Entropic Optimal Transport Solutions of the Semigeostrophic Equations

Jean-David Benamou, Colin Cotter, Hugo Malamut

Entropic Optimal Transport Solutions of the Semigeostrophic Equations

J.-D. Benamou, C. Cotter, H. Malamut

[a] *INRIA, rue Simone Iff, Paris, 75012, , France*
[b] *Department of Mathematics, Imperial College London, South Kensington Campus, London, SW7 2AZ, , United Kingdom of Great Britain and Northern Ireland*
[c] *ENS, rue Simone Iff, Paris, 75012, , France*

Preprint submitted to *January 27, 2023*

Entropic Optimal Transport Solutions of the Semigeostrophic Equations

J.-D. Benamou, C. Cotter, H. Malamut

[d] *INRIA, rue Simone Iff, Paris, 75012, , France*
[e] *Department of Mathematics, Imperial College London, South Kensington Campus, London, SW7 2AZ, , United Kingdom of Great Britain and Northern Ireland*
[f] *ENS, rue Simone Iff, Paris, 75012, , France*

Abstract

The Semigeostrophic equations are a frontogenesis model in atmospheric science. Existence of solutions both from the theoretical and numerical point of view is given under a change of variable involving the interpretation of the pressure gradient as an Optimal Transport map between the density of the fluid and its push forward. Thanks to recent advances in numerical Optimal Transportation, the computation of large scale discrete approximations can be envisioned. We study here the use of Entropic Optimal Transport and its Sinkhorn algorithm companion.

Keywords:
2000 MSC: 78A46, 49Q22, 65K10

1. Introduction

The Semigeostrophic (SG) model provided the first mathematical explanation of frontogenesis, which is the process by which fronts (discontinuities in temperature and velocity) emerge in the atmosphere out of smooth conditions. The equations approximates the slow large scale dynamics of rotating fluids, without explicitly representing rapidly oscillating fast internal waves. This model is valid in the regime of slow, large scale solutions with small aspect ratio. The SG equations are not uniformly valid over the whole of Earth's atmosphere, so they are not used to make weather forecasts. However, they can be extremely useful in providing reference solutions to examine more standard un-approximated equations, as was proposed in [12], and was carried out in [29, 30].

The "Geostrophic coordinate" reformulation [10], later understood in terms of Optimal Transport (OT) has been instrumental in the study of these equations and the proof of existence since [2]. For more details, motivations and comprehensive review of the theory, see [11]. The numerical solution of the OT-SG formulation requires to solve OT problems at all time to recover the velocity field. Numerical solutions were obtained in [11] under a Lagrangian discretization but available

OT solvers at that time were at best of cubic complexity and the resolution was coarse/costly.

Later advances in *Semi-Discrete Optimal* transport [22], leading to linear cost OT solvers have generated a renewed interest in solving numerically the Semigeostrophic equations [5] and higher resolution solutions have been obtained in [14]. Another efficient OT numerical resolution approach based on an *Entropic penalization* [13] [20] is developing rapidly, see [25] and also [23]. Our main contribution is the introduction of a new solver based on Entropic OT. The implementation uses a GPU back end and provides a fine resolution of the Eady slice test problem.

Outline: Section (2) recalls the optimal transport formulation of the Semigeostrophic equations, starting with the Semigeostrophic approximation of the three dimensional incompressible Boussinesq equations, before also considering the vertical slice incompressible Eady equations (which are a simplified two dimensional model exhibiting frontogenesis). Section 3 provides a minimal presentation of the Entropic regularization of OT and formulates the "Entropic" SG-OT equations, obtained by replacing standard OT by Entropic OT. Section 4 presents the discretization of the problem and the *Sinkhorn algorithm* used to solve the Entropic OT problems. We finally provide a numerical study in section 5 and a list of numerical and mathematical open problems and tasks in the conclusion.

2. The OT formulation of the SG equation

2.1. The Semigeostrophic approximation for the incompressible Boussinesq equations

The incompressible Boussinesq equations in Cartesian coordinates on the f-plane (constant Coriolis force f approximation) take the form

$$\frac{Du_1}{Dt} - fu_2 = -\frac{\partial p}{\partial x_1}, \tag{2.1}$$

$$\frac{Du_2}{Dt} + fu_1 = -\frac{\partial p}{\partial x_2}, \tag{2.2}$$

$$\frac{Du_3}{Dt} = b - \frac{\partial p}{\partial x_3}, \tag{2.3}$$

$$\frac{Db}{Dt} = 0, \tag{2.4}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0, \tag{2.5}$$

where (u_1, u_2, u_3) is the velocity field, p is the pressure field, f is the (here, constant) Coriolis parameter, and b is the buoyancy field (due to small density differences caused by changes in temperature), and where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u_1\frac{\partial}{\partial x_1} + u_2\frac{\partial}{\partial x_2} + u_3\frac{\partial}{\partial x_3}. \tag{2.6}$$

In this paper we consider the solution in a domain Ω with boundary $\partial\Omega$ with either periodic or rigid (or a combination of the two) boundary conditions $(u, v, w) \cdot n = 0$, where n is the unit outward pointing normal to the boundary.

We make the hydrostatic approximation by neglecting Du_3/Dt in 2.3, which is valid when the flow is slow and the aspect ratio (height of Ω divided by the width) is small. The geostrophic approximation, which is valid on the large scale when the flow is slow, also neglects Du_1/Dt and Du_1/Dt in (2.1-2.2), leading to

$$-fu_{2,g} = -\frac{\partial p}{\partial x_1}, fu_{1,g} = -\frac{\partial p}{\partial x_2}. \tag{2.7}$$

However, this is a purely diagnostic equation that does not predict dynamics. In the Semigeostrophic approximation of (2.1-2.5) we replace Du_1/Dt, Du_2/Dt by $Du_{1,g}/Dt$, $Du_{2,g}/Dt$ (whilst retaining u_1, u_2, u_3 in 2.6), leading to

$$\frac{Du_{1,g}}{Dt} - fu_2 = -\frac{\partial p}{\partial x_1}, \tag{2.8}$$

$$\frac{Du_{2,g}}{Dt} + fu_1 = -\frac{\partial p}{\partial x_2}, \tag{2.9}$$

$$0 = b - \frac{\partial p}{\partial x_3}, \tag{2.10}$$

$$\frac{Db}{Dt} = 0, \tag{2.11}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0. \tag{2.12}$$

To clarify how this equation might be solved, we use 2.7 and 2.10 to eliminate $u_{1,g}$, $u_{2,g}$ and b to obtain

$$-\frac{D}{Dt}\left(\frac{\partial p}{\partial x_2} - fy\right) = -\frac{\partial p}{\partial x_1}, \tag{2.13}$$

$$\frac{D}{Dt}\left(\frac{\partial p}{\partial x_2} + fx\right) = -\frac{\partial p}{\partial x_2}, \tag{2.14}$$

$$\frac{D}{Dt}\frac{\partial p}{\partial x_3} = 0, \tag{2.15}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0. \tag{2.16}$$

In these equations, (u_1, u_2, u_3) must be instantaneously solved for such that (2.13-2.15) are consistent with the time evolution of the gradient ∇p, subject to the incompressibility constraint. The optimal transport formulation aims to avoid the computational and analytical difficulties with this challenging nonlinear system.

2.2. Optimal transport formulation

The optimal transport formulation starts from the change of *physical variables* $X = (x_1, x_2, x_3) \in \Omega$ to *Geostrophic variables* $G = (G_1, G_2, G_3)$, defined by

$$G_1 = x_1 + u_{2,g}/f,\ G_2 = x_2 - u_{1,g}/f,\ G_3 = b/f^2. \tag{2.17}$$

The Geostrophic domain is defined as the image of the physical domain under this map. This means that the Geostrophic domain boundary can move as time progresses.

We notice that

$$G = (G_1, G_2, G_3) = \nabla P, \quad P = \frac{1}{2}(x_1^2 + x_2^2) + \frac{p}{f^2}. \tag{2.18}$$

The Cullen Stability Principle introduced in [10] notices that P must remain a convex function in order for the Energy of the system to remain in a minimal state and (2.13-2.15) to be solvable. It implies in particular that the map $X \mapsto G(X) = \nabla P(X)$ is well defined. The optimal transport formulation below gives a mathematical framework to this remark.

