_{1_{ij}}, \mu_{2_{ij}}) = (\Phi_1(\boldsymbol{\lambda}_{ij}, \Delta t), \Phi_2(\boldsymbol{\lambda}_{ij}, \Delta t))$ the final position of the particle ij. The final disposition of tracers is thus given by

$$g(x_1, x_2) = r \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} \frac{1}{k_{\boldsymbol{\mu}_{ij}}} \exp\Big(-18M^2 s^2 \big[\big(x_1 - \mu_{1_{ij}}\big)^2 + \big(x_2 - \mu_{2_{ij}}\big)^2 \big] \Big) + (1 - r),$$

where the normalization constant is defined similarly:

$$k_{\boldsymbol{\mu}_{ij}} = (M-1)^2 \int_{[0,1]^2} \exp\Big(-18M^2 s^2 \big[\big(x_1 - \mu_{1_{ij}}\big)^2 + \big(x_2 - \mu_{2_{ij}}\big)^2 \big] \Big) \mathrm{d}\mathbf{x}.$$

The densities f and g are essentially sums of Gaussian distributions which have been scaled, pushed away from 0 and cut to "fit" the bounded domain Ω. The integrals $k_{\boldsymbol{\lambda}_{ij}}$ and $k_{\boldsymbol{\mu}_{ij}}$ are normalization constants ensuring that f and g are well-defined probability densities on $[0,1]^2$. Finally, the parameter r is introduced to push these densities away from 0 by a factor $1 - r$. It is well known (see [25]) that the optimal solution to the L^2 time-dependent transport problem is given by

$$\mathbf{T}_t(\mathbf{x}) = \left(1 - \frac{t}{\Delta t}\right)\mathbf{x} + \frac{t}{\Delta t}\mathbf{T}(\mathbf{x}),$$

so that when $t = \Delta t$ the optimal map is $\mathbf{T}(\mathbf{x})$, and when $t = 0$ it is just the identity. The interpolation $\mathbf{T}_t$ was first introduced by McCann in [16]. We have

$$\frac{\partial}{\partial t}\mathbf{T}_t(\mathbf{x}) = \frac{\mathbf{T}(\mathbf{x}) - \mathbf{x}}{\Delta t} \approx \mathbf{v}(\mathbf{x}, t),$$

provided Δt is small enough. Here, we use $\mathbf{v}_{\text{approx}} = (\mathbf{T}(\mathbf{x}) - \mathbf{x})/\Delta t$ as a first order approximation for the actual fluid flow velocity.

2.2. Optimal transport algorithm. We now introduce our procedure to solve the L^2 optimal transport problem in $\mathbb{R}^d$, $d \geq 1$. For simplicity, we enforce periodic boundary conditions and assume that the initial and target domains are equal to $\Omega = [0,1]^d$. In this setting, it has been proven in [7] that, provided $f, g \in \mathcal{C}^\alpha$ for $\alpha > 0$ and f, g are bounded away from 0, the optimal map $\nabla\Psi$ is additive, unique up to a constant, and $\mathcal{C}^{2,\beta}(\Omega)$ for $0 < \alpha < \beta$. By additivity, we mean $\nabla\Psi(\mathbf{x}+\mathbf{p}) = \nabla\Psi(\mathbf{x}) + \mathbf{p}$ for every $\mathbf{x} \in \mathbb{R}^d$ and for all $\mathbf{p} \in \mathbb{Z}^d$. This property implies that $\Psi(x)$ can be written in the form $\Psi(\mathbf{x}) = |\mathbf{x}|^2/2 + u(\mathbf{x})$ where u is a periodic function on Ω. The optimal map can thus be expressed as $T(\mathbf{x}) = \nabla\Psi(\mathbf{x}) = \mathbf{x} + \nabla u(\mathbf{x})$. Assuming Ψ is convex, the Monge-Ampère equation becomes

$$g(\mathbf{x} + \nabla u(\mathbf{x})) \det(\mathcal{I} + \mathrm{D}^2 u(\mathbf{x})) = f(\mathbf{x}), \tag{2.1}$$

and the approximated velocity field is $\mathbf{v}_{\text{approx}} = \nabla u / \Delta t$. To solve (2.1), we use a damped Newton method based on a generalization of the work of [15]. More details on this procedure are found in [21, 22]. The main steps are outlined below for the sake of completeness.

Damped Newton algorithm

$$\begin{cases} \text{With } u_0 \text{ given, loop over } n \in \mathbb{N} \\ \text{Compute } f_n = g(\mathbf{x} + \nabla u_n) \det(\mathcal{I} + D^2 u_n) \\ \text{Solve the linearized Monge-Ampère equation} \\ \qquad L_n \theta_n := D_{u_n} M \cdot \theta_n = \dfrac{1}{\tau}(f - f_n) \\ \text{Update the solution: } u_{n+1} = u_n + \theta_n. \end{cases} \tag{2.2}$$

Here, $D_{u_n} M \cdot \theta_n$ is the linearization of the Monge-Ampère operator in the direction of θ_n:

$$D_{u_n} M \cdot \theta_n = g(\mathbf{x} + \nabla u) \operatorname{Tr}(\operatorname{Adj}(\mathcal{I} + D^2 u) D^2 \theta_n) + \det(\mathcal{I} + D^2 u) \nabla g(\mathbf{x} + \nabla u) \cdot \nabla \theta_n, \tag{2.3}$$

where $\operatorname{Adj}(A) = \det(A) \cdot A^{-1}$. We will denote $D_{u_n} M \cdot \theta_n$ by $L_n \theta_n$ for simplicity. Moreover, the damping factor τ is introduced as a parameter controlling the "step-size" of the Newton iterations, which is necessary to ensure that we remain within the class of convex potentials while iterating towards the solution. For the resolution of the linearized Monge-Ampère equation (2.2), which is an elliptic linear PDE with non-constant coefficients, we employ a technique introduced in [23] which consists in first preconditioning the system in the following way:

$$\begin{cases} L_n \overline{L}_n^{-1} \sigma(x) = h(x), \\ \qquad \overline{L}_n \theta(x) = \sigma(x), \end{cases} \tag{2.4}$$

where $\overline{L}_n$ is the preconditioner, obtained from L_n by averaging the coefficients on $[0,1]^d$. Then, (2.4) is solved by the fast Fourier transform (FFT) algorithm [21, 22]. We use the following stopping criterion for the Newton algorithm iterations:

$$\|u_n - u_{n-1}\|_{l^2_{\text{norm}}} + \|f - f_n\|_{l^2_{\text{norm}}} < \text{TOL}_{\text{Newton}}, \tag{2.5}$$

where $\mathrm{TOL}_{\mathrm{Newton}}$ is a given tolerance and $\|\cdot\|_{l^2_{\mathrm{norm}}}$ is the normalized l^2 norm defined for $\mathbf{z}=(z_1,z_2,\cdots,z_m)\in\mathbb{R}^{m\times k}$ by

$$\|\mathbf{z}\|_{l^2_{\mathrm{norm}}}:=\sqrt{\frac{1}{m}\sum_{i=1}^{m}\|z_i\|_{l^2}^2}.$$

The quality of the approximation will of course depend on the parameters r, σ, and Δt. Our first task is thus to study the sensitivity of the approximation on these parameters for the simple case of a single particle in both 1D and 2D (the results for higher dimensional cases follow easily). We start with the model in dimension one to prove that the error converges linearly in $(1-r)/r$, σ, and Δt, and then generalize the analysis to dimension two. In Section 5, we present numerical experiments to demonstrate that this result remains valid for the case of multiple particles.

3. One particle in dimension one

3.1. Error estimates. We consider the case where there is only one particle in dimension one, i.e. $\Omega=[0,1]$. In this situation, we can derive an explicit representation of the optimal map T through the exact solution of the Monge-Ampère equation (1.4). This will provide us with precious insight on the behaviour of T in higher dimensions, for the particle image velocimetry problem. The one-dimensional and one particle equivalent to the densities f and g introduced in Subsection 2.1 are given by

$$f(x)=\frac{r}{k_\lambda}e^{-\frac{(x-\lambda)^2}{2\sigma^2}}+(1-r),\quad g(x)=\frac{r}{k_\mu}e^{-\frac{(x-\mu)^2}{2\sigma^2}}+(1-r),\tag{3.1}$$

where

$$k_\lambda=\int_0^1 e^{-\frac{(t-\lambda)^2}{2\sigma^2}}\,\mathrm{d}t,\quad k_\mu=\int_0^1 e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t,\tag{3.2}$$

λ, μ being two points in Ω and $r\in[0,1]$. For simplicity, in this section, we will directly work with the parameter σ instead of s and M (which is 2 here). Because our technique consists in approximating the velocity vector at λ by $(T(\lambda)-\lambda)/\Delta t$, we would hope that $T(\lambda)$ is equal to μ. However, the parameters σ and r induce an error that sends $T(\lambda)$ off the target μ. In fact, we have the following theorem.

THEOREM 3.1. *Assume $2/3<r\le 1$, $0<\epsilon\le\lambda\le\mu\le 1-\epsilon$ for $0<\epsilon<0.5$, and $0<\sigma<1$. If T is the optimal map solving the one-dimensional L^2 transport problem with densities f and g given by* (3.1)*, then*

$$0\le\mu-T(\lambda)\le C_1\sigma\frac{(1-r)}{r}(\mu-\lambda)+C_2\frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma}(\mu-\lambda),\tag{3.3}$$

where C_1 and C_2 are positive constants independent of σ, λ, μ, r, and ϵ. Moreover, if all other parameters are fixed, $\mu-T(\lambda)$ does not decrease faster than linearly when $\sigma\to 0$.

Before we present the proof of Theorem 3.1, we state and prove the following technical lemma.

LEMMA 3.2. *Under the assumptions of Theorem 3.1, there exists a constant C_2 independent of σ, λ, μ, r, and ϵ such that*

$$0 \leq \frac{k_\mu \left(\frac{1}{k_\mu} \int_0^\mu e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt - \frac{1}{k_\lambda} \int_0^\lambda e^{-\frac{(t-\lambda)^2}{2\sigma^2}}\, dt \right)}{k_\mu \frac{(1-r)}{r} + \int_0^1 e^{-\frac{t^2 (T(\lambda)-\mu)^2}{2\sigma^2}}\, dt} \leq C_2 \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma} (\mu - \lambda), \tag{3.4}$$

where k_λ and k_μ are given in (3.2).

