# Well-posedness and convergence of entropic approximation of semi-geostrophic equations

Guillaume Carlier,* Hugo Malamut†

April 29, 2024

**Abstract**

We prove existence and uniqueness of solutions for an entropic version of the semi-geostrophic equations. We also establish convergence to a weak solution of the semi-geostrophic equations as the entropic parameter vanishes. Convergence is also proved for discretizations that can be computed numerically in practice as shown recently in [6].

**Keywords:** optimal transport, entropic optimal transport, semi-geostrophic equations.

**MS Classification:** 49Q22, 35D30, 35Q35.

## 1 Introduction

The semi-geostrophic equations are used in meteorology to describe frontogenesis for large scale atmospheric flows. Initially proposed by Eliassen [18] in the late 1940s and subsequently developed by Hoskins [21] and Cullen, see [12], [13], [14], they have gained a lot of attention by mathematicians because of their connections with optimal transport theory. The seminal work of Brenier [8] indeed enabled Benamou and Brenier [5] to define a notion of weak solutions and establish the existence of such solutions. This was later generalized by Loeper [24] in the framework of measure-valued

*CEREMADE, UMR CNRS 7534, Université Paris Dauphine, PSL, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

†CEREMADE, UMR CNRS 7534, Université Paris Dauphine, PSL, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE and INRIA-Paris, MOKAPLAN, hugo.malamut@ens.psl.eu

solutions (also see Feldman and Tudorascu [19, 20]). It is also worth mentioning here that Ambrosio and Gangbo [3] developed a general and self-contained theory of Hamiltonian systems in the Wasserstein space which contains semi-geostrophic equations as a special case. Regarding stronger notions of solutions, thanks to the deep regularity theory for the Monge-Ampère equation, Ambrosio, Colombo, De Philippis and Figalli were able to obtain Eulerian solutions in the two-dimensional periodic case [1] and for convex three-dimensional domains [2].

Given $\Omega$ a Lipschitz bounded open subset of $\mathbb{R}^3$, and $\alpha_0$ a Borel measure on $\mathbb{R}^3$ with total mass $|\Omega|$ the semi-geostrophic system reads as the coupling of the continuity equation

$$\partial_t \alpha + \operatorname{div}(\alpha J(\operatorname{id} - \nabla \psi)) = 0, \ \alpha(0, \cdot) = \alpha_0, \ \text{with } J := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (1.1)$$

with the Monge-Ampère equation

$$\det(D^2 \psi_t) = \alpha_t, \ \psi_t \text{ convex.} \quad (1.2)$$

Before going further, let us briefly explain the origin of (1.1)-(1.2), we refer to [12], [13] [14] and the references therein for a modern and complete exposition and a derivation from incompressible Euler equations with a large Coriolis force. The starting point is the following system

$$\begin{cases} D_t \boldsymbol{v}^g + \nabla p = -J\boldsymbol{v} + \theta e_3, \ \operatorname{div}(\boldsymbol{v}) = 0, \ \text{in } (0, \infty) \times \Omega \\ \qquad\qquad \boldsymbol{v}^g = J\nabla p, \ D_t \theta = 0, \ \text{in } (0, \infty) \times \Omega, \\ \qquad \boldsymbol{v} \cdot \nu_\Omega = 0, \ \text{on } (0, \infty) \times \partial\Omega, \ p(0, \cdot) = p_0, \end{cases}$$

where $D_t = \partial_t + \boldsymbol{v} \cdot \nabla$ denotes the convective derivative, $e_3 = (0, 0, 1)^T$ is the vertical direction and $\nu_\Omega$ denotes the outward unit normal to $\partial\Omega$. Note that the first equation enables one to relate the pressure $p$ to the *buoyancy* $\theta$ through the simple relation $\partial_3 p = \theta$ and that the so-called semi-geostrophic balance equation $\boldsymbol{v}^g = J\nabla p$ imposes that the *semi-geostrophic velocity field* $\boldsymbol{v}^g$ is horizontal. Cullen's stability criterion [14] further requires that