Returning to (2.8-2.12), and eliminating p for $u_{1,g}$, $u_{2,g}$ and b, we get

$$\frac{DG_1}{Dt} = u_{1,g} := f(G_2 - x_2), \tag{2.19}$$

$$\frac{DG_2}{Dt} = u_{2,g} := -f(G_1 - x_1), \tag{2.20}$$

$$\frac{DG_3}{Dt} = 0, \tag{2.21}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0, \tag{2.22}$$

which is a closed system when combined with 2.18.

To make further progress we need to consider the Lagrangian description of the equations. The Lagrangian description of fluid dynamics is formulated in terms of a time dependent flow map $\mathcal{X}_t : \Omega \to \Omega$ such that $\mathcal{X}_t(a)$ describes the time evolution of a moving fluid particle for each fixed $a \in \Omega$. Since particles move with velocity $U = (u_1, u_2, u_3)$, we have

$$\frac{\partial}{\partial t}\mathcal{X}_t = U_t \circ \mathcal{X}_t. \tag{2.23}$$

For simplicity here we define $\mathcal{X}_0 = I$ but it could be equally defined as any diffeomorphism on Ω. The time dependent density μ_t of particles is the image (or push forward) measure of the initial distribution μ_0 by $\mathcal{X}_t$, denoted $\mathcal{X}_t \# \mu_0$ and defined as

$$\int \phi \, d\mu_t := \int \phi \, d(\mathcal{X}_t \# \mu_0) = \int \phi \circ \mathcal{X}_t \, d\mu_0,\ \forall \phi \in \mathcal{C}(\Omega) \tag{2.24}$$

The divergence free incompressibility constraint 2.12 directly ensures that $\frac{\partial}{\partial t}\mu_t = 0$, particles move but their density, usually chosen as the uniform Lebesgue measure $\mu_0 = \mathcal{L} \llcorner \Omega$ is preserved by the flow.

The image of the particle trajectories in the Geostrophic space in Lagrangian coordinates is:

$$\mathcal{G}_t = \nabla P \circ \mathcal{X}_t \tag{2.25}$$

Using (2.23) and (2.19-2.21) gives

$$\frac{\partial}{\partial t}\mathcal{G}_t = V \circ \mathcal{G}_t, \tag{2.26}$$

where V is the velocity in Geostrophic coordinates given by

$$V = \underbrace{f \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{=J} \cdot \Big((\nabla P)^{-1} - G \Big). \tag{2.27}$$

Equation (2.25) couples the Lagrangian systems in physical and Geostrophic coordinates. The distribution of particles in Geostrophic space is defined as $\sigma_t = \nabla P \# \mu_t$, the push-forward or image measure of μ_t. As already mentioned the incompressibility constraint implies that the density $\mu_t = \mu_0$ is conserved and does not depend on t. The convexity assumption on P allows to use Brenier's theorem [6] (see [26] for an extensive presentation) to reformulate the SG equations in Geostrophic coordinates i.e. in terms of σ_t or equivalently $\mathcal{G}_t$, for all t.

Proposition 1 (Brenier Th.)**.** *With the particular choice of μ_0 and assuming σ_t has a probability density with finite second moment, there is a unique map onto the support of σ_t , $X \mapsto G^*(X) = \nabla P(X)$ and P convex, solution of the Monge optimal transport problem*

$$OT(\mu_0, \sigma_t) := \inf_{G,\, \sigma_t = G\#\mu_0} \frac{1}{2} \int_\Omega \|G(X) - X\|^2 \, \mu_0(\mathrm{d}\, X). \tag{2.28}$$

(see (2.24) for the definition of the push forward constraint).

Moreover, one can consider the reverse transport problem and $G \mapsto X^*(G) = (\nabla P)^{-1}(G) = \nabla Q(G)$ (Q beeing the Legendre-Fenchel transform of P) is the solution

$$OT(\sigma_t, \mu_0) := \inf_{X,\, X\#\sigma_t = \mu_0} \frac{1}{2} \int_{\mathbb{R}^d} \|G - X(G)\|^2 \, \sigma_t(\mathrm{d}\, G). \tag{2.29}$$

Equations (2.26-2.27) can therefore be written in in Lagrangian Geostrophic coordinates alone: t*f*he variational problem (2.29) yields $X^* = (\nabla P)^{-1}$ at all times from σ_t the density of $\mathcal{G}_t$. In Eulerian Geostrophic coordinates, one can further eliminate $\mathcal{G}_t$ and replace (2.26) by the continuity equation (in Geostrophic G coordinates),

$$\frac{\partial}{\partial t}\sigma_t + \nabla \cdot (\sigma_t \, V) = 0. \tag{2.30}$$

Remark 1 (Existence of solutions in Geostrophic coordinates)**.** *Existence and continuity of the Monge map depends on the convexity of the support of the target measure [7], i.e. where the density is bounded away from* 0*. For the Semigeostrophic equations, the density σ_t must be supported on the (evolving) Geostrophic domain.*
Working in Geostrophic coordinates, as in (2.30 - 2.27 - 2.29) is therefore easier as one controls μ_0. Existence of solutions in this setting was established in [2].

Remark 2 (Semigeostrophic energy)**.** *Using (2.17), (2.29) is formally equivalent to extremising the Semigeostrophic energy,*

$$E_{SG} = \int \frac{1}{2}\left(u_{1,g}^2 + u_{2,g}^2\right) - b\, x_3 \,\mathrm{d}\, X, \tag{2.31}$$

over P, which is how the optimal formulation was originally derived in [10].

Remark 3 (Going back to physical coordinates and regularity issues)**.** *To compute physical quantities $u_{1,g}$, $u_{2,g}$ and b we need to solve the forward optimal transport problem 2.28, and then obtain*

$$(u_{2,g}, -u_{1,g}, b/f) = G(X) - X. \tag{2.32}$$

Computing the full velocity u (instead of u_g) requires to rewrite (2.8-2.12) in the form

$$\frac{D}{Dt}(G_1 - x_1) = f(u_{1,g} - u_1), \quad \frac{D}{Dt}(G_2 - x_2) = f(u_{2,g} - u_2). \tag{2.33}$$

The left hand side can be computed in Geostrophic coordinates G and mapped back to physical coordinates X using the optimal map. Finally, u_3 can be recovered by integrating the divergence free constraint.

The regularity of σ_t and its support is however a serious obstacle to give a meaning to these solutions (in opposition to remark 1). See [19] for a review.

2.3. Optimal transport formulation of the Semigeostrophic Eady vertical slice model

The Eady vertical slice model considers a solution of (2.1-2.5) in a domain with infinite extent in the y direction. Here, the domain is periodic in the x direction with periodicity $2L$, and the vertical extent of the domain is $0 \leq z \leq H$. We write $b = -s\, x_2 + b'$, where $s > 0$. This describes a background horizontal temperature gradient that models the North-South temperature gradient on planet Earth due to differential solar heating. We then write $p = -s\, x_2\, (x_3 - H/2) + \phi$ so that

$$b - \frac{\partial p}{\partial x_3} = b' - \frac{\partial \phi}{\partial x_3}. \tag{2.34}$$

Finally, we assume that u_1, u_2, u_3, b' and ϕ are all independent of x_2. This produces the following consistent set of equations,

$$\frac{Du_1}{Dt} - fu_2 = -\frac{\partial \phi}{\partial x_1}, \tag{2.35}$$

$$\frac{Du_2}{Dt} + fu_1 = s\,(x_3 - H/2) - \frac{\partial \phi}{\partial x_2}, \tag{2.36}$$

$$\frac{Du_3}{Dt} = b' - \frac{\partial \phi}{\partial x_3}, \tag{2.37}$$

$$\frac{Db'}{Dt} + su_2 = 0, \tag{2.38}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_3}{\partial x_3} = 0, \tag{2.39}$$

where now

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u_1 \frac{\partial}{\partial x_1} + u_3 \frac{\partial}{\partial x_3}. \tag{2.40}$$

This system can be solved entirely in two dimensions (x_1, x_3); this means that it is a useful system for testing out numerical schemes in a less computationally intensive framework. It is still very challenging because it exhibits fronts that form in finite time.

The Semigeostrophic version of this model is intended to describe frontogenesis in the x_1 direction, so the scales are different in the x_1 and x_2 directions. We assume that we can totally neglect Du_1/Dt and Du_3/Dt, whilst approximating $Du_2/Dt \approx Du_{2,g}/Dt$. This produces the Semigeostrophic vertical slice Eady model, where

$$fu_2 = fu_{2,g} := -\frac{\partial \phi}{\partial x_1}, \tag{2.41}$$

$$\frac{Du_{2,g}}{Dt} + fu_1 = s\,(x_3 - H/2), \tag{2.42}$$

$$0 = b' - \frac{\partial \phi}{\partial x_3}, \tag{2.43}$$

$$\frac{Db'}{Dt} + su_{2,g} = 0, \tag{2.44}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_3}{\partial x_3} = 0, \tag{2.45}$$

$$\tag{2.46}$$

whilst preserving u_1 and u_3 in D/Dt.