Proof. If $\lambda = \mu$, the result is trivial. Assume $\lambda < \mu$. We introduce the function

$$G(\gamma) = \frac{\int_0^\gamma e^{-\frac{(t-\gamma)^2}{2\sigma^2}}\, \mathrm{d}t}{\int_0^1 e^{-\frac{(t-\gamma)^2}{2\sigma^2}}\, \mathrm{d}t}.$$

We observe that its derivative,

$$G'(\gamma) = \frac{e^{-\frac{\gamma^2}{2\sigma^2}} \int_\gamma^1 e^{-\frac{(t-\gamma)^2}{2\sigma^2}}\, \mathrm{d}t + e^{-\frac{(\gamma-1)^2}{2\sigma^2}} \int_0^\gamma e^{-\frac{(t-\gamma)^2}{2\sigma^2}}\, \mathrm{d}t}{\left(\int_0^1 e^{-\frac{(t-\gamma)^2}{2\sigma^2}}\, \mathrm{d}t \right)^2},$$

is always positive, which means that G is an increasing function. By the mean-value theorem,

$$\exists \xi : \lambda \leq \xi \leq \mu \quad \text{and} \quad G(\mu) - G(\lambda) = G'(\xi)\,(\mu - \lambda).$$

Moreover, because $\lambda, \mu \in [\epsilon, 1-\epsilon]$, we can bound G' in the following way:

$$G'(\xi) \leq \frac{e^{-\frac{\epsilon^2}{2\sigma^2}} \int_\xi^1 e^{-\frac{(t-\xi)^2}{2\sigma^2}}\, \mathrm{d}t + e^{-\frac{\epsilon^2}{2\sigma^2}} \int_0^\xi e^{-\frac{(t-\xi)^2}{2\sigma^2}}\, \mathrm{d}t}{\left(\int_0^1 e^{-\frac{(t-\xi)^2}{2\sigma^2}}\, \mathrm{d}t \right)^2} = \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\int_0^1 e^{-\frac{(t-\xi)^2}{2\sigma^2}}\, \mathrm{d}t}.$$

Combining the above estimate and the fact that the optimal map T sends $[0,1]$ to $[0,1]$, we have

$$\begin{aligned}
\frac{k_\mu (G(\mu) - G(\lambda))}{k_\mu \frac{(1-r)}{r} + \int_0^1 e^{-\frac{t^2 (T(\lambda)-\mu)^2}{2\sigma^2}}\, \mathrm{d}t} &\leq \frac{\int_0^1 e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, \mathrm{d}t}{\int_0^1 e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t} \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\int_0^1 e^{-\frac{(t-\xi)^2}{2\sigma^2}}\, \mathrm{d}t} (\mu - \lambda) \\
&\leq \frac{\int_0^1 e^{-\frac{(t-1/2)^2}{2\sigma^2}}\, \mathrm{d}t}{\int_0^1 e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t} \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\int_0^1 e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t} (\mu - \lambda) \\
&= \frac{\int_{-1/2}^{1/2} e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t}{\int_0^1 e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t} \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma \int_0^{1/\sigma} e^{-\frac{t^2}{2}}\, \mathrm{d}t} (\mu - \lambda),
\end{aligned}$$

where in the last line we performed two changes of variables. Finally, from the symmetry of the Gaussian distribution and the assumption $\sigma < 1$, we get

$$\frac{\int_{-1/2}^{1/2} e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t}{\int_0^1 e^{-\frac{t^2}{2\sigma^2}}\, \mathrm{d}t} \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma \int_0^{1/\sigma} e^{-\frac{t^2}{2}}\, \mathrm{d}t} (\mu - \lambda) \leq 2 \frac{1}{\int_0^1 e^{-\frac{t^2}{2}}\, \mathrm{d}t} \frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma} (\mu - \lambda).$$

The inequality (3.4) follows with $C_2 = 2/\int_0^1 e^{-\frac{t^2}{2}}\,\mathrm{d}t$. ∎

Proof. (Theorem 3.1.) We divide the proof into three main parts. First, we will show that

$$0 \le \mu - T(\lambda) \le K(\sigma)\frac{(1-r)}{r}(\mu-\lambda) + C_2\frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma}(\mu-\lambda), \tag{3.5}$$

where $K(\sigma)$ is a function of σ which is bounded by 2, and C_2 is a constant. Second, we will prove that there exists a constant C_1 such that $K(\sigma) \le C_1\sigma$, and finally, we will show that $T(\lambda)-\mu$ cannot decay faster than linearly in σ.

To prove the first claim, we introduce the cumulative distribution functions for $f(x)$ and $g(x)$,

$$\begin{aligned} F(x) &= \frac{r}{k_\lambda}\int_0^x e^{-\frac{(t-\lambda)^2}{2\sigma^2}}\,\mathrm{d}t + (1-r)x \\ \text{and} \quad G(x) &= \frac{r}{k_\mu}\int_0^x e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t + (1-r)x. \end{aligned} \tag{3.6}$$

Because the solution of the L^2 transport problem is the gradient of a convex function (which in dimension one corresponds to a function with nonnegative second derivative), in this simple 1D case the Monge-Ampère equation (1.4) becomes

$$g(\Psi'(x))\Psi''(x) = f(x),$$

which integrates to

$$T(x) \equiv \Psi'(x) = G^{-1}(G(T(0)) + F(x)).$$

Now, because the optimal map T is increasing and invertible, we have $T(0)=0$ and $G(0)=0$, so $T(x)=G^{-1}(F(x))$. We want to estimate $\mu - T(\lambda) = \mu - G^{-1}(F(\lambda))$. Let $y = T(\lambda)$. According to (3.6), we have $G(y) = F(\lambda)$, i.e.,

$$\frac{1}{k_\mu}\int_0^y e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t + \frac{(1-r)}{r}y = \frac{1}{k_\lambda}\int_0^\lambda e^{-\frac{(t-\lambda)^2}{2\sigma^2}}\,\mathrm{d}t + \frac{(1-r)}{r}\lambda. \tag{3.7}$$

Next, we consider

$$N(y) = \frac{1}{k_\mu}\int_0^y e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t,$$

and compute its Taylor expansion about μ with integral remainder:

$$\begin{aligned} N(y) &= N(\mu) + N'(\mu)(y-\mu) + (y-\mu)^2\int_0^1 (1-t)N''(\mu + t(y-\mu))\,\mathrm{d}t \\ &= \frac{1}{k_\mu}\int_0^\mu e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t + \frac{(y-\mu)}{k_\mu} - \frac{(y-\mu)^3}{k_\mu\sigma^2}\int_0^1 t(1-t)e^{-\frac{t^2(y-\mu)^2}{2\sigma^2}}\,\mathrm{d}t \\ &= \frac{1}{k_\mu}\int_0^\mu e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t + \frac{(y-\mu)}{k_\mu}\int_0^1 e^{-\frac{t^2(y-\mu)^2}{2\sigma^2}}\,\mathrm{d}t. \end{aligned}$$

Note that we used integration by parts to obtain the last line. Isolating the linear factors in $\mu - y$ (in (3.7)) and multiplying by k_μ yields

$$\left(k_\mu\frac{(1-r)}{r} + \int_0^1 e^{-\frac{t^2(y-\mu)^2}{2\sigma^2}}\,\mathrm{d}t\right)(\mu-y)$$

$$=k_\mu\frac{(1-r)}{r}(\mu-\lambda)+k_\mu\left(\frac{1}{k_\mu}\int_0^\mu e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t-\frac{1}{k_\lambda}\int_0^\lambda e^{-\frac{(t-\lambda)^2}{2\sigma^2}}\,\mathrm{d}t\right). \tag{3.8}$$

From Lemma 3.2, we know that the right-hand side is nonnegative, which implies that the first inequality in (3.3) is satisfied. Moreover, because T maps $[0,1]$ to itself, we have $\mu-y\le 1$, and thus $\exp(-\frac{t^2}{2\sigma^2})\le\exp(-\frac{t^2(y-\mu)^2}{2\sigma^2})$. We get

$$\frac{k_\mu}{k_\mu\frac{(1-r)}{r}+\int_0^1 e^{-\frac{t^2(y-\mu)^2}{2\sigma^2}}\,\mathrm{d}t}\le\frac{\int_0^1 e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t}{\int_0^1 e^{-\frac{t^2(y-\mu)^2}{2\sigma^2}}\,\mathrm{d}t}\le\frac{\int_0^1 e^{-\frac{(t-1/2)^2}{2\sigma^2}}\,\mathrm{d}t}{\int_0^1 e^{-\frac{t^2}{2\sigma^2}}\,\mathrm{d}t}\le 2, \tag{3.9}$$

from which we conclude the existence of $K(\sigma)$ in (3.5). In addition, the second term on the right of (3.8) leads to the expression in Lemma 3.2, which in turn yields (3.5).

Second, we prove the linearity in σ. Let $z(\sigma)=\mu-T(\lambda)$. Notice that (3.7) can be rewritten as

$$\frac{\int_{-\frac{\mu}{\sigma}}^{-\frac{z(\sigma)}{\sigma}} e^{-\frac{t^2}{2}}\,\mathrm{d}t}{\int_{-\frac{\mu}{\sigma}}^{\frac{(1-\mu)}{\sigma}} e^{-\frac{t^2}{2}}\,\mathrm{d}t}+\frac{(1-r)}{r}(\mu-\lambda)=\frac{\int_{-\frac{\lambda}{\sigma}}^{0} e^{-\frac{t^2}{2}}\,\mathrm{d}t}{\int_{-\frac{\lambda}{\sigma}}^{\frac{(1-\lambda)}{\sigma}} e^{-\frac{t^2}{2}}\,\mathrm{d}t}+\frac{(1-r)}{r}z(\sigma). \tag{3.10}$$

We proceed by contradiction. Assume that for every $n\in\mathbb{N}$, there exists a σ_n such that $z(\sigma_n)>n\sigma_n$. Because $z(\sigma)<1$ for every σ, the sequence $1/\sigma_n$ is bounded below by n and therefore $1/\sigma_n\to\infty$ as $n\to\infty$. Inserting σ_n in (3.10), taking the limit with respect to n, and using the fact that both λ and μ are bounded away from 0 and 1, respectively, yields

$$\frac{(1-r)}{r}(\mu-\lambda)=\frac{1}{2}+\frac{(1-r)}{r}\lim_{n\to\infty}z(\sigma_n)\ge\frac{1}{2}.$$

The last inequality results from $z(\sigma)$ being nonnegative for every σ, as shown in (3.5). Because $\mu-\lambda$ is smaller than 1, we get

$$\frac{(1-r)}{r}\ge\frac{1}{2},$$

which is a contradiction for $r>2/3$. We conclude the existence of a positive constant C_1 such that $K(\sigma)\le C_1\sigma$.

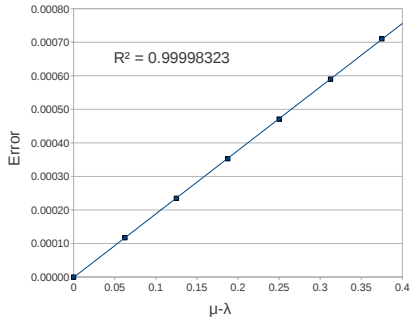
Finally to prove that $z(\sigma)=\mu-T(\lambda)$ cannot decay faster than linearly in σ, we proceed again by contradiction. Indeed, if we assume $\lim_{\sigma\to 0}z(\sigma)/\sigma=0$, and take the limit with respect to σ on both sides of (3.10), we get

$$\frac{1}{2}+\frac{(1-r)}{r}(\mu-\lambda)=\frac{1}{2},$$

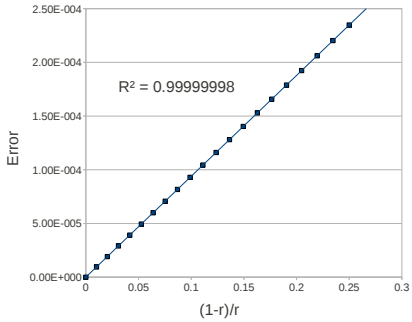
which yields another contradiction for fixed $r\neq 1$ and $\lambda\neq\mu$. ∎

REMARK 3.1. Observe that for any fixed $\epsilon>0$, the decay of the second term is always going to be exponential when σ goes to 0. However, if one wants to vary ϵ as a function of σ, then we need to take $\epsilon\ge\sigma^{1-\alpha}$ for $\alpha>0$ to preserve this exponential decay.