$$P_t(x) = p_t(x) + \frac{1}{2}(x_1^2 + x_2^2)$$

is convex. Substituting $\boldsymbol{v}^g = J\nabla p$ in the first equation of the system yields

$$D_t \nabla P = J\nabla p = J(\nabla P - \operatorname{id}).$$

Let us then denote by $(t, x) \mapsto X_t(x)$ the flow of $\boldsymbol{v}$. This flow remains in the physical domain $\Omega$ (since $\boldsymbol{v}$ is tangential to $\partial\Omega$) and is measure preserving,

i.e. $X_{t\#}\mathscr{L}_\Omega = \mathscr{L}_\Omega$ where $\mathscr{L}_\Omega$ denotes the Lebesgue measure on $\Omega$ (since $\mathrm{div}(\boldsymbol{v}) = 0$). At least formally, we have

$$\partial_t \nabla P_t(X_t(x)) = J(\nabla P_t(X_t(x)) - X_t(x))$$
$$= J(\mathrm{id} - \nabla P_t^*)(\nabla P_t(X_t(x)))$$

where $P_t^*$ is the Legendre transform of $P_t$. So if we set $\alpha_t := \nabla P_{t\#}\mathscr{L}_\Omega$ and $\psi_t := P_t^*$ (so that $\nabla \psi_t = (\nabla P_t)^{-1}$ and $\mathscr{L}_\Omega = \nabla \psi_{t\#}\alpha_t$ if $\nabla P_t$ is invertible and smooth), the semi-geostrophic system reads as the coupling of the continuity equation (1.1) (with initial condition $\alpha_0 = \nabla P_{0\#}\mathscr{L}_\Omega$) with the Monge-Ampère equation (1.2). Note that (1.1)-(1.2) is a reformulation of the initial semi-geostrophic system in *semi-geostrophic* coordinates i.e. after the change of variables $x \mapsto \nabla P_t(x)$. Conversely, note that (again formally) from a solution of (1.1)-(1.2) one recovers $\nabla P = \nabla \psi^*$, $\theta$ by $\theta = \partial_3 p = \partial_3 P$, $\boldsymbol{v}^g = J\nabla p$, and the velocity $\boldsymbol{v}$ by

$$\boldsymbol{v}_t(x) = \partial_t \nabla \psi_t(\nabla P_t(x)) + (D^2\psi_t(\nabla P_t(x)))J(\nabla P_t(x) - x).$$

In general, $\psi_t$ and $P_t$ may fail to be smooth or strongly convex so that the previous computations only remain formal and one has to resort to some notion of weak solution, following the seminal optimal transport approach of Benamou-Brenier [5] (which will be recalled in section 2) based on the fact that $\nabla P_t$ is the optimal transport from $\mathscr{L}_\Omega$ to $\alpha_t$.

The present paper is motivated by recent research on optimal transport methods to numerically solve semi-geostrophic equations. On the one hand, in the spirit of the geometric approach pioneered by Cullen and Purser [14], Bourne, Egan, Pelloni and Wilkinson proposed a semi-discrete optimal transport strategy [7] that was implemented in [17]. On the other hand, Benamou, Cotter and Malamut [6] developed and implemented an alternative method based on entropic approximation of optimal transport and the popular Sinkhorn algorithm (see [15], [27]). In [7], the authors have established well-posedness of semi-discrete approximations (for suitably discretized initial conditions) as well as their convergence. Concerning entropic approximation of semi-geostrophic equations, well-posedness and convergence issues are not addressed in [6]. The purpose of our paper is precisely to study these two questions. The paper is organized as follows. In section 2, we recall fundamentals from optimal transport theory, the weak formulation of semi-geostrophic equations and introduce their entropic approximations with regularization parameter $\varepsilon > 0$. We establish well-posedness of these approximations for every $\varepsilon > 0$ in section 3. Section 4 is devoted to convergence as $\varepsilon \to 0^+$, including the convergence analysis of the explicit discretization scheme recently proposed in [6].