To obtain the optimal transport formulation, we make the change of variables

$$(G_1, G_3) = (x_1 + u_{2,g}/f, b/f^2) = \nabla P, \quad P = \frac{x_1^2}{2} + \frac{\phi^2}{2}. \tag{2.47}$$

Following very similar steps to the previous subsection, we obtain

$$\frac{DG_1}{Dt} = u_{1,g} := \frac{s}{f}(x_2 - H/2), \tag{2.48}$$

$$\frac{DG_3}{Dt} = u_{2,g} := -\frac{s}{f}(x_1 - X_1), \tag{2.49}$$

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0. \tag{2.50}$$

Again, introducing the pushforward measure $\sigma = \nabla P \# \mu_0$, we obtain the following optimal transport formulation.

Definition 1 (Optimal transport formulation of the Semigeostrophic Eady vertical slice equations)**.** *The optimal transport formulation of the Semigeostrophic Eady vertical slice equations is given by the initial value problem for σ such that*

$$\sigma_t + \nabla_G \cdot (V\sigma) = 0, \tag{2.51}$$

where

$$V(G) = \frac{s}{f} J \begin{pmatrix} X_1^*(G) - G_1 \\ X_3^*(G) - H/2 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.52}$$

and

$$X^* = \arg \inf_{X,\, X\#\sigma = \mathcal{L}} \frac{1}{2} \int_\Omega \|X(G) - G\|^2 \,\mathrm{d}\,\sigma(G). \tag{2.53}$$

As in remark 2, this is in fact equivalent to extremizing the Semigeostrophic energy

$$E_{SG} = \int \frac{1}{2} u_{1,g}^2 - b'\, x_3 \,\mathrm{d}\,\mu_0(X), \tag{2.54}$$

3. Kantorovich formulation and its entropic regularization

3.1. Kantorovich formulation

Kantorovich relaxation relies on optimizing on the larger space of all possible symmetric couplings $(X, G) \mapsto \pi(X, G)$ instead of maps $X \mapsto G(X)$ or $G \mapsto G(X)$ in proposition 1. Here π is a probability measure on the product *physical* $\times$ *Geostrophic* space, $\pi(X, G)$ is the portion of mass transported between $\mu_0(X)$ and $\sigma(G)$ assuming these are densities and its marginals are constrained :

$$\int_{\mathbb{R}^d} \pi(X, \mathrm{d}\, G) = \mu_0(X), \quad \int_\Omega \pi(\mathrm{d}\, X, G) = \sigma_t(G) \tag{3.1}$$

Then one has

Proposition 2. *The Kantorovich problem*

$$OT(\mu_0, \sigma_t) := \inf_{\pi \geq 0,\ s.t.\ (3.1)} \frac{1}{2} \int_{\Omega \times \mathbb{R}^d} \|G - X\|^2 \, \pi(\mathrm{d}\, X, \mathrm{d}\, G). \tag{3.2}$$

has a solution and defines the same (proposition 1) Optimal Transport distance. We simplified the problem by minimizing on positive measures as the normalisations (3.1) enforce the probability measure property. Moreover one has:

(i) The relaxation is tight when the Monge Problem has a solution, the unique optimal plan concentrates on the graph of the maps

$$\pi^*(X, G) = \mu_0(X)\, \delta_X \times \delta_{G = G^*(X)} = \sigma_t(G)\, \delta_G \times \delta_{X = X^*(G)} \tag{3.3}$$

(δ is here the dirac measure).

(ii) Kantorovich has a simpler dual formulation on (Φ, Ψ), continuous functions with compact support: $OT(\mu_0, \sigma_t) :=$

$$\sup_{\substack{\Phi \in \mathcal{C}(\Omega), \Psi \in \mathcal{C}(\mathbb{R}^d)\ s.t. \\ \Phi(X) + \Psi(G) \leq \|G - X\|^2,\ \forall (X, G)}} \int_\Omega \Phi\, \mu_0(\mathrm{d}\, X) + \int_{\times \mathbb{R}^d} \Psi\, \sigma_t(\mathrm{d}\, G) \tag{3.4}$$

(iii) the primal π^* and dual (Φ^*, Ψ^*) (a.k.a. Kantorovich potentials) optimal variables provide the following barycentric map *characterization of the optimal* displacement

$$\nabla \Psi^*(G) = \int_\Omega X \, \frac{\pi^*(\mathrm{d}\, X, G)}{\mu_0(X)} - G = X^*(G) - G,\ \forall G, \tag{3.5}$$

and the symmetric characterization

$$\nabla \Phi^*(X) = \int_{\mathbb{R}^d} G \, \frac{\pi^*(X, \mathrm{d}\, G)}{\sigma_t(G)} - X = G^*(X) - X,\ \forall X. \tag{3.6}$$

Remark 4 (Periodic case)**.** *In the periodic case (as in the Eady slice problem), we need to give a meaning to the integrals in (3.5,3.6), because the sum of points has no meaning except in a cartesian setting, where the points are identified with vectors. To this end, we write (3.5) under the form $\nabla \Psi^*(G) = \int_\Omega (X - G) \, \frac{\pi^*(\mathrm{d}\, X, G)}{\mu_0(X)}$, then it extends to the periodic setting, with $X - G$ being the smaller vector joining G to X, and the integral running on vectors (see section 1.3.2 in [26]).*

Remark 5 (Wasserstein Hamiltonian Systems)**.** *A differential structure on the space of probability measures known as "Otto calculus" is leveraged in [1] to eliminate V (2.27) from the SG equations 2.30. In a nutshell: defining the "Hamiltonian" $H(\sigma) = OT(\sigma, \mu_0)$, its first variation with respect to σ is the dual continuous function (see (3.4))*

$$\delta_\sigma H = \Psi^*. \tag{3.7}$$

Using (3.5), (2.30) can written as a PDE (in the Wasserstein metric) in σ_t

$$\frac{\partial}{\partial t}\sigma_t + \nabla \cdot (\sigma_t\, J \cdot \nabla(\delta_\sigma H(\sigma_t)) = 0. \tag{3.8}$$

It is easy to check (at least formally) that $\frac{\partial}{\partial t}H(\sigma_t) = 0$, hence the Hamiltonian system denomination. This technique of course applies to other Hamiltonians and in particular to the Entropic regularization of Optimal Transport below. To get a more detailed introduction to Wasserstein gradient and hamiltonian flows, see chapter 8 of [28] and especially part 8.3.2 for other examples of Wasserstein hamiltonian dynamics.

3.2. Entropic regularization

Kantorovich relaxation is a landmark theoretical result (see [26] for comprehensive presentation) providing a simpler linear programming (LP) formulation of the non-linear Monge problem. From the computational point of view, however it does not scale well with large discretization problems. The fundamental obstacle is the complexity of LP solvers and the size of the primal unknowns or the similar number of dual constraints.

The relative entropy (a.k.a. Kullback Leibler divergence) between measures q and r is defined as:

$$\mathrm{KL}(q \,|\, r) := \begin{cases} \int \log\left(\frac{dq}{dr}\right) \mathrm{d}q \text{ if } q \ll r \\ +\infty \text{ otherwise.} \end{cases} \tag{3.9}$$

The $q \ll r$ (q is absolutely continuous with respect to the reference measure r) means that q has density $\frac{dq}{dr}$, a function of q) thus allowing to take its logarithm. The sum spans the support of q and will be discrete if (as below) q is a discrete measure. The penalization of (3.2) using relative entropy can be written in the form $OT_\varepsilon(\mu_0, \sigma_t) :=$

$$\inf_{\pi_\varepsilon \text{ s.t. } (3.1)} \frac{1}{2} \int_{\Omega \times \mathbb{R}^d} \|G - X\|^2 \, \pi_\varepsilon(\mathrm{d}\, X, \mathrm{d}\, G) + \varepsilon \, \mathrm{KL}(\pi_\varepsilon \,|\, \mu_0 \times \sigma_t) \tag{3.10}$$

Note that as $\mathcal{K}L(q|r)$ is a strictly convex function of q with positive infinite slope at 0, this problem now has a unique solution and most importantly we got rid of the positivity constraint $\pi \geq 0$. There is some freedom in the choice of the reference measure, it is customary to choose the tensor product of the marginal measure as the "best" a priori approximation containing the support of optimal plans.