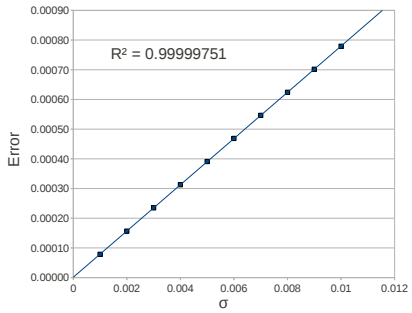
3.2. Numerical validation of 1D error estimates. In order to verify the linear convergence of $T(\lambda)$ to μ when $\mu-\lambda$, $(1-r)/r$, and σ go to 0, we solve for $T(\lambda)$ the equation $G(T(\lambda))=F(\lambda)$ using the function fzero of MATLAB, when the



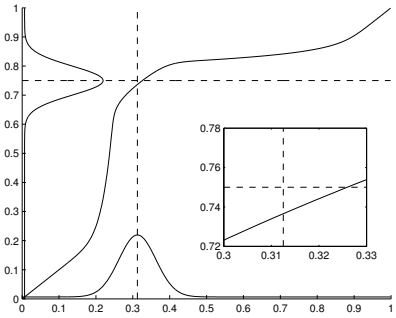
(a) $|T(\lambda)-\mu|$ as a function of $|\mu-\lambda|$, for $r=0.8$ (i.e. $(1-r)/r=0.25$) and $\sigma=0.003$.



(b) $|T(\lambda)-\mu|$ as a function of $(1-r)/r$, for $\sigma=0.003$, $\lambda=5/16$, and $\mu=7/16$.



(c) $|T(\lambda)-\mu|$ as a function of σ, for $r=0.8$, $\lambda=5/16$, and $\mu=7/16$.



(d) Plot of T for $\lambda=5/16$, $\mu=12/16$, $r=0.8$, and $\sigma=0.05$.

FIG. 3.1. *Results of the one-dimensional numerical experiment. For the three error graphs, the linear regression line passing through the data is included, with the corresponding coefficient of determination. For the plot of T, we also added a vertical line at λ, a horizontal line at μ, the densities f and g (normalized to fit in the box) and a zoombox displaying the area of the graph around (λ,μ).*

three parameters are varied one at the time. Notice that in this case, varying $\mu-\lambda$ is equivalent to varying the time-step Δt. The results are presented in figure 3.1 (a), (b), and (c). On these graphs, we also include the least square linear regression lines and the corresponding coefficients of determination R^2. Recall that the coefficient of determination is defined by

$$R^2 := 1 - \frac{\sum_{i=1}^{m}(z_i - l_i)^2}{\sum_{i=1}^{m}(z_i - \overline{z})^2},$$

where m is the number of data points, $z=(z_1,\cdots,z_m)$ are the data pointss, $l=(l_1,\cdots,l_m)$ are the corresponding predicted points on the line, and $\overline{z}$ is the mean of the observed data. We see that in all three cases, R^2 is very close to 1, which confirms the linear behaviour in dimension one. We therefore conclude that the image of the initial position of the center of the particle is sent closer to the real position of the center of the particle after Δt units of time when either $r\to 1$, $\sigma\to 0$, or $|\mu-\lambda|\to 0$,

and that the only way to get an exact answer is to take $\Delta t=0$ (i.e. $|\mu-\lambda|=0$) or $\sigma=0$. However, taking $r=1$ yields an error that decays exponentially with σ and linearly in $|\mu-\lambda|$.

In figure 3.1 (d), we plot the map T for a typical set of parameters. This map was obtained by laying out a uniform grid $x_i, i=0,\cdots,N$ on $[0,1]$ with $N=1024$ and solving $G(y_i)-F(x_i)=0$ for y_i, at every grid point x_i (once we computed y_i, we assigned $T(x_i)=y_i$).

4. One particle in dimension two

4.1. Preliminary lemmas. In higher dimensions, it is not possible to obtain the transport map explicitly, which makes it harder to rigorously derive error bounds. Nonetheless, we will be able to present in this section convincing arguments on why we expect the same error behavior as in the one-dimensional case. For the case of one particle in 2D, the densities f, g are given by

$$f(x_1,x_2)=\frac{r}{k_\lambda}\exp\Big(-18M^2s^2\big[(x_1-\lambda_1)^2+(x_2-\lambda_2)^2\big]\Big)+(1-r),$$
$$g(x_1,x_2)=\frac{r}{k_\mu}\exp\Big(-18M^2s^2\big[(x_1-\mu_1)^2+(x_2-\mu_2)^2\big]\Big)+(1-r), \tag{4.1}$$

where

$$k_\lambda=\int_{[0,1]^2}\exp\Big(-18M^2s^2\big[(x_1-\lambda_1)^2+(x_2-\lambda_2)^2\big]\Big)\mathrm{d}\mathbf{x},$$
$$k_\mu=\int_{[0,1]^2}\exp\Big(-18M^2s^2\big[(x_1-\mu_1)^2+(x_2-\mu_2)^2\big]\Big)\mathrm{d}\mathbf{x},$$

and $M=2$. Let $\sigma=1/(6Ms)$ and let $\mathbf{T}(x_1,x_2)=(T_1(x_1,x_2),T_2(x_1,x_2))$ be the optimal transport map. We introduce the sets

$$E_1=\{(x_1,x_2)\in[0,1]^2:x_1\le T_1(\lambda_1,\lambda_2)\}$$
$$\text{and}\qquad S_1=\mathbf{T}^{-1}(E_1)=\{(x_1,x_2)\in[0,1]^2:T_1(x_1,x_2)\le T_1(\lambda_1,\lambda_2)\}.$$

These sets will be useful to recover an analogue of our one-dimensional arguments for the first coordinate of the optimal transport map (the same arguments will apply as well for the second coordinate). Before we see how, let us look at an illustration in figure 4.1. Similarly to the one-dimensional case, we expect the first coordinate of the image of (λ_1,λ_2) to be smaller than μ_1 and thus the dashed line $x_1=T_1(\lambda_1,\lambda_2)$ is before the line $x_1=\mu_1$. The set E_1 corresponds in this case to the part of the domain which is to the left of this dashed line. Now, if we consider the set of all points for which the first coordinate of the transport map is smaller than this dashed line, which we called S_1, then we expect to get a region like the shaded region in figure 4.1. Indeed, close to the center of the particle, the optimal map should move the mass towards the final position (μ_1,μ_2) of the particle. However, for points further than several standard deviations to the center, the mass should stay roughly at the same position. We will discuss these claims in more detail by looking at some numerical examples of these sets in Subsection 4.2. Now, going back to our analysis, by the conservation of mass constraint (1.2), taking $E=E_1$, we have

$$\int_0^1\int_0^{T_1(\lambda_1,\lambda_2)}g(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2=\int\int_{S_1}f(x_1,x_2)\mathrm{d}\mathbf{x}.$$

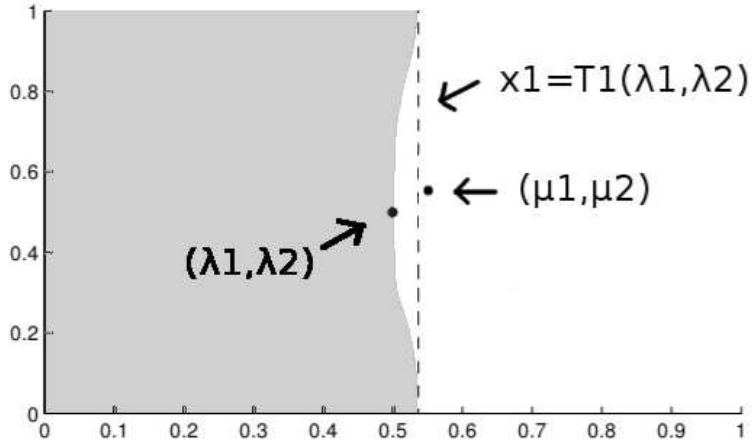


FIG. 4.1. *Illustration of the sets E_1 and S_1. The set E_1 corresponds to the part of $[0,1]^2$ which is on the left of the dashed line $x_1 = T_1(\lambda_1,\lambda_2)$ whereas the set S_1 corresponds to the shaded region. The initial position (λ_1,λ_2) and the final position (μ_1,μ_2) of the particle are also indicated.*

This in turn yields

$$\frac{1}{k_\mu}\int_0^1\int_0^{T_1(\lambda_1,\lambda_2)} \exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\mu_1)^2+(x_2-\mu_2)^2\big]\Big)\mathrm{d}x_1\mathrm{d}x_2+\frac{(1-r)}{r}T_1(\lambda_1,\lambda_2)$$
$$=\frac{1}{k_\lambda}\int\int_{S_1}\exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\lambda_1)^2+(x_2-\lambda_2)^2\big]\Big)\mathrm{d}\mathbf{x}+\frac{(1-r)}{r}\mathrm{m}(S_1). \tag{4.2}$$

Here, $\mathrm{m}(S_1)$ denotes the Lebesgue measure of the set S_1. Observe that (4.2) is very close to (3.7), the equality we used in the one-dimensional proof. Then, we can rewrite (4.2) in a similar way as in (3.8) by using a Taylor expansion, we get

$$(\mu_1-T_1(\lambda_1,\lambda_2))\Bigg[\frac{(1-r)}{r}$$
$$+\frac{1}{k_\mu}\int_0^1\int_0^1\exp\Big(-\frac{1}{2\sigma^2}\big[t^2(T_1(\lambda_1,\lambda_2)-\mu_1)^2+(x_2-\mu_2)^2\big]\Big)\mathrm{d}t\mathrm{d}x_2\Bigg]$$
$$=\frac{1}{k_\mu}\int_0^1\int_0^{\mu_1}\exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\mu_1)^2+(x_2-\mu_2)^2\big]\Big)\mathrm{d}x_1\mathrm{d}x_2$$
$$-\frac{1}{k_\lambda}\int\int_{S_1}\exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\lambda_1)^2+(x_2-\lambda_2)^2\big]\Big)\mathrm{d}\mathbf{x}+\frac{(1-r)}{r}\big[\mu_1-\mathrm{m}(S_1)\big]. \tag{4.3}$$

Before we pursue the analysis further, we shall investigate the properties of the set S_1. We have the following lemma.