# 2 Optimal and entropic optimal transport for semi-geostrophic equations

**Notations** We shall denote by $\mathscr{P}(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$, by $\mathscr{P}_c(\mathbb{R}^d)$ the set of compactly supported Borel probability measures on $\mathbb{R}^d$, for $R > 0$, $B_R$ denotes the closed ball of radius $R$ of $\mathbb{R}^d$, centered at $0$ and $\mathscr{P}(B_R)$ the set of Borel probability measures supported on $B_R$. For $x \in \mathbb{R}^d$, we denote by $|x|$ the euclidean norm of $x$ and if $A$ is a $d \times d$ matrix, $\|A\|$ stands for the euclidean operator norm of $A$.

**Optimal transport (OT) and weak solutions of the semi-geostrophic system** Given $R > 0$, $\alpha$ and $\mu$ in $\mathscr{P}(B_R)$, denoting by $\Pi(\alpha, \mu)$ the set of transport plans between $\alpha$ and $\mu$ i.e. the set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having $\alpha$ and $\mu$ as marginals, the squared-2-Wasserstein distance between $\alpha$ and $\mu$ is defined as the value of the quadratic OT problem

$$W_2^2(\alpha, \mu) := \inf_{\gamma \in \Pi(\alpha, \mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma(\mathrm{d}x, \mathrm{d}y). \tag{2.1}$$

It is a classical result of Brenier [8] that, as soon as $\mu$ is absolutely continuous with respect to the Lebesgue measure, (2.1) admits a unique optimal plan $\gamma$ which is characterized by $\gamma = (\nabla \varphi, \mathrm{id})_{\#} \mu$ i.e.

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x, y) \gamma(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}^d} f(\nabla \varphi(y), y) \mu(\mathrm{d}y), \ \forall f \in C_c(\mathbb{R}^d \times \mathbb{R}^d)$$

with $\varphi$ convex (and $\nabla \varphi$ unique up to $\mu$-negligible sets). Note in particular that $\nabla \varphi_{\#} \mu = \alpha$ and defining $\psi$ as the Legendre transform of $\varphi$:

$$\psi(x) := \sup_{y \in B_R} \{x \cdot y - \varphi(y)\}, \ \forall x \in \mathbb{R}^d$$

the optimal plan $\gamma$ is supported by pairs $(x, y)$ for which $y \in \partial \psi(x)$ (where $\partial \psi(x)$ denotes the subdifferential of the convex function $\psi$ at $x$, see [28]). In particular, if we disintegrate $\gamma$ with respect to its first marginal $\alpha$:

$$\gamma(\mathrm{d}x, \mathrm{d}y) = \alpha(\mathrm{d}x) \gamma(\mathrm{d}y|x)$$

so that $\gamma(.|x)$ is a probability measure on $B_R$, its conditional expectation

$$\int_{\mathbb{R}^d} y \gamma(\mathrm{d}y|x)$$

belongs to $\partial \psi(x)$ by convexity of $\partial \psi(x)$. Note also that $\psi$ can be seen, at least formally, as a weak solution of the Monge-Ampère equation

$$\det(D^2 \psi) \mu(\nabla \psi) = \alpha.$$

Let us now fix $R_0 > 0$, $\mu_0 \in \mathscr{P}(B_{R_0})$, $\alpha_0$ in $\mathscr{P}(B_{R_0})$; let $A$ be a $d \times d$ matrix and $T > 0$ be a time-horizon. We are interested in the following evolution problem for a curve of probability measures $\alpha$: $t \in [0, T] \mapsto \alpha_t \in \mathscr{P}_c(\mathbb{R}^d)$:

$$\partial_t \alpha + \operatorname{div}(\alpha A(\operatorname{id} - \