The minimization (3.10) has an equivalent dual maximization formulation:

$$OT_\varepsilon(\mu_0, \sigma_t) := \sup_{\Phi_\varepsilon \in \mathcal{C}(\Omega), \Psi_\varepsilon \in \mathcal{C}(\mathbb{R}^d)} \mathcal{J}_\varepsilon(\Phi_\varepsilon, \Psi_\varepsilon) \tag{3.11}$$

where

$$\begin{aligned}\mathcal{J}_\varepsilon(\Phi_\varepsilon, \Psi_\varepsilon) := \quad & \int_\Omega \Phi_\varepsilon \, \mu_0(\mathrm{d}\, X) + \int_{\times \mathbb{R}^d} \Psi_\varepsilon \, \sigma_t(\mathrm{d}\, G) - \\ & \varepsilon \int_{\Omega \times \mathbb{R}^d} (e^{\frac{1}{\varepsilon}(\Phi_\varepsilon(X) + \Psi_\varepsilon(G) - \frac{1}{2}\|G-X\|^2)} - 1) \, \mu_0(\mathrm{d}\, X) \times \sigma_t(\mathrm{d}\, G)\end{aligned} \tag{3.12}$$

The entropic regularization method has replaced the dual constraints by a soft barrier penalization. The new unconstrained problem is downsized to just the discretization of the dual variables.

We gather properties (similar to prop. 2) for the optimal primal and dual variables:

Proposition 3. *(i) The unique optimal plan has an explicit form as a function of the dual variable.*

$$\pi_\varepsilon^*(X, G) = e^{\frac{1}{\varepsilon}(\Phi_\varepsilon^*(X) + \Psi_\varepsilon^*(G) - \frac{1}{2}\|G-X\|^2)} \mu_0(X) \times \sigma_t(G) \tag{3.13}$$

(ii) The optimal dual variables $(\Phi_\varepsilon^*(X), \Psi_\varepsilon^*)$ are unique up to an additive constant C ($(\Phi_\varepsilon^*(X) + C, \Psi_\varepsilon^* - C)$ is also a solution) and satisfy the optimality system

$$\delta_{\Phi_\varepsilon} \mathcal{J}_\varepsilon(\Phi_\varepsilon^*, \Psi_\varepsilon^*) = 0 \quad \delta_{\Psi_\varepsilon} \mathcal{J}_\varepsilon(\Phi_\varepsilon^*, \Psi_\varepsilon^*) = 0 \tag{3.14}$$

the 0 elements above are in the space of functions, (3.14) gives an explicit representation of each variable as a function of the other:

$$\Phi_\varepsilon^*(X) = LSE_\varepsilon^{\sigma_t, \Psi_\varepsilon^*}(X), \, \forall X \quad \Psi_\varepsilon^*(G) = LSE_\varepsilon^{\mu_0, \Phi_\varepsilon^*}(G), \, \forall G \tag{3.15}$$

where we introduced the log/sum/exp operator:

$$LSE_\varepsilon^{\rho, u}(X) := -\varepsilon \log(\int e^{\frac{1}{\varepsilon}(u(X) - \frac{1}{2}\|G-X\|^2)} \rho(\mathrm{d}\, G)) \tag{3.16}$$

(X and G play symmetric roles according on the measure arguments in (3.15).

(iii) The barycentric map *characterizations (prop. 2 (iii)) still hold*

$$\nabla \Psi_\varepsilon^*(G) = \int_\Omega X \, \frac{\pi_\varepsilon^*(\mathrm{d}\, X, G)}{\mu_0(X)} - G, \, \forall G, \tag{3.17}$$

$$\nabla \Phi_\varepsilon^*(X) = \int_{\mathbb{R}^d} G \, \frac{\pi_\varepsilon^*(X, \mathrm{d}\, G)}{\sigma_t(G)} - X, \, \forall X. \tag{3.18}$$

but the plan π_ε^* is diffuse and has no Monge Map displacement interpretation.

(iv) The Wasserstein first variation equation (3.7) also holds , choosing $H = OT_\varepsilon$ we get

$$\delta_\sigma H = \Psi_\varepsilon^*. \tag{3.19}$$

This paper is based on using the efficient Sinkhorn solver (see section 4) and replacing $\nabla \Psi^*$ by $\nabla \Psi_\varepsilon^*$ in (3.8).

(v) The convergence of problem (3.11) towards (3.2) as $\varepsilon \to 0$ as been largely studied both in the continuous and discrete case. We refer to [21] and [25] for comprehensive ref. and postpone more precise results we use below due to [4] to section 4.

Remark 6 (Assymptotic ε considerations)**.** *At the continuous level, first order asymptotic expansion as $\varepsilon \to 0$ of the entropic transport cost has been established in [9], [24] :*

$$OT_\varepsilon(\mu,\nu) = OT_0(\mu,\nu) - \varepsilon \ln(\Lambda_\varepsilon) - \frac{\varepsilon}{2}(\mathrm{KL}(\mu \mid \mu_0)) + \mathrm{KL}(\nu \mid \mu_0)) + o(\varepsilon) \tag{3.20}$$

here Λ_ε is a constant and only depends on ε.
The result only holds for smooth and compactly supported measures (μ,ν) with μ_0 the reference Lebesgue measure on both metric spaces. We can always choose $\mu = \mu_0$ but using (3.20) with $\nu = \sigma_t$ remains purely formal. It nevertheless provides an encouraging view of the entropic bias induced when replacing $\nabla\Psi^*$ by $\nabla\Psi^*_\varepsilon$. Indeed taking Wasserstein variation in the first order terms yields :

$$\delta_\nu(-\varepsilon \ln(\Lambda_\varepsilon) - \frac{\varepsilon}{2}(\mathrm{KL}(\mu \mid \mu_0)) + \mathrm{KL}(\nu \mid \mu_0))|_{(\mu,\nu)=(\mu_0,\sigma_t)} = -\ln(\sigma_t) \tag{3.21}$$

where we have overloaded the notation and used $\sigma_t = \frac{d\sigma_t}{d\mu_0}$ for its density with respect to the Lebesgue measure μ_0. Only the last term contributes and one readily checks that $\nabla \cdot (\sigma_t\, J \cdot \nabla \ln(\sigma_t)) = 0$. We can therefore expect that the impact on the dynamics in (3.8) of the Entropic regularization is weak for small ε.

4. Discretization and Entropic OT solver

4.1. Sinkhorn Algorithm

Sinkhorn algorithm is the iterative coordinate-wise ascent solution of (3.11). Iterate over k

$$\Phi_\varepsilon^{k+1} = LSE_\varepsilon^{\sigma_t,\Psi_\varepsilon^k}, \quad \Psi_\varepsilon^{k+1} = LSE_\varepsilon^{\mu_0,\Phi_\varepsilon^{k+1}}. \tag{4.1}$$

This solver has been thoroughly investigated (see the monograph [21] and [25]). It is known to converge linearly to the solution of (3.15) but the rate of convergence degrades with ε as a $\mathcal{O}(1-\varepsilon)$. When μ_0 and σ_t are discrete measures supported on N Diracs, the two functional equations in (4.1) becomes a $2N$ system of equations and the integral in the LSE'_ε operators sums. The naive cost of one iteration of of order $\mathcal{O}(N^2)$. When the discrete problem arises from the discretization of continuous measures for which a Monge (graph) solution exists, it is possible to construct a continuation multi-scale method increasing N and decreasing ε which requires a finite number of k iterations at each scale. This heuristic has been followed in [27] and also [16]. The resulting parallel implementation is numerical shown to provide a linear solver in term of the fine (N,ε) target.

In this paper we use the corresponding GPU optimized library Geomloss [15] as well as the standard Pytorch AutoGrad tool to approximate the displacements (3.5) (same for (3.6)) by

$$\nabla\Psi^* \underset{\substack{k\to\infty\\ \varepsilon\to 0}}{\simeq} \nabla\Psi_\varepsilon^k \tag{4.2}$$

the limits in k and ε are of course finite in practice. This is discussed in the next section.