LEMMA 4.1. *There exists a continuous function $h_1:[0,1]\to[0,1]$ such that $h_1(\lambda_2)=\lambda_1$ and*

a) $\mathbf{T}\big\{(x_1,x_2)\in\Omega : x_1 = h_1(x_2)\big\} = \big\{(x_1,x_2)\in\Omega : x_1 = T_1(\lambda_1,\lambda_2)\big\}$.

b) $S_1 = \bigcup_{x_2\in[0,1]}\Big([0,h_1(x_2)]\times\{x_2\}\Big)$.

Before we prove this lemma, we state and prove the following intermediate result.

LEMMA 4.2. *Let* $\mathbf{T}=(T_1,T_2)$ *be the unique solution to the optimal transport problem* (1.1) *on* $\Omega=[0,1]^2$ *with densities* f *and* g *bounded away from* 0 *and at least* $C^\alpha(\Omega)$ *for* $\alpha>0$. *Then* $\mathbf{T}$ *maps every side of the boundary to itself. More precisely,* $T_2(x_1,0)=0$, $T_2(x_1,1)=1$ *for every* $x_1\in[0,1]$, *and* $T_1(0,x_2)=0$, $T_1(1,x_2)=1$ *for every* $x_2\in[0,1]$. *As a consequence,* $\mathbf{T}$ *maps every corner of the square* $[0,1]^2$ *to itself.*

Proof. Recall that $\mathbf{T}=\nabla\Psi$ where Ψ is a convex function. Because f and g are positive densities bounded away from 0, by the Monge-Ampère equation (1.4), $\det(D^2\Psi)>0$ and thus Ψ is strictly convex. This yields

$$(\mathbf{T}(x)-\mathbf{T}(y))\cdot(x-y)=(\nabla\Psi(x)-\nabla\Psi(y))\cdot(x-y)>0, \tag{4.4}$$

for $x\neq y$, $x,y\in[0,1]^2$. We first proceed by contradiction to show that $T_1(0,x_2)=0$. Assume there is a point $(0,x_2)\in[0,1]^2$ for which $\mathbf{T}(0,x_2)=(y_1,y_2)$ and $y_1>0$. Then, because the optimal map $\mathbf{T}$ is invertible (hence bijective), there exists another point $(z_1,z_2)\in[0,1]^2$ which gets mapped to $(0,y_2)$. Using these two points in (4.4), we get

$$(\mathbf{T}(0,x_2)-\mathbf{T}(z_1,z_2))\cdot((0,x_2)-(z_1,z_2))=(y_1,0)\cdot(-z_1,x_2-z_2)=-y_1z_1>0,$$

which yields a contradiction because $y_1z_1\geq 0$ by assumption. We conclude that $T_1(0,x_2)=0$ for every $x_2\in[0,1]$. Using similar arguments, we can prove that the other sides of the square get mapped to themselves. ∎

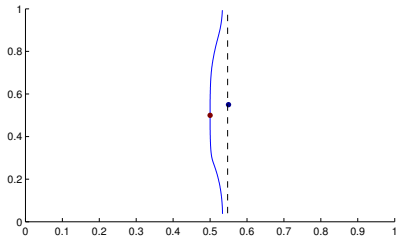
Proof. (Lemma 4.1.) Consider the line $\mathcal{L}=\{(x_1,x_2)\in[0,1]^2:x_1=T_1(\lambda_1,\lambda_2)\}$. Because $\mathbf{T}$ is continuous, the inverse image of that line by $\mathbf{T}$ has to be a continuous curve lying in $[0,1]^2$. Denote this curve by $\mathcal{C}$. From Lemma 4.2 and by the bijectivity of $\mathbf{T}$, we have that the inverse image of $(T_1(\lambda_1,\lambda_2),1)$ is on the line $[0,1]\times\{1\}$ and the inverse image of $(T_1(\lambda_1,\lambda_2),0)$ is on the line $[0,1]\times\{0\}$. Therefore, by continuity, for every $x_2\in[0,1]$ there exists a $x_1\in[0,1]$ such that $\mathbf{T}(x_1,x_2)\in\mathcal{L}$. Moreover, thanks to (4.4), for every fixed $x_2\in[0,1]$, $T_1(x_1,x_2)$ is a strictly increasing function of x_1. This implies that for any $x_2\in[0,1]$, there is a unique $x_1\in[0,1]$ such that $(x_1,x_2)\in\mathcal{C}$, which in turn proves the existence of a continuous function h_1 of x_2 such that $C=\{(x_1,x_2)\in[0,1]^2:x_1=h_1(x_2)\}$, that is $\mathcal{C}$ is the graph of h_1. In addition, when $x_2=\lambda_2$, $T_1(h(\lambda_2),\lambda_2)=T_1(\lambda_1,\lambda_2)$. This yields $h_1(\lambda_2)=\lambda_1$ because $x_1\mapsto T_1(x_1,\lambda_2)$ is a bijection on $[0,1]$, and statement a) is thus verified. For b), we have

$$\begin{aligned}
S_1&=\{(x_1,x_2)\in[0,1]^2:T_1(0,x_2)\leq T_1(x_1,x_2)\leq T_1(\lambda_1,\lambda_2)\}\\
&=\{(x_1,x_2)\in[0,1]^2:T_1(0,x_2)\leq T_1(x_1,x_2)\leq T_1(h_1(x_2),x_2)\}\\
&=\{(x_1,x_2)\in[0,1]^2:0\leq x_1\leq h_1(x_2)\}\\
&=\bigcup_{x_2\in[0,1]}\Big([0,h_1(x_2)]\times\{x_2\}\Big),
\end{aligned}$$

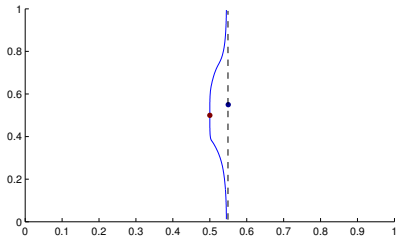
because $0=T_1(0,x_2)$. This concludes the proof. ∎

Note that by symmetry, we can derive similar results for μ_2, $T_2(\lambda_1,\lambda_2)$, and the set $S_2=\{(x_1,x_2)\in[0,1]^2:T_2(x_1,x_2)\leq T_2(\lambda_1,\lambda_2)\}$ with the function h_1 replaced by its analogue h_2.

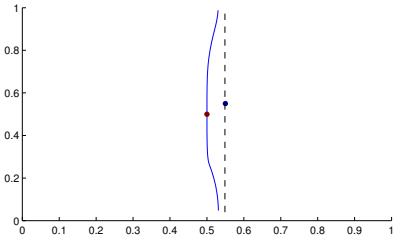
4.2. Numerical investigation of S_1. In order to better understand the behavior of the function h_1, and thus the set S_1, we present in figure 4.2 some examples of h_1 for different values of the parameters. These results were obtained by using the algorithm of Subsection 2.2 to find the optimal map $\mathbf{T}$ associated with the densities f and g given by (4.1). From figure 4.2, we observe that for values of x_2 close to λ_2,



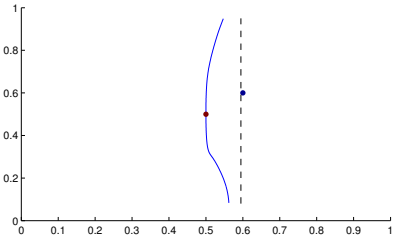
(a) $\lambda_1=\lambda_2=0.5$, $\mu_1=\mu_2=0.55$, $\sigma=1/12$, and $r=0.8$.



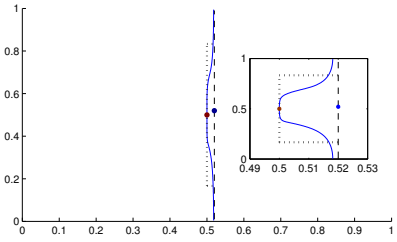
(b) Same parameters as (a), but $\sigma=1/24$.



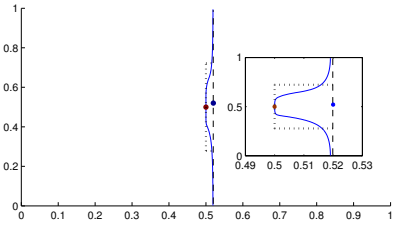
(c) Same parameters as (a), but $r=0.9$.



(d) Same parameters as (a), but $\mu_1=\mu_2=0.6$.



(e) $\lambda_1=\lambda_2=0.5$, $\mu_1=\mu_2=0.52$, $\sigma=1/24$, and $r=0.8$.



(f) Same parameters as (e), but $\sigma=1/36$.

FIG. 4.2. *Some examples of the function h_1, which is represented here by the solid line. The vertical axis is x_2 and the horizontal axis is x_1. The dots represent the initial and final position of the particle. The dashed lines represent $x_1=T_1(\lambda_1,\lambda_2)$. For (e) and (f), the vertical dotted line is the line connecting $\lambda_2-8\sigma$ to $\lambda_2+8\sigma$ and the two horizontal dotted lines link the bottom and top of the previous line segment to the line $x_1=T_1(\lambda_1,\lambda_2)$. In addition, for (e) and (f), a subplot is displayed to better visualize the figure.*

h_1 is essentially a vertical line, which means that $\mathbf{T}$ acts as a translation in a neighborhood of λ_2. The length of this neighborhood depends on the standard deviation σ and seems to suggest that the map $\mathbf{T}$ is a translation up to the point where the part of the mass that is due to the particle becomes negligible. In addition, outside this neighborhood of λ_2, the function h_1 appears to converge to another vertical line as x_2

goes to 0 or 1, and this line is close to the line $x_1 = \mathbf{T}(\lambda_1, \lambda_2)$. Note that this line lays very close to (and before) $x_1 = \mu_1$. This means that the map $\mathbf{T}$ is almost the identity map when taking points for which the Gaussian contribution to the densities f and g is negligible. This confirms the intuition that the optimal map translates the mass around the particle's location and leaves the rest of the domain (almost) unchanged. Accordingly, the plots in figure 4.2 suggest the following assertions:

1) $\mu_1 \geq T_1(\lambda_1, \lambda_2)$,
2) $h_1(x_2) \geq \lambda_1$ for any $x_2 \in [0,1]$,
3) There exist constants $C > 0$ and $\sigma_0 > 0$ such that $C\sigma_0 \leq \min\{\lambda_2, 1-\lambda_2\}$, and for every $\sigma < \sigma_0$, we have

$$T_1(\lambda_1, \lambda_2) - 2C\sigma(T_1(\lambda_1, \lambda_2) - \lambda_1) \leq \mathrm{m}(S_1).$$

In Subsection 4.3, we will assume that these properties hold to complete the proof of the linearity behavior of σ and $(1-r)/r$ in dimension two.