4.2. Berman convergence

The only (to the best of our knowledge) theoretical convergence result in (N, ε, k) is given in [4]. Summarizing and simplifying corollary 1.3 [4]:

Theorem 4. *Set* $N = [\frac{1}{\varepsilon^d}]$. *Assume,* $(\mu_N, \nu_N) \to (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ *(tech. assumptions missing) and* $(\mu, \nu) \in \mathcal{C}^{2,\alpha}$ *and bounded below. Then*

(i) There exists positive constants A, C *such that :*

For $k \geq m_\varepsilon := [-A\frac{\log(\varepsilon)}{\varepsilon}]$ *, we have*

$$\|\Psi_\varepsilon^k - \Psi^*\|_\infty,\ \|\Phi_\varepsilon^k - \Phi^*\|_\infty \leq C\,\varepsilon\log(\varepsilon) \tag{4.3}$$

(ii) Setting

$$\pi_\varepsilon^k(X,G) = e^{\frac{1}{\varepsilon}(\Phi_\varepsilon^k(X)+\Psi_\varepsilon^k(G)-\frac{1}{2}\|G-X\|^2)}\,\mu_N(X)\times\nu_N(G).$$

There is a constant p *depending only on the marginals such that, again for* $k \geq m_\varepsilon$ *and* $\forall X$ *:*

$$\pi_\varepsilon^k(X,G) \leq \frac{p}{\varepsilon^p}\, e^{\frac{1}{2\,\varepsilon\,p}(\|G-X^*(G)\|^2)}\,\mu_N(X)\times\nu_N(G), \quad \forall G \tag{4.4}$$

and the symmetric result using the other optimal Monge map $(X \mapsto X^*(G))$ *(see proposition 1).*

These results depends on the existence and regularity of the Monge Maps and on fine time convergence of non-linear parabolic equations. We will not discuss this here but rater give a a few comments allowing to understand this result:

- The results gives a scaling between discretization N, entropic regularization ε and minimum number of Sinkhorn iterations m_ε to guarantee joint convergence.
- When running Sinkhorn, equations (4.1) holds on the finite discrete support. The functions $(\Phi_\varepsilon^k, \Psi_\varepsilon^k)$ in (4.3) are the continuous extensions of their discrete counterpart using the formula (3.16)

- Estimate (4.4) is obtained using (4.3) but also relies on the log concavity of the slices $G \mapsto \pi_\varepsilon^k(X, G)$. It gives the precious quantitative information that the support of the entropic plan π_ε^k converges to the graph of the Monge map exponentially fast, actually like a Gaussian of standard deviation $\sqrt{p\varepsilon}$. This behavior is the instrument to discard points in multi-scale continuation method mentioned in subsection 4.1. A theoretical analysis of this strategy is available in [3].

4.3. Lagrangian Semigeostrophic discretization

We attack the continuous Lagrangian SG equations (2.26-2.27) here rewritten using (3.5):

$$\frac{\partial}{\partial t}\mathcal{G}_t = f\,J \cdot (X^*(\mathcal{G}_t) - \mathcal{G}_t) = f\,J \cdot \nabla\Psi^*(\mathcal{G}_t) \tag{4.5}$$

We recall that Ψ^* is the optimal Kantorovich potential for the $OT(\mu_0, \sigma_t)$ problem. Let $\mu_{0,N}$ be a N point sampling of the Lebesgue measure on Ω, we approximate σ_t by a N equi-weighted particles $\{\mathcal{G}_{t,i}\}$ moving in time

$$\sigma_{t,N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\mathcal{G}_{t,i}}. \tag{4.6}$$

We will consider the following *space* discretization of (4.5):

$$\frac{\partial}{\partial t}\mathcal{G}_{t,i} = f\,J \cdot \nabla\Psi_\varepsilon^{k^*}(\mathcal{G}_{t,i}), \quad \forall i \tag{4.7}$$

where now the optimal transport part is approximated (see 4.2) using converged Sinkhorn potentials corresponding to the $OT_\varepsilon(\mu_{0,N}, \sigma_{t,N})$ problem.

At each time, Sinkhorn algorithm is stopped after k^* iterations. We discuss the choice of k^* in subsection 5.4. Finally, we note that proving convergence of the discrete scheme (4.7) towards (4.5) is open as 1) the required regularity assumptions on σ_t to use proposition 4) are not available, 2) Convergence estimates are only available on the potentials and not their gradients.

5. Numerical study

We use an explicit time discretization of (4.7) (mid point rule or 4*th* order Runge-Kutta) with constant time step dt. The use of implicit symplectic scheme are left for later investigations.

We tested the numerical method on the 2D Eady slice model (section 2.3) which shares the same mathematical structure (2.51-2.52) as the 3D problem used in the above sections. This is the test case used in the few existing numerical studies [10] [14]. The initial conditions are *exactly* the same as in [14] section 5.1 (table 1 column 1) and corresponds to an unstable perturbation of a stationary solution (stationary means $u_{2,g} = 0$ in (2.47)). The perturbation generates the non convex deformation observed on the left frame figure 1.

5.1. Frontogenesis dynamics

Figure 1 to 7, show the time evolution of the Geostrophic particles over a period of approximately 20 days. We used $N = 65536$, $\sqrt{\varepsilon} = 0.0625$ and 320 time steps. It ran for 8 hours on a single AMD EPIC 7302 GPU with 16 cores, 3-3.3GHz and 192 GBytes of RAM. The left column shows the uniform weighted samples $\sigma_{t,N}$ (4.6) in the Geostrophic space (dubbed *Geoverse* in the figures), $\mu_{0,N}$ is not shown and remains an unchanged sample of $\mu_0 = \mathcal{L} \llcorner \Omega$. The middle and right column position in the physical space (dubbed the *Universe*) the image of the $\mathcal{G}_{t,i}$ particles by the approximate entropic barycentric map (see (3.17)):

$$\mathcal{X}_{t,i} \simeq \mathcal{G}_{t,i} + \nabla \Psi_\varepsilon^{k^*}(\mathcal{G}_{t,i}). \tag{5.1}$$

We use as a color code the temperature, i.e the second coordinate of $\mathcal{G}_{t,i}$, as a color code in the middle column and the geostrophic speed $u_{2,g}$, i.e. the first coordinate of $\mathcal{G}_{t,i} - \mathcal{X}_{t,i}$ (see (2.47)) in the right column. Remember that in the the Eady Slice model the temperature of the particles is not preserved in time.

The results show a vertical shear developing associated to moving temperature and geostrophic speed discontinuities. This is precisely the expected behavior of the model: generating "fronts" from smooth perturbation. Note that the "map" $\nabla \Psi_\varepsilon^{k^*}$ remains continuous unlike its inverse $\nabla \Phi_\varepsilon^{k^*}$. The "front" formation is generated by the discontinuity of the transport map, well understood when the support of the target measure, here σ_t is non convex (see [18] for a discussion and a review of transport maps between non convex domains).

Our results are consistent with [14] but our resolution catches finer detail, in particular on figure 2 the apparition of 2 fronts from top and bottom boundaries. They will move in and merge. The choice of using a fixed weight sampling generates the line artifacts in the Geoverse and seem to contribute the propagation of errors in the Universe.

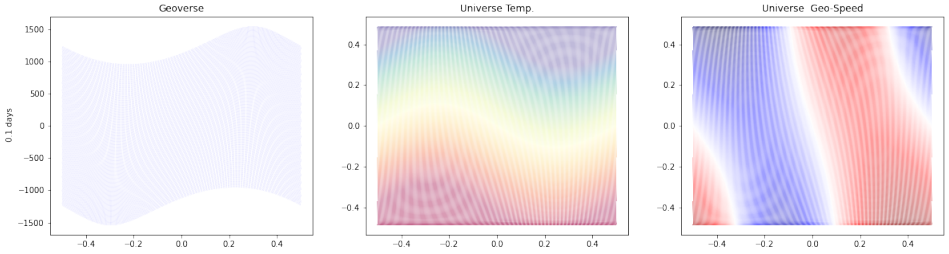

Figure 1: 0.1 Days.

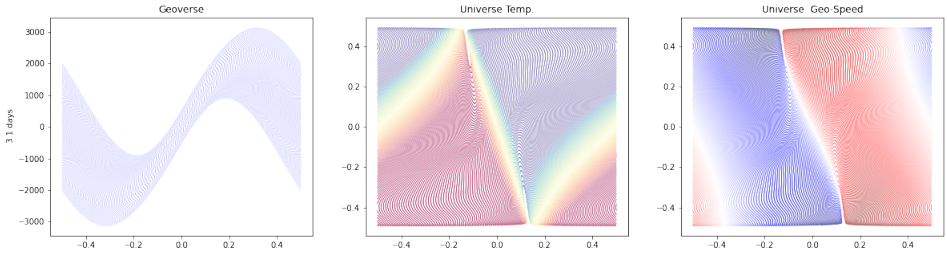

Figure 2: 3.1 Days.

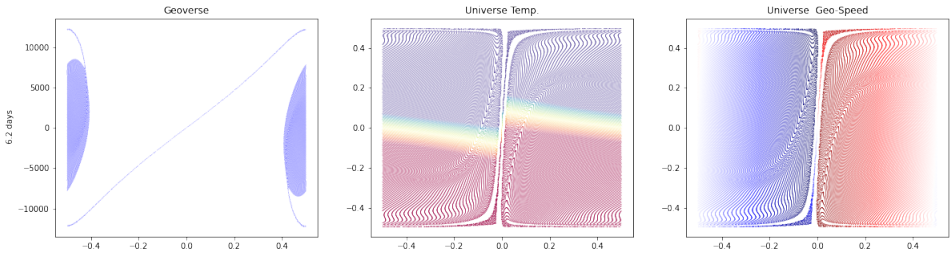

Figure 3: 6.2 Days.

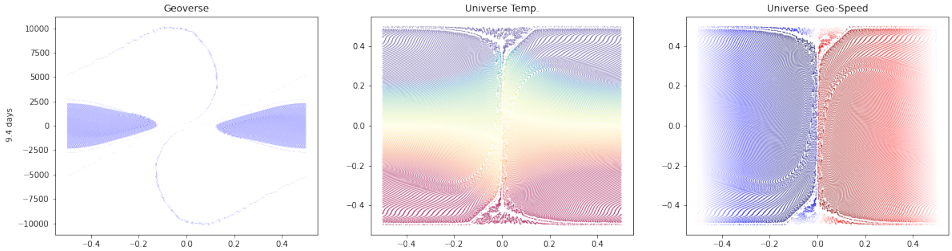

Figure 4: 9.4 Days.

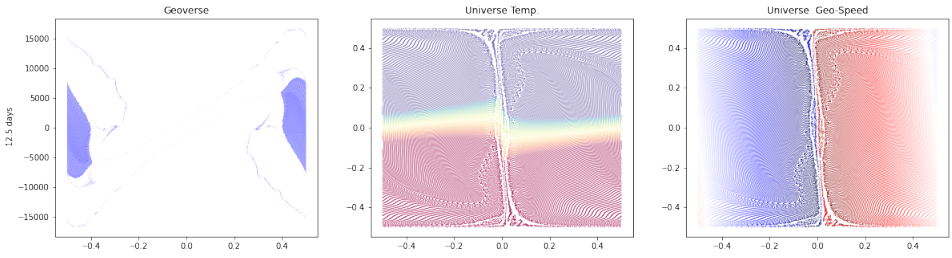

Figure 5: 12.5 Days.

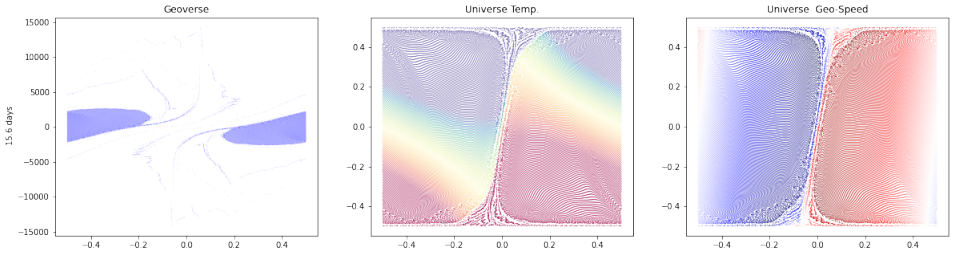

Figure 6: 15.6 Days.

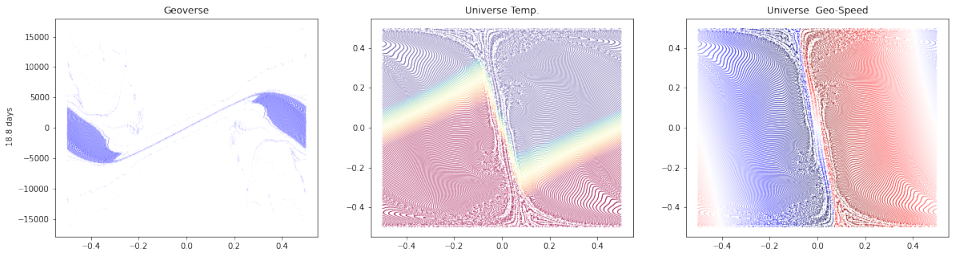

Figure 7: 18.8 Days.

5.2. Convergence

We first study the convergence of the computed solutions depending on ε and $N_\varepsilon = [\mathcal{O}(\frac{1}{\varepsilon^2})]$ linked by the Berman rule (Theorem 4). We compare, at time $t_2 = 2\,days$, for different values of ε, $\sigma_{t_0,N_\varepsilon}$ with the finest solutions (larger N_ε) obtained for $\varepsilon_{min} = 10^{-4}$.

The comparison uses Wasserstein 1 and 2 distances $\mathcal{W}_{1,2}$ (actually their Sinkhorn Divergence approximation which are also implemented in Geomloss [15] see [17] for details). **To sum up figure 8 left** shows the "pseudo-residual error" curves

$$\sqrt{\varepsilon} \mapsto \mathcal{W}_{1,2}(\sigma_{t_2,N_{\varepsilon_{min}}}, \sigma_{t_2,N_\varepsilon}) \tag{5.2}$$

in $log - log$ scale as well as a slope 2 line. This is an indication of first, respectively second, order convergence in $\mathcal{W}_1$, respectively $\mathcal{W}_2$ distance.

We also investigate the dependance on the time step dt for two time discretization: mid-point rule and Runge-Kutta 4. We compare, at time $t_2 = 2\,days$, different values of dt, $\sigma_{t_2,dt}$ with the finest solution obtained with a discretization $dt_{min} = 0.037\,days$. We keep a fixed $\varepsilon = 10^{-4}$ and corresponding N_ε. **Figure 8 right** shows the "pseudo-residual error" curves

$$dt \mapsto \mathcal{W}_1(\sigma_{t_2,dt_{min}}, \sigma_{t_2,dt}) \tag{5.3}$$

in $log - log$ scale as well as a slope 1 line. This is an indication of first order convergence in $\mathcal{W}_1$ for both(second and fourth order) time integrators. The speed $\nabla \Psi_\varepsilon^*$ is probably no better that $\mathbb{C}^1$ in time.

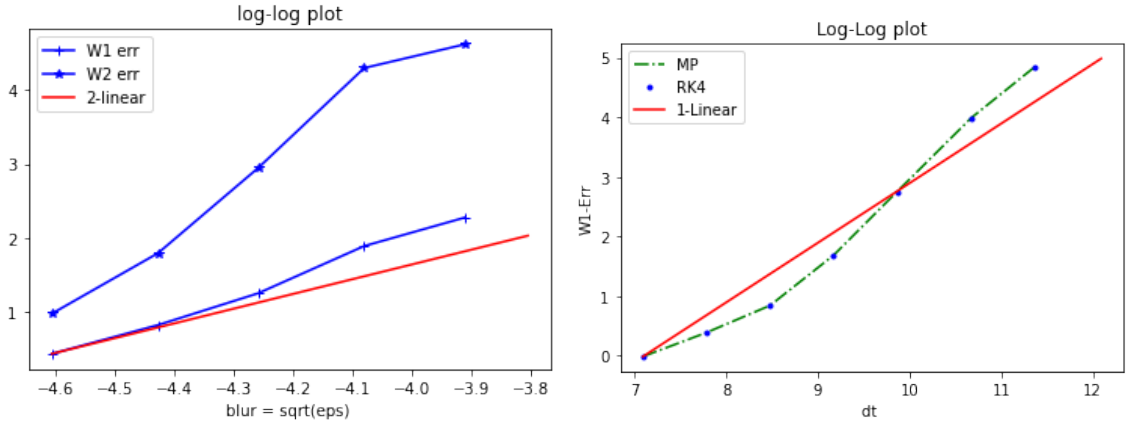

Figure 8: Left: ε convergence study, see (5.2). Right: dt convergence study, see (5.3).