The first two are easily observable in plots (a) to (f) of figure 4.2, whereas the third one is a bit more technical and in order to better visualize it we included the dotted lines in plots (e) and (f). This third property actually states that the Lebesgue measure of the set

$$\tilde{S}_1 = \Big\{ (x_1, x_2) \in [0,1]^2 : 0 \leq x_1 \leq \lambda_1 \text{ if } \lambda_2 - C\sigma \leq x_2 \leq \lambda_2 + C\sigma \\ \text{and } 0 \leq x_1 \leq T_1(\lambda_1, \lambda_2) \text{ else} \Big\}$$

is smaller than $\mathrm{m}(S_1)$. The vertical dotted line in (e) and (f) therefore corresponds to the interval $[\lambda_2 - C\sigma, \lambda_2 + C\sigma]$ in x_2, where we took $C = 8$ as an example. We observe that in both cases, $\mathrm{m}(\tilde{S}_1)$ is indeed smaller than $\mathrm{m}(S_1)$. In the limiting case where $\sigma \to 0$, that is when the particle becomes a point mass (or Dirac delta function), the optimal plan only moves the mass at the particle's location and thus assertion 3 is trivially satisfied. Despite all these observations in favor of assertions 1, 2, or 3, obtaining a rigorous proof showing that they actually hold true for σ and $1-r$ small enough remains an open problem. However, intuitively we argue that all these assertions make sense, because of the optimality of the transport map $\mathbf{T}$. Moreover, we proved in the previous section that assertion 1 holds in dimension one, whereas 2 and 3 are trivial in this case.

4.3. Error estimates. We show here that, assuming properties 1 and 2 hold, we can get the linear behavior of the error with respect to $(1-r)/r$, as in Theorem 3.1. Moreover, if in addition we assume 3 holds, we recover the linear behavior in σ.

THEOREM 4.3. *Assume $2/3 < r \leq 1$, $0 < \epsilon \leq \lambda_1 \leq \mu_1 \leq 1-\epsilon$, and $0 < \epsilon \leq \lambda_2 \leq \mu_2 \leq 1-\epsilon$ for $0 < \epsilon < 0.5$, and $0 < \sigma < 1$. Let $\mathbf{T} = (T_1, T_2)$ be the optimal map solving the two-dimensional L^2 transport problem with densities f and g given by (4.1). Suppose in addition that properties 1 and 2 hold. Then we have the following inequality:*

$$\mu_1 - T_1(\lambda_1, \lambda_2) \leq C_1(\sigma)\frac{(1-r)}{r}(\mu_1 - \lambda_1) + C_2\frac{e^{-\frac{\epsilon^2}{2\sigma^2}}}{\sigma}(\mu_1 - \lambda_1), \tag{4.5}$$

where C_1 is a positive function of σ which is bounded by 2, and C_2 is a positive constant. Both C_1 and C_2 are independent of λ_1, λ_2, μ_1, μ_2, r, and ϵ. If in addition

we assume that 3 holds, then

$$C_1(\sigma) \leq 4C\sigma, \tag{4.6}$$

where C is the same constant as in assumption 3. A similar result holds for $\mu_2 - T_2(\lambda_1, \lambda_2)$.

Proof. We prove the result for $\mu_1 - T_1(\lambda_1, \lambda_2)$; by symmetry, similar arguments can be used for $\mu_2 - T_2(\lambda_1, \lambda_2)$. The starting point is equation (4.3). By Assumption 2, we have $\lambda_1 \leq h(x_2)$ for any $x_2 \in [0,1]$, so that

$$\begin{aligned}\int_0^1 \int_0^{\lambda_1} \exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\lambda_1)^2 + (x_2-\lambda_2)^2\big]\Big)\mathrm{d}x_1\,\mathrm{d}x_2 \\ \leq \int_0^1 \int_0^{h_1(x_2)} \exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\lambda_1)^2 + (x_2-\lambda_2)^2\big]\Big)\mathrm{d}x_1\,\mathrm{d}x_2 \\ = \int\int_{S_1} \exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\lambda_1)^2 + (x_2-\lambda_2)^2\big]\Big)\mathrm{d}\mathbf{x}.\end{aligned}$$

Moreover, using Lemma 4.1 and Assumption 2, we have $[0,\lambda_1]\times[0,1] \subset S_1$ and thus $\lambda_1 \leq \mathrm{m}(S_1)$. Combining these estimates with the fact that both k_μ and k_λ can be written as the product of two integrals, (4.3) implies

$$\begin{aligned}\mu_1 - T_1(\lambda_1,\lambda_2) \leq \Bigg[\frac{(1-r)}{r}(\mu_1-\lambda_1) + \frac{1}{k_{\mu_1}}\int_0^{\mu_1} \exp\Big(-\frac{1}{2\sigma^2}(x_1-\mu_1)^2\Big)\mathrm{d}x_1 \\ - \frac{1}{k_{\lambda_1}}\int_0^{\lambda_1} \exp\Big(-\frac{1}{2\sigma^2}(x_1-\lambda_1)^2\Big)\mathrm{d}x_1\Bigg] k_{\mu_1} \\ \left(k_{\mu_1}\frac{(1-r)}{r} + \int_0^1 \exp\big(-\frac{1}{2\sigma^2}t^2(T_1(\lambda_1,\lambda_2)-\mu_1)^2\big)\mathrm{d}t\right)^{-1}\end{aligned} \tag{4.7}$$

where k_{λ_1} and k_{μ_1} are defined by respectively replacing λ and μ by λ_1 and μ_1 in (3.2). Notice now that we can split the right-hand side of (4.7) into two terms, one for which bound (3.4) in Lemma 3.2 applies, and one for which the arguments displayed in (3.9) apply. This yields the existence of C_1 and C_2, as claimed. Finally, if in addition we assume that Assumption 3 holds, then

$$\begin{aligned}\mu_1 - m(S_1) &\leq \mu_1 - T_1(\lambda_1,\lambda_2) + (T_1(\lambda_1,\lambda_2)-\lambda_1)2C\sigma \\ &\leq \mu_1 - T_1(\lambda_1,\lambda_2) + (\mu_1-\lambda_1)2C\sigma.\end{aligned}$$

Using this in (4.3), we cancel the terms $(\mu_1 - T_1(\lambda_1,\lambda_2))(1-r)/r$ on both sides to obtain

$$\begin{aligned}(\mu_1 - T_1(\lambda_1,\lambda_2)) \times \frac{1}{k_\mu}\int_0^1\int_0^1 \exp\Big(-\frac{1}{2\sigma^2}\big[t^2(T_1(\lambda_1,\lambda_2)-\mu_1)^2 + (x_2-\mu_2)^2\big]\Big)\mathrm{d}t\,\mathrm{d}x_2 \\ \leq \frac{1}{k_\mu}\int_0^1\int_0^{\mu_1} \exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\mu_1)^2 + (x_2-\mu_2)^2\big]\Big)\mathrm{d}x_1\mathrm{d}x_2 \\ - \frac{1}{k_\lambda}\int\int_{S_1} \exp\Big(-\frac{1}{2\sigma^2}\big[(x_1-\lambda_1)^2 + (x_2-\lambda_2)^2\big]\Big)\mathrm{d}\mathbf{x} + \frac{(1-r)}{r}(\mu_1-\lambda_1)2C\sigma.\end{aligned}$$

We see that this previous inequality is very close to (4.7), provided we remove the term $k_{\mu_1}(1-r)/r$ in the denominator. Even if this term is not removed, we can still

get the same upper bounds as in Lemma 3.2 and as in the arguments in (3.9), which yields the result. ∎

We will present in Subsection 5.3 numerical simulations validating the linear behavior with respect to $(1-r)/r$ and σ, for the case where we consider multiple particles in dimension two. The results will be similar to the results in the one-dimensional case, presented in Subsection 3.2. Finally, it is worth mentioning that even though we restricted ourselves to dimension 2, all the arguments presented can be employed to get the same results in any dimension.

5. Numerical experiments using synthetic flows

In this section, we present numerical simulations in 2D, with multiple particles. We consider two synthetic flow velocities: one is a simple plane shear and the other consists of an array of vortices. We use the algorithm of Subsection 2.2 to solve the associated PIV problem for Δt small enough so that we remain in the realm of PTV. These simulations constitute a first step towards analyzing the potential of the OT framework for PIV and our plan is to investigate in the future more involved algorithms to treat real data and more PIV-like images. This is followed in Subsection 5.3 by the validation of the error behavior of the procedure with respect to the different parameters involved for the vortices example, as it was done in Subsection 3.2 for the one-dimensional case. As mentioned in Subsection 2.1, we take $(M-1)^2$ particles with evenly distributed initial positions and we use the MATLAB ode45 procedure to find their corresponding final positions due to their displacement by the synthetic flow, at time $t=\Delta t$. From the change of variables $\mathbf{T}(\mathbf{x})=\mathbf{x}+\nabla u(\mathbf{x})$ introduced in Subsection 2.2, we have $\mathbf{v}_{\text{approx}}=\nabla u/\Delta t$. We use a tolerance $\text{TOL}_{\text{GMRES}}=10^{-1}$ for the GMRES algorithm employed to solve the linear equation at every Newton step, and we took the tolerance $\text{TOL}_{\text{Newton}}=10^{-6}$ for the main Newton algorithm. Note that taking $\text{TOL}_{\text{GMRES}}=10^{-1}$ significantly decreases the number of computations, and even in the cases presented below where the algorithm did not converge, using a smaller tolerance (e.g. $\text{TOL}_{\text{GMRES}}=10^{-2}$ or $\text{TOL}_{\text{GMRES}}=10^{-4}$) did not change the outcome. To increase the speed of the algorithm, we use different values for the damping factor according to the precision reached in the stopping criterion. We select an initial value τ_i of τ sufficiently large to guarantee that the damped Newton algorithm proceeds in the right direction. We change this value to a smaller τ_f after the stopping criterion reaches a certain threshold, denoted τ_{t}. From the experiments conducted here, we found that $\tau_{\text{t}}=10^{-2}$ was typically a good value to take and we employed it to obtain all the results presented in this work. We present two different types of errors, namely the total relative error e_{total} between $\mathbf{v}$ and $\mathbf{v}_{\text{approx}}$ defined by

$$e_{\text{total}} := \frac{\|\mathbf{v}-\mathbf{v}_{\text{approx}}\|_{l^2_{\text{norm}}}}{\|\mathbf{v}\|_{l^2_{\text{norm}}}}$$

and the relative error due to the direction e_{dir} between $\mathbf{v}$ and $\mathbf{v}_{\text{approx}}$ defined by

$$e_{\text{dir}} := \frac{\|\tilde{\mathbf{v}}-\tilde{\mathbf{v}}_{\text{approx}}\|_{l^2_{\text{norm}}}}{\|\tilde{\mathbf{v}}\|_{l^2_{\text{norm}}}},$$

where $\tilde{\mathbf{v}}$ is the locally normalized vector field

$$\tilde{\mathbf{v}}(x_1,x_2) := \frac{\mathbf{v}(x_1,x_2)}{\|\mathbf{v}(x_1,x_2)\|_{l^2}},$$

if $\tilde{\mathbf{v}}(x_1,x_2)\neq 0$, for any $(x_1,x_2)\in[0,1]^2$ ($\tilde{\mathbf{v}}_{\text{approx}}$ is defined similarly). The error e_{dir} will be useful to demonstrate that most of the error in our procedure is in the length of the individual vectors, and not in their directions.

5.1. Plane shear flow. For this first experiment, we take the vector field to be

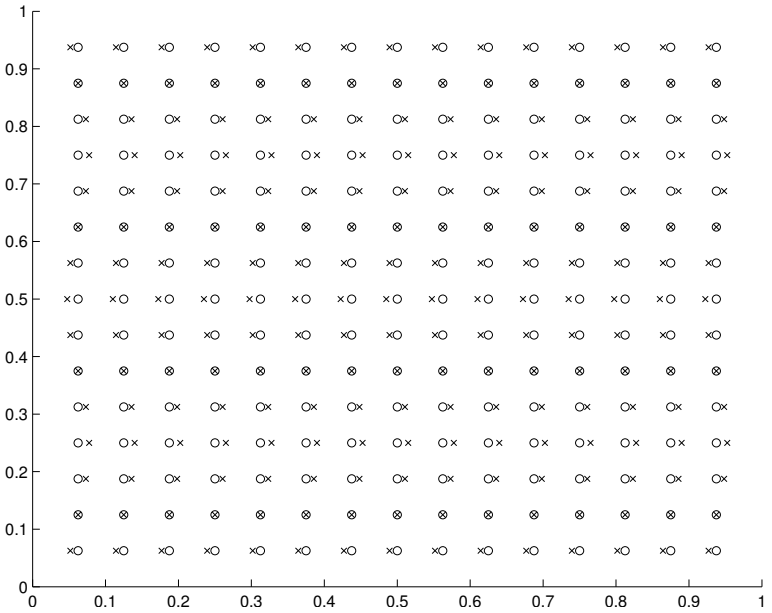
$$\mathbf{v}(x_1,x_2,t)=(-\cos(2\pi\gamma x_2),0),$$

where $\gamma=2$. Our goal is to investigate the behaviour of the approximation with respect to the different parameters involved. We invite the reader to keep in mind that for real data, these parameters would be dictated by the PIV image (whether some pre-processing was applied to it or not) and thus it would not necessarily be possible to vary them at the user's leisure. We first fix $r=0.8$ and we vary both σ and Δt. The results are presented in table 5.1. We observe that the relative

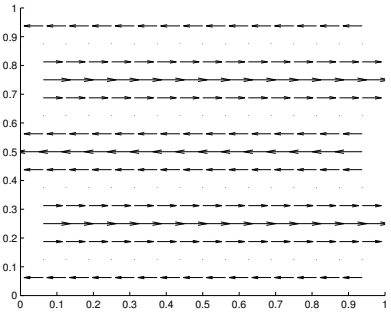
M=16, r=0.8			e_{total}			e_{dir}		
			Δt			Δt		
			0.01	0.005	0.0025	0.01	0.005	0.0025
s=1	N	512	0.2493	0.2267	0.2201	0.0239	0.0203	0.0192
($\sigma=1/96$)		1024	0.2496	0.2268	0.2202	0.0239	0.0203	0.0192
s=1.5	N	512	0.0899	0.0819	0.0806	0.0046	0.0048	0.0076
($\sigma=1/144$)		1024	0.0902	0.0820	0.0806	0.0045	0.0048	0.0076
s=2	N	512	0.0455	0.0441	0.0438	0.0036	0.0027	0.0046
($\sigma=1/192$)		1024	0.0476	0.0444	0.0438	0.0018	0.0025	0.0046

TABLE 5.1. *Some results for the shear experiment. In the first part of this table, we display the relative error between* $\mathbf{v}_{approx}$ *and* $\mathbf{v}$ *for different values of* s*. In the second part, we display the relative error for* $\tilde{\mathbf{v}}_{approx}$ *and* $\tilde{\mathbf{v}}$*.*

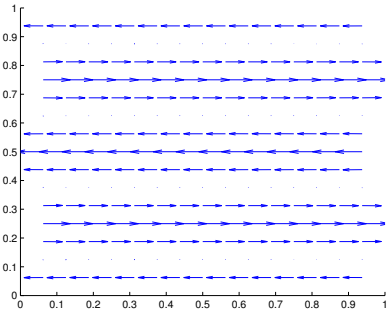
error e_{tol} between $\mathbf{v}_{\text{approx}}$ and $\mathbf{v}$ decreases as we decrease σ (increase s), and that it seems to converge to a constant as $\Delta t\to 0$. As we will see below, this is to be expected because the error in the position of the centers converges to 0 linearly with respect to both σ and $\lambda_2-\lambda_1$. Simulations in Subsection 5.3 will further validate this claim. The fairly large grid size employed is required for the algorithm to converge for large values of s. Indeed, if we consider that a particle's physical width spans over approximately 6 standard deviations (6σ), then for $s=1$ ($\sigma=1/96$), taking $N=128$ gives about 8 grid points per particle and this was sufficient to reduce the error to the desired tolerance. However, for $s=2$ ($\sigma=1/192$), the algorithm did not converge for $N=128$ (4 points per particles) nor $N=256$ (8 points per particle) when $\Delta t=0.01$. We therefore selected a higher ratio of 16 ($N=512$) or 32 ($N=1024$) grid points per particle to ensure a good enough resolution so that the variations of the density are properly captured by the underlying mesh. For the locally normalized vector fields, the relative error e_{dir} is much smaller than e_{total}, which means that most of the error is in the length of the individual vectors, as previously stated. We can visualize this by looking at a specific case in figure 5.1, where one example of the field $\mathbf{v}_{\text{approx}}$ is presented beside the target field $\mathbf{v}$. Of course, this could also be due to the particular nature of the vector field considered in the current experiment, and thus the vortices experiment presented in the next Subsection will tell us more about this. Note also that the maximum time interval selected for this experiment was such that the particles did not move much from their initial positions. We will look at a case where the tracers move further from their initial position in the experiment with vortices. Observe next that the relative error for the locally normalized vector fields



(a) Initial and final distributions of particles. The circles represent the initial positions and the crosses the final positions. Note that the standard deviations are not represented on this plot.



(b) Target vector field v.



(c) Approximate vector field v_{approx}.

FIG. 5.1. *Results for the shear, when* $M=16$, $s=2$ *(*$\sigma=1/192$*),* $r=0.8$, $N=1024$, *and* $t=0.01$. *The vector fields were scaled in order to better visualize these results. The maximum magnitude is* 1 *for the target vector field and* 0.9569 *for the approximate vector field.*

e_{dir} also decreases with σ, but ceases to decrease beyond a certain level with respect to Δt for a fixed σ. This can be explained by the fact that for the specific values of Δt taken, the relative error due to the finite differences approximation is negligible, but increasing. More specifically, for $\Delta t=0.01$, $\Delta t=0.005$, and $\Delta t=0.0025$, this error is respectively 2.4×10^{-14}, 5.8×10^{-14}, and 1.6×10^{-13} (for comparison, the machine epsilon used was 2.2×10^{-16}). In fact, we can understand the effect of both errors through the following inequalities (for one particle):

$$|v_1 - v_{\text{approx}1}| = \left| v_1 - \frac{T_1(\lambda_1)-\lambda_1}{\Delta t} \right| = \left| v_1 - \frac{\mu_1-\lambda_1}{\Delta t} + \frac{\mu_1 - T_1(\lambda_1)}{\Delta t} \right|$$

M=16, s=1.5 ($\sigma=1/144$)			e_{total}			$\|\mathbf{T}(\boldsymbol{\lambda})-\boldsymbol{\mu}\|_{l^2_{\text{norm}}}$		
			Δt			Δt		
			0.01	0.005	0.0025	0.01	0.005	0.0025
r=0.825	N	512	0.0816	0.0731	0.0713	5.58e-04	2.50e-04	1.22e-04
		1024	0.0818	0.0732	0.0714	5.59e-04	2.50e-04	1.22e-04
r=0.85	N	512	0.0726	0.0639	0.0628	4.96e-04	2.18e-04	1.07e-04
		1024	0.0728	0.0640	0.0629	4.97e-04	2.19e-04	1.07e-04
r=0.875	N	512	0.0637	0.0551	0.0542	4.35e-04	1.88e-04	9.25e-05
		1024	0.0643	0.0551	0.0542	4.39e-04	1.88e-04	9.25e-05

TABLE 5.2. *Some results for the shear experiment. In the first part of this table, we display the relative error between* $\mathbf{v}_{approx}$ *and* $\mathbf{v}$ *for different values of* r*. In the second part, we display* $\|\mathbf{T}(\boldsymbol{\lambda})-\boldsymbol{\mu}\|_{l^2_{norm}}$*, where* $\boldsymbol{\mu}$ *is a vector containing all* $\boldsymbol{\mu}_{ij}$ *for every particle* ij *and* $\mathbf{T}(\boldsymbol{\lambda})$ *is a vector containing all* $\mathbf{T}(\boldsymbol{\lambda}_{ij})$ *for every particle* ij*.*

$$\begin{aligned}&\leq \left| v_1 - \frac{\mu_1-\lambda_1}{\Delta t}\right| + C(r,\sigma)\left|\frac{\mu_1-\lambda_1}{\Delta t}\right| \\ &\leq \mathcal{O}(\Delta t) + C(r,\sigma)\|v_1\|_\infty,\end{aligned}$$

where $C(r,\sigma)$ is the bound given in Theorem 4.3, and v_1 and $v_{\text{approx}\,1}$ are respectively the first components of the target vector field $\mathbf{v}$ and the approximate vector field $\mathbf{v}_{\text{approx}}$ (a similar bound can be obtained for the second components of the vectors). We see that we are in a regime where the velocity error is dominated by the effect of σ and r alone, because the term in $\mathcal{O}(\Delta t)$ becomes negligible for Δt small enough, which is the case here. We will present a different regime for which the finite differences error is significant in Subsection 5.2.

Let us now look at results for which only r is varying; see table 5.2. The relative error also decreases as r increases and as Δt decreases. In the second part of this table, we display the total error in the final position of the particles, that is the $\|\cdot\|_{l^2_{\text{norm}}}$ norm of the vector containing the errors $\boldsymbol{\mu}_{ij}-\mathbf{T}(\boldsymbol{\lambda}_{ij})$ for every particle ij. We observe that this error in the final position also decreases as r increases and as Δt decreases, as expected. Note finally that all the results obtained in table 5.1 and table 5.2 were obtained using a multiresolution approach, where for example in table 5.1, the final solution obtained for $s=1$ was taken to be the starting solution for $s=1.5$. In every case, we took $\tau_i=3$ and $\tau_f=1$. It took about 11 hours in total to compute all three approximate vector fields corresponding to $s=1$, $s=1.5$, and $s=2$ in the case where $N=1024$, $r=0.8$, and $\Delta t=0.01$. We used a single processor of a 2.67-GHz Xeon x5550 core with 24 GB of RAM. At this stage, the current implementation of the algorithm is still serial and has not been optimized in terms of speed, so the computing time remains fairly high. We will discuss this further in the conclusion.