5.3. *Energy conservation*

As solutions of the Hamiltonian system (3.8), the SG solutions conserve in time t the Hamiltonian/Energy, in our case $OT_\varepsilon(\sigma_t, \mu_0)$ solution of the primal (3.10) and dual (3.11) problem. Convex duality ensures in particular (see [8] section 6.4) that for admissible plans π_ε the dual cost is less that the primal:

$$\begin{gathered} \int_\Omega \Phi_\varepsilon \, \mu_0(\mathrm{d}\, X) + \int_{\times \mathbb{R}^d} \Psi_\varepsilon \, \sigma_t(\mathrm{d}\, G) - \\ \varepsilon \int_{\Omega \times \mathbb{R}^d} (e^{\frac{1}{\varepsilon}(\Phi_\varepsilon(X) + \Psi_\varepsilon(G) - \frac{1}{2}\|G-X\|^2)} - 1) \, \mu_0(\mathrm{d}\, X) \times \sigma_t(\mathrm{d}\, G) \\ \leq \\ \tfrac{1}{2} \int_{\Omega \times \mathbb{R}^d} \|G - X\|^2 \, \pi_\varepsilon(\mathrm{d}\, X, \mathrm{d}\, G) + \varepsilon \, \mathrm{KL}(\pi_\varepsilon \,|\, \mu_0 \times \sigma_t) \end{gathered} \tag{5.4}$$

and is equal at optimality :

$$\begin{gathered} \int_\Omega \Phi_\varepsilon^* \, \mu_0(\mathrm{d}\, X) + \int_{\times \mathbb{R}^d} \Psi_\varepsilon^* \, \sigma_t(\mathrm{d}\, G) \\ = \\ \tfrac{1}{2} \int_{\Omega \times \mathbb{R}^d} \|G - X\|^2 \, \pi_\varepsilon^*(\mathrm{d}\, X, \mathrm{d}\, G) + \varepsilon \, \mathrm{KL}(\pi_\varepsilon^* \,|\, \mu_0 \times \sigma_t) \end{gathered} \tag{5.5}$$

(the dual entropic part vanishes

$$\varepsilon \int_{\Omega \times \mathbb{R}^d} e^{\frac{1}{\varepsilon}(\Phi_\varepsilon^*(X) + \Psi_\varepsilon^*(G) - \frac{1}{2}\|G-X\|^2)} \, \mu_0(\mathrm{d}\, X) \times \sigma_t(\mathrm{d}\, G) = 0). \tag{5.6}$$

In the Semi-Geostrophic/OT dictionary the energy E_{SG} (2.54) is splited into the Kinetic Energy

$$E_K = \int \frac{1}{2} u_{1,g}^2 \, \mathrm{d}\, \mu_0(X) = \frac{1}{2} \int_{\Omega \times \mathbb{R}^d} \|G_1 - X_1\|^2 \, \pi_\varepsilon^*(\mathrm{d}\, X, \mathrm{d}\, G) \tag{5.7}$$

and the Potential Energy

$$E_P = -\int b' \, x_3 \, \mathrm{d}\, \mu_0(X) = -\frac{1}{2} \int_{\Omega \times \mathbb{R}^d} G_2 \, X_2 \, \pi_\varepsilon^*(\mathrm{d}\, X, \mathrm{d}\, G) \tag{5.8}$$

As we are using OT_ε and not OT_0 we have a third Energy component corresponding the the relative Entropy of the OT plan:

$$E_E = \varepsilon \, \mathrm{KL}(\pi_\varepsilon^* \,|\, \mu_0 \times \sigma_t) \tag{5.9}$$

Figure 9 left shows these Energie as well as their sum $E_K + E_P + E_E$ corresponding to the primal cost, the first line of (5.5). The second line of (5.5), the dual cost is natively returned by Geomloss [15] The simulation used the mid-point rule time integration, $dt = 0.0635\, days$, $\varepsilon = 10^{-4}$ and $N = 65536$. Energy conservation is conserved and the Kinetic Potential separation consistent with [14]). As expected the convergence error places the dual cost below the primal cost. We also note that the relative Entropy seems to be conserved, this is consistent with the asymptotic analysis developed in remark 6: the Entropy of σ_t is conserved and is a first order approximation of (5.9). **Figure 9 center** shows the same in *log* scale.

We now focus on the time evolution of another preserved quantity this time in the physical space: using (5.1) we can define

$$\mu_{t,N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{t,i}}, \tag{5.10}$$

an approximate sampling of the imcompressible fluid density $\mu_0 = \mathcal{L} \llcorner \Omega$. **Figure 9 right** shows (in *log* scale) the "imcompressibilty default" of the solver.

$$t \mapsto \mathcal{W}_1(\mu_{t,N}, \mu_{0,N}) \tag{5.11}$$

which is well preserved.

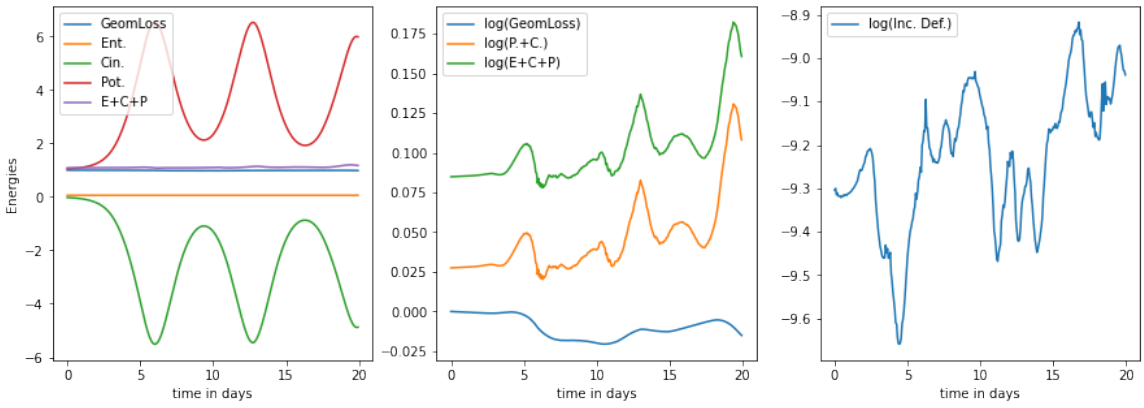

Figure 9: Left: the energies over time, Cin.=(5.7), Pot.=5.8, Ent. = 5.9, Geomloss is the dual cost in (5.5). Center: same in *log* scale. Right: incompressibility default (5.11).

5.4. Performance and choice of k^*

Berman's convergence (theorem 4) suggests to choose an ε dependent $k^* = m_\varepsilon$ but Geomloss implementation is turned towards performance and does not allow for this. It relies on and ε downscaling strategy to overcome the $\mathcal{O}(\frac{1}{\varepsilon})$ number of iteration curse. Given a $Scaling$ parameter $0 < Scaling < 1$, it runs a finite number of Sinkhorn iterations for a sequence of n scales $\varepsilon^n = Scaling^n$ decreasing to the prescribed finest ε. The algorithm then stops after $k^* = \mathcal{O}(\frac{\log \varepsilon}{\log Scaling})$ iterations. The number of points N_{ε^n} increases as in Berman's theorem 4. The potentials $(\Phi_\varepsilon, \Psi_\varepsilon)$ are interpolated from coarse n to fine level. This algorithm guarantees speed (we indeed observed a linear time cost w.r.t. to the number of samples for large problems) but the user needs to experiment with $Scaling$ to obtain convergence. Geomloss documentation suggest $Scaling = .99$ to ensure "good" convergence and we used this number in the results presented above. We also compared with $Scaling = .999$, this increases run time by an order of magnitude.

Figure 10 tests, on a two days simulation, the conservation of the Semi-Geostrophic energy (also discussed in subsection 5.3). We used mid-point rule time integration and a uniform $dt = 0.0635\, days$. We show the normalised energy for different values of ε and for $Scaling = .99$ or $= .999$ (Dashed or Solid lines). We gain an order of magnitude in accuracy. As a consequence of poorer Sinkhorn convergence ($Scaling = .99$) the auto-grad differentiation is not returning and accurate gradient of the Energy and conservation is lost.