5.2. Array of vortices. For the second experiment, we take the vector field to be

$$v(x_1,x_2,t)=(\sin(2\pi\gamma_1 x_1),\cos(2\pi\gamma_2 x_2)+\pi/2),$$

where $\gamma_1=2$ and $\gamma_2=4$. In table 5.3 and figure 5.2 we present the results of similar experiments as the ones presented in the previous case. We observe a similar behavior for e_{total} as for the shear experiment. However, several differences arise. First, e_{dir} now decreases with Δt as opposed to the previous case. This can be explained by the fact that the finite differences relative error is now 0.0821, 0.0406, and 0.0202 for $\Delta t=0.01$, $\Delta t=0.005$, and $\Delta t=0.0025$, respectively. Observe also that e_{total} is

close to this finite differences error, which means that our algorithm gives good results. The error in direction e_{dir} is now a bigger portion of the relative error, but as s increases and Δt decreases, this portion decreases significantly. Second, in the case where $N=512$ and $\Delta t=0.01$, the algorithm did not converge anymore for $s=1.5$ and $s=2$. In both cases, we let the Newton algorithm evolve even if the GMRES algorithm did not converge to a tolerance of $\text{TOL}_{\text{GMRES}}=0.1$ on the relative error for some iterates (the maximum number of iterations was set to 50 for GMRES). We then stopped the algorithm when this relative residual in the GMRES algorithm became bigger than 0.99, that is when the solution computed is almost the same as the one computed in the previous iteration. This happened after $n=27$ Newton iterations for $s=1.5$ and after $n=24$ for $s=2$. In the former case, we reduced the errors to $\|f_n-f\|_{l^2_{\text{norm}}}=0.0041$, $\|u_n-u_{n-1}\|_{l^2_{\text{norm}}}=7.0084e-08$, and $\|f_n-f\|_{l^2_{\text{norm}}}/\|f\|_{l^2_{\text{norm}}}=3.2496e-06$, whereas we got $\|f_n-f\|_{l^2_{\text{norm}}}=0.0327$, $\|u_n-u_{n-1}\|_{l^2_{\text{norm}}}=3.0946e-07$, and $\|f_n-f\|_{l^2_{\text{norm}}}/\|f\|_{l^2_{\text{norm}}}=1.2206e-04$ for the latter case.

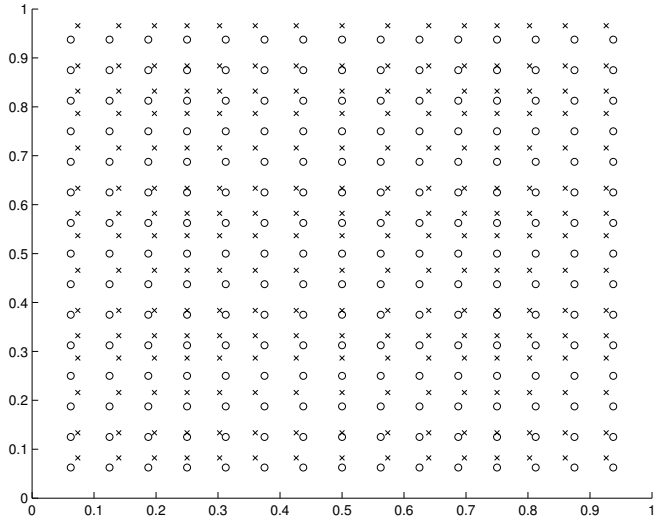
M=16, r=0.8			e_{total}			e_{dir}		
			Δt			Δt		
			0.01	0.005	0.0025	0.01	0.005	0.0025
s=1	N	512	0.2513	0.1957	0.1803	0.1685	0.1156	0.1018
($\sigma=1/96$)		1024	0.2523	0.1959	0.1803	0.1697	0.1157	0.1018
s=1.5	N	512	0.1218*	0.0865	0.0743	0.0522*	0.0321	0.0240
($\sigma=1/144$)		1024	0.1296	0.0867	0.0744	0.0591	0.0322	0.0240
s=2	N	512	0.0888*	0.0563	0.0450	0.0376*	0.0200	0.0122
($\sigma=1/192$)		1024	0.0980	0.0593	0.0454	0.0400	0.0206	0.0123

TABLE 5.3. *Some results for the vortices experiment. In the first part of this table, we display the relative error between* $\mathbf{v}_{approx}$ *and* $\mathbf{v}$ *for different values of* s*. In the second part, we display the relative error for the locally normalized* $\mathbf{v}_{approx}$ *and* $\mathbf{v}$*. If a * is displayed, it means that the algorithm did not converge to the given tolerance (it stagnated).*

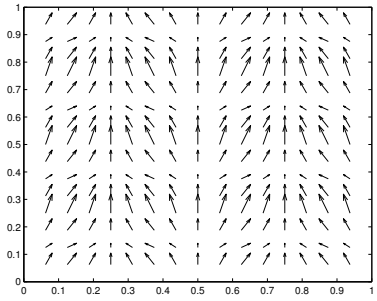
M=32 N=1024	e_{total}			$\|u_n-u_{n-1}\|_{l^2_{\text{norm}}}$		
	Δt			Δt		
	0.01	0.005	0.0025	0.01	0.005	0.0025
s=1, r=0.8	0.9558*	0.2752	0.2129	1.92e-08	7.43e-12	3.82e-12
s=1, r=0.85	0.9625*	0.2544	0.1878	2.32e-07	3.77e-12	1.60e-11
s=1.5, r=0.8	1.0005*	0.2138*	0.0852	2.74e-06	6.48e-08	1.08e-11
s=1.5, r=0.85	1.0011*	0.2039*	0.0680	1.70e-06	1.81e-06	2.68e-10

M=32 N=1024	$\|f_n-f\|_{l^2_{\text{norm}}}$			$\frac{\|f_n-f\|_{l^2_{\text{norm}}}}{\|f\|_{l^2_{\text{norm}}}}$		
	Δt			Δt		
	0.01	0.005	0.0025	0.01	0.005	0.0025
s=1, r=0.8	0.0012	6.50e-07	7.13e-07	6.14e-07	1.82e-13	2.20e-13
s=1, r=0.85	0.0677	8.67e-07	5.00e-07	0.0018	3.02e-13	1.00e-13
s=1.5, r=0.8	0.4408	0.0655	7.10e-07	0.0408	9.01e-04	1.06e-13
s=1.5, r=0.85	0.5505	0.1312	9.46e-07	0.0578	0.0033	1.70e-13

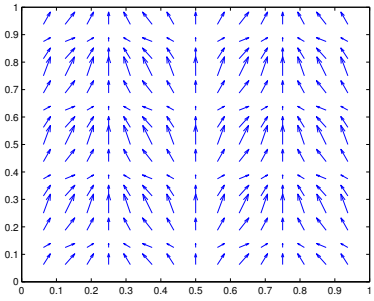
TABLE 5.4. *Some results for the vortices experiment for the case* $M=32$ *and* $N=1024$*. If a * is displayed, it means that the algorithm did not converge to a tolerance of* $TOL_{Newton}=10^{-6}$ *in the stopping criterion (it stagnated). We also present in every case several examples of the norms involved in the stopping criterion, where in each case* n *is the number of Newton iterations required to either reach the tolerance or stagnate. Note that* $s=1$ *and* $s=1.5$ *correspond to* $\sigma=1/192$ *and* $\sigma=1/288$*, respectively.*



(a) Initial and final distributions of particles. The circles represent the initial positions and the crosses the final positions. Note that the standard deviations are not represented on this plot.



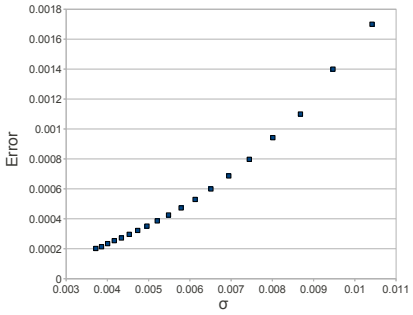
(b) Target vector field.



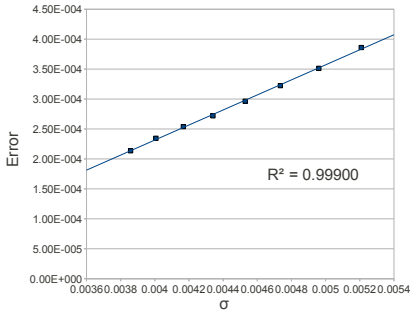
(c) Approximate vector field.

FIG. 5.2. *Results for the vortices, when* $M=16$, $s=2$ *(*$\sigma=1/192$*),* $r=0.8$, $N=1024$, *and* $t=0.01$. *The vector fields were scaled in order to better visualize these results. The maximum magnitude is* 2.7584 *for the target vector field and* 2.5722 *for the approximate vector field.*

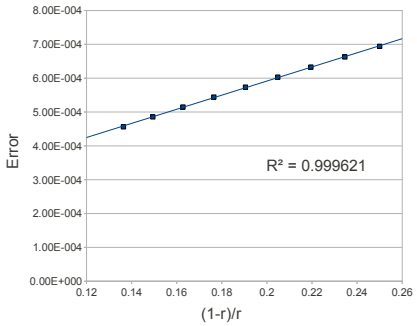
Nevertheless, as we can see in the table, the relative error e_{total} is quite close to the one for $N=1024$. Therefore, we realize that a tolerance of $\text{TOL}_{\text{Newton}}=10^{-6}$ on the stopping criterion (2.5) might be too stringent and that acceptable results could be obtained with fewer iterations. In addition, the number of grid points required to get convergence could be large for larger values of s, and thus a mesh adaptation implementation would be better suited in order to increase the performances of our method. We shall come back to this later. We also point out that varying the tolerance originally set to $\text{TOL}_{\text{GMRES}}=10^{-1}$ for the GMRES algorithm does not prevent this stagnation from occurring. As for the computing times, it took on the same CPU as the one previously mentioned about 33 hours to compute the vector fields associated with $s=1$, $s=1.5$, and $s=2$, using a multiresolution approach in the case where $r=0.8$, $N=1024$, and $\Delta t=0.01$. This longer computing time can be explained by the fact that compared to the case of the shear experiment, the optimal solution is



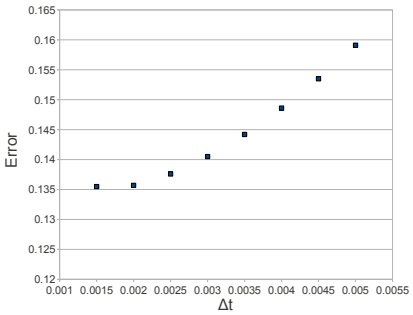
(a) $\|\mathbf{T}(\boldsymbol{\lambda})-\boldsymbol{\mu}\|_{l^2_{\text{norm}}}$ as a function of $\sigma=1/(6Ms)$ for $N=1024$, $M=16$, $r=0.8$, and $\Delta t=0.005$.



(b) Tail of the error graph 5.3 (a).