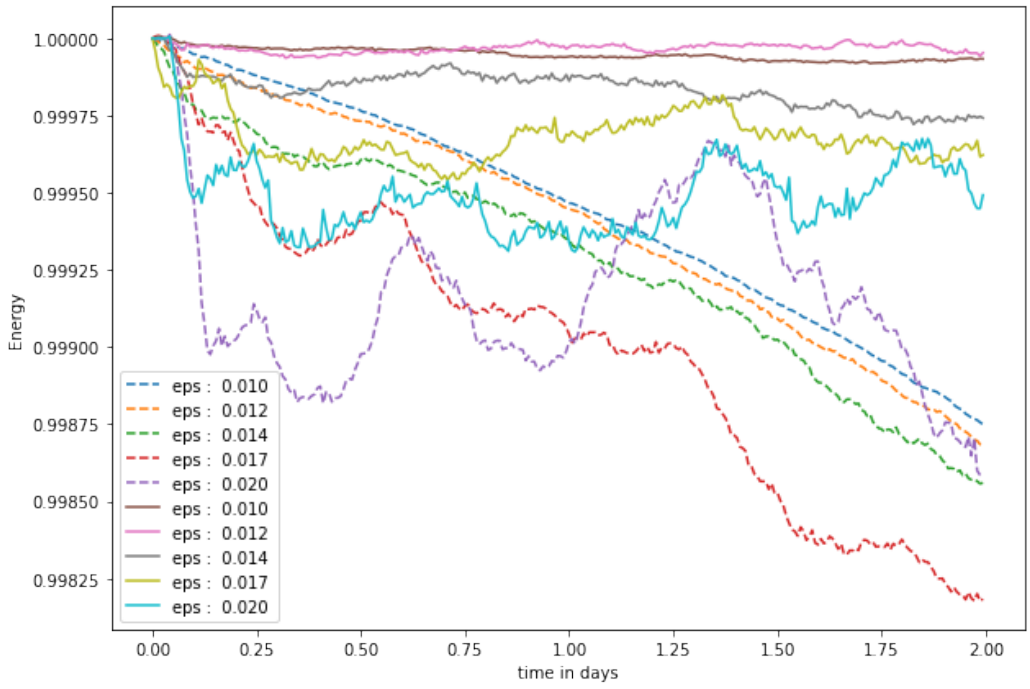

Figure 10: Energy conservation. $Scaling = .99$ or $= .999$ (Dashed or Solid lines).

6. Conclusion

These first experiments show that Entropic OT is an efficient and computationally relevant solution method. On the implementation side, a possible improvements would be to introduce a control on the convergence of Geomloss (this is possible using the lower level Keops library). Thanks to the the good scalability of the GPU implementation we also plan to design and solve a frontogenesis test case for the 3D SG equations.

We finally list open theoretical questions that are left for further investigations:

- For a fixed ε, this paper focus on the "Entropic" SG equations obtained by replacing $\nabla\Psi^*$ by its smoother Entropic version $\nabla\Psi^*_\varepsilon$ in (3.8). Existence (in the Brenier sense) obtained using a time discretization (as in [2]) seems straightforward. Uniqueness (still open for standard SG equations) would also probably be enforced by the Entropic penalization. The interpretation of the Entropic SG model in the physical domain is unclear.

- Letting $\varepsilon \to 0$ and using the current OT ε asymptotic results (remark 6) indicates that the Entropic penalization impact the dynamics only at second order. This is probably important to perform a joint dt and ε convergence study.

- The Lagrangian space discretization (4.7) has been used (in the standard $\varepsilon = 0$ case, see [5]) to prove existence of solutions (in the Aleksandrov sense). A more general result would be to used Berman convergence framework (theorem 4) to prove the joint convergence in space, time and ε of the solutions of (4.7).

The OT background used in the last two bullets currently rely on the assumption that the prescribed marginal remains smooth on both side (say $\mathcal{C}^{2,\alpha}$) with compact support. This is true for μ_0 but questionable for σ_t. Whether these conditions can be relaxed (one one side) or the use of the Entropic transport map confers sufficient regularity is not known.

Acknowledgements

We thank Jean Feydy for his help for using the Geomloss software. We also acknowledge the support of Imperial College through the Nelder Fellowship awarded to JD Benamou and the support of the Imperial-CNRS Collaboration Fund.

References

[1] Luigi Ambrosio and Wilfrid Gangbo. Hamiltonian odes in the wasserstein space of probability measures. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(1):18–53, 2008.

[2] J.-D. Benamou and Y. Brenier. Weak existence for the semigeostrophic equations formulated as a coupled Monge–ampère/transport problem. *SIAM J. Appl. Math.*, 58:1450–1461, October 1998.

[3] Jean-David Benamou and Mélanie Martinet. Capacity Constrained Entropic Optimal Transport, Sinkhorn Saturated Domain Out-Summation and Vanishing Temperature. working paper or preprint, May 2020.

[4] Robert J. Berman. The sinkhorn algorithm, parabolic optimal transport and geometric monge–ampère equations. *Numer. Math.*, 145(4):771–836, aug 2020.

[5] David P. Bourne, Charles P. Egan, Beatrice Pelloni, and Mark Wilkinson. Semi-discrete optimal transport methods for the semi-geostrophic equations. *Calculus of Variations and Partial Differential Equations*, 61(1), February 2022. Funding Information: DPB would like to thank the UK Engineering and Physical Sciences Research Council (EPSRC) for financial support via the grant EP/R013527/2 Designer Microstructure via Optimal Transport Theory. CPE is supported by The Maxwell Institute Graduate School in Analysis and its Applications, a Centre for Doctoral Training funded by the EPSRC (Grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University and the University of Edinburgh. BP and MW gratefully acknowledge the support of the EPSRC via the Grant EP/P011543/1 Analysis of models for large-scale geophysical flows. Publisher Copyright: © 2021, The Author(s).

[6] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[7] Luis Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 5:99–104, 01 1992.

[8] G. Carlier. *Classical and Modern Optimization*. Advanced textbooks in mathematics. World Scientific, 2021.

[9] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications, 2019.

[10] M. J. P. Cullen and R. J. Purser. An extended Lagrangian theory of semi-geostrophic frontogenesis. *J. Atmos. Sci.*, 41:1477–1497, 1984.

[11] Michael John Priestley Cullen. *A mathematical theory of large-scale atmosphere/ocean flow*. World Scientific, 2006.

[12] Mike Cullen. Modelling atmospheric flows. *Acta Numerica*, 16:67–154, 2007.

[13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc., 2013.

[14] Charlie P. Egan, David P. Bourne, Colin J. Cotter, Mike J. P. Cullen, Beatrice Pelloni, Steven M. Roper, and Mark Wilkinson. A new implementation of the geometric method for solving the eady slice equations, 2022.

[15] Jean Feydy. Geometric loss functions between sampled measures, images and volumes. 2019. Available at https://www.kernel-operations.io/geomloss/.

[16] Jean Feydy. *Analyse de données géométriques, au delà des convolutions*. PhD thesis, 2020. Thèse de doctorat dirigée par Trouvé, Alain Mathématiques appliquées université Paris-Saclay 2020.

[17] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 2019)*, volume 89 of *Proceedings of Machine Learning Research*, pages 2681–2690. PMLR, 16–18 Apr 2019.

[18] Alessio Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Communications in Partial Differential Equations*, 35(3):465–479, 2010.

[19] Alessio Figalli. *Global Existence for the Semigeostrophic Equations via Sobolev Estimates for Monge-Ampère*, pages 1–42. Springer International Publishing, Cham, 2018.

[20] Alfred Galichon and Bernard Salanié. Cupid's invisible hand: Social surplus and identification in matching models, 2021.

[21] Pedro Machado Manhães De Castro, Quentin Mérigot, and Boris Thibert. Far-field reflector problem and intersection of paraboloids. *Numerische Mathematik*, 134(2):389–411, October 2016.

[22] Quentin Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1584–1592, August 2011. 18 pages.

[23] Quentin Merigot and Boris Thibert. Optimal transport: discretization and algorithms, 2020.

[24] Soumik Pal. On the difference between entropic cost and the optimal transport cost, 2019.

[25] G. Peyré and M. Cuturi. Computational Optimal Transport. *ArXiv e-prints*, March 2018.

[26] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.

[27] Bernhard Schmitzer. Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems. *SIAM J. Sci. Comput.*, 41(3), 2019.

[28] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[29] AR Visram, CJ Cotter, and MJP Cullen. A framework for evaluating model error using asymptotic convergence in the eady model. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1629–1639, 2014.

[30] Hiroe Yamazaki, Jemma Shipton, Michael JP Cullen, Lawrence Mitchell, and Colin J Cotter. Vertical slice modelling of nonlinear eady waves using a compatible finite element method. *Journal of Computational Physics*, 343:130–149, 2017.