(c) $\|\mathbf{T}(\boldsymbol{\lambda})-\boldsymbol{\mu}\|_{l^2_{\text{norm}}}$ as a function of $(1-r)/r$ for $N=1024$, $M=16$, $s=1.5$, and $\Delta t=0.005$.



(d) $\|\mathbf{v}-\mathbf{v}_{\text{approx}}\|_{l^2_{\text{norm}}}$ as a function of Δt for $N=1024$, $M=16$, $s=1.5$, and $r=0.8$.

FIG. 5.3. *Numerical validation of the error analysis in dimension two. In the cases where the data follows a linear trend, the linear regression line passing through the data is included, with the corresponding coefficient of determination.*

further from the initial guess of $\mathbf{0}$ by the Newton algorithm.

The only parameter we have not varied so far is the number of tracers M. In table 5.4, we present some results for $M=32$, or 31^2 particles. We fix the grid size to $N=1024$. From the table, we see that the procedure did not converge in any case when $\Delta t=0.01$ or for two cases for $\Delta t=0.05$. To see how far the algorithm was from reaching the tolerance of $\text{TOL}_{\text{Newton}}=10^{-6}$, we also presented in table 5.4 the value of some errors involved for the last iterate. Due to this higher number of particles, it appears that a larger grid size is required in these situations, which gives more arguments in favour of an adaptive mesh implementation. Note that it would also be possible to reduce the computational cost of the algorithm by splitting the region containing the tracers into smaller interrogation windows and by applying the OT algorithm on each of them.

5.3. Numerical validation of 2D error estimates. Here, we investigate in more details the error behaviour with respect to the parameters r, σ, and Δt for the case of multiple particles in 2D. For simplicity, we consider only the case of vortices when $M=16$. The results are shown in figure 5.3. As we see in figure 5.3 (c), the error

decreases linearly with respect to $(1-r)/r$. For the standard deviation, figure 5.3 (a) depicts that for the larger values on the graph, we are in an exponential regime. Then, as σ decreases, we enter a linear regime, as seen in figure 5.3 (b), which is consistent with the results of Theorem 4.3. Finally, we displayed in figure 5.3 (d) the error as a function of Δt. We see that, because the grid size was fixed to $N=1024$, as Δt decreases, the displacements of the tracers are not all detected by the grid. The vector field becomes under-resolved and the error stagnates. Before this transition, e_{total} decreases with Δt, but not linearly. This is expected because the trajectories in the flow are not linear.

6. Concluding remarks

We introduced in this work a method for approximating the velocity field from measurements obtained by particle image velocimetry using optimal transportation. Even though we investigated only the case where the average displacement of tracers is smaller than about half of the average distance between them in our numerical experiments, this work gives good reasons to believe that more involved algorithms based on similar principles could be successful in recovering the velocity field associated to general PIV problems. The technique presented has several advantages: it can be used on problems requiring any number of physical dimensions; it is a global approach and thus there is no need to split the domain into several interrogation windows; it can be applied to virtually any distribution of tracers; and finally it appears that most of the error introduced by this method is concentrated in the length of the vectors, not in their directions. In addition, the very general framework provided by the optimal transportation theory could provide even more accurate methods. Indeed, we took here the cost function c to be the square of the Euclidean distance. This particular distance is easier to deal with due to the nice convexity properties of the solution, but other costs might be better suited for specific situations (depending on the model taken for the densities or the properties of the fluid under consideration).

The two major difficulties we found for this method are the fact that the optimal map $\mathbf{T}$ does not map the particle's location exactly to its target location (for the specific model we used) and the fact that in its current stage of development, it is still quite computationally demanding. However, by modelling the uncertainty in the measurement of a particle's location using a Gaussian distribution with standard deviation σ and the noise in the light measurements by $1-r$, we proved that this error in the position decreases linearly with respect to both parameters. We also saw that the error between the target and approximate vector fields is limited mostly by the finite differences used to approximate the velocity using the optimal map.

Using smaller values of σ and bigger values of r significantly reduced the error in position and therefore we can think that for a small uncertainty in the measurement of the particle's location and a low noise level, this position error would become negligible. Getting the algorithm to converge in such a case would require a very high grid size which could be difficult to achieve with a uniform grid. Due to the nature of the densities f and g, when s is big enough (that is, the standard deviation σ is small enough) many grid points are not in fact needed for the computations. Indeed, as about 99.7% of the data is within three standard deviations of the mean in a normal distribution, outside circles of radius 3σ centered at the particle's location, the value of f (or g) is almost constant with value $1-r$, and thus we don't need to keep track of all the related grid points. For example, in the case where $M=16$ and $s=1$, the total area covered by these circles corresponds to about 69% of the area of the domain $[0,1]^2$, whereas when $M=16$ and $s=2$, this percentage decreases to about

17.3%. Therefore, we see that a mesh adaptation method would significantly increase the efficiency of the algorithm for smaller standard deviations by laying a grid with a larger concentration of points inside these circles. It is also worth mentioning that we employed a serial algorithm and it is clear that some parts of the procedure could be computed much faster using a parallel implementation.

The model based on Gaussian distributions employed in this work was taken primarily because it provides a realistic model for simple PIV or PTV situations which is at the same time compatible with the selected optimal transport numerical algorithm. Several procedures available in the literature to solve the optimal transport map require the densities to be bounded away from 0, which makes the parameter r necessary. Nevertheless, the mathematical analysis we presented is independent of the choice of algorithm and explains the effect of a uniformly distributed noise and of the uncertainty in the measurement of the tracer particles. For more complicated PIV data in which the distributions of tracers are much more random, one way to use our model would be to first apply a clustering algorithm to identify defining clusters of particles. Then, one could find the centroid of the clusters and use it for the center of the Gaussians. The standard deviation could be obtained by computing the average distance from the boundary of every cluster to its center. One significant issue to investigate for this procedure would occur when a cluster representing several tracers splits into several clusters on the image. In subsections 5.1 and 5.2, the time-scales and particle sizes taken are usually considered more for PTV than PIV. However, in the near future, our goal is to apply the method we presented in this paper to real PIV data and then further investigate different strategies to approximate the velocity field.

At last, there are a few mathematical open problems left to solve in order to complete the proof of the position error's behaviour in dimension two. As we saw in Section 4, we had to introduce three assumptions to prove the linear convergence in dimension 2. Nevertheless, we provided some numerical evidence of its validity for the more complex case of multiple particles.

Acknowledgment. The authors wish to mention that all the computations were done using the Nestor cluster from the Westgrid consortium, which is part of Compute Canada. In addition, the authors are supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] R.J. Adrian, *Twenty years of particle image velocimetry*, Exper. Fluids, 39, 159–169, 2005.
[2] J.D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84, 375–393, 2000.
[3] J.D. Benamou, B. Froese, and A. Oberman, *Numerical solution of the optimal transportation problem using the Monge-Ampère equation*, J. Comput. Physics, 260, 107–126, 2014.
[4] D. Bertsekas, *The auction algorithm: A distributed relaxation method for the assignment problem*, Ann. Op. Res., 14(1), 105–123, 1988.
[5] Y. Brenier, *Décomposition polaire et réarrangement monotone de champs de vecteurs*, C.R. Acad. Sci. Paris Ser. I, 305, 19, 805–808, 1987.
[6] M. Burger, M. Franek, and C.B. Schönlieb, *Density estimation, Smoothing based on Optimal Transport*, Applied Mathematics Research Express, abs007, 45, DOI:10.1093/amrx/abs007, 2012.
[7] D. Cordero-Erausquin, *Sur le transport de mesures périodiques*, C.R. Acad. Sci. Paris Ser. I, 329, 199–202, 1999.
[8] R. Chartrand, B. Wholberg, K. Vixie, and E. Bollt, *A gradient descent solution to the Monge-Kantorovich problem*, Appl. Math. Sci., 3, 21-24, 1071–1080, 2009.

[9] G.E. Elsinga, F. Scarano, B. Wieneke, and B.W. van Oudheusden, *Tomographic particle image velocimetry*, Exper. Fluids, 41, 933–947, 2006.
[10] S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, and J.F. Aujol, *Regularized discrete optimal transport*, in Scale Space and Variational Methods in Computer Vision, Springer Berlin Heidelberg, 428–439, 2013.
[11] D. Goldfarb, *Efficient dual simplex algorithms for the assignment problem*, Math. Prog., 33(2), 187–203, 1985.
[12] E. Haber, T. Rehman, and A. Tannenbaum, *An efficient numerical method for the solution of the L^2 optimal mass transfer problem*, SIAM J. Sci. Comput., 32(1), 197–211, 2010.
[13] R.D. Keane and R.J. Adrian, *Theory of cross-correlation analysis of PIV images*, Appl. Sci. Res., 49, 191–215, 1992.
[14] H.W. Kuhn, *The Hungarian method for the assignment problem*, Naval Res. Logist. Quart., 2, 83–97, 1955.
[15] G. Loeper and F. Rapetti, *Numerical solution of the Monge-Ampère equation by a Newton's algorithm*, C.R. Acad. Sci. Paris Ser. I, 340, 319–324, 2005.
[16] R.J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128, 153–179, 1997.
[17] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, 666–704, 1781.
[18] O. Museyko, M. Stiglmayr, K. Klamroth, and G. Leugering, *On the application of the Monge-Kantorovich problem to image registration*, SIAM J. Imaging Sci., 2, 1068–1097, 2009.
[19] J. Rabin, G. Peyré, J. Delon, and M. Bernot, *Wasserstein barycenter and its application to texture mixing*, in Scale Space and Variational Methods in Computer Vision, Springer Berlin Heidelberg, 435–446, 2012.
[20] M. Raffel, C.E. Willert, S.T. Wereley, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide*, Springer, Heidelberg New York, 2007.
[21] L.P. Saumier, *An Efficient Numerical Algorithm for the L^2 Optimal Transport Problem with Applications to Image Processing*, M.Sc. Thesis, University of Victoria, Canada, 2010.
[22] L.P. Saumier, M. Agueh, and B. Khouider, *An Efficient Numerical Algorithm for the L^2 Optimal Transport Problem with Periodic Densities*, IMA J. Appl. Math. DOI: 10.1093/imamat/hxt032, 2013.
[23] J. Strain, *Fast spectrally-accurate solution of variable-coefficients elliptic problems*, Proc. A.M.S., 122(3), 843–850, 1994.
[24] A. Dominitz and A. Tannenbaum, *Texture Mapping via optimal mass transport*, IEEE Trans. Vis. Comp. Graphics, 16, 419–433, 2010.
[25] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, AMS, Providence, USA, 58, 2003.
[26] C. Villani, *Optimal Transport, Old and New*, Grundlehren der Mathematischen Wissenshaften, Springer-Verlag Berlin Heidelberg, 338, 2009.
[27] J. Westerweel, *Digital Particle Image Velocimetry Theory and Application*, Ph.D. Dissertation, Delft University Press, 1993.
[28] Y. Zemel, *Optimal Transportation: Continuous and Discrete*, Master's Thesis, EPFL, 2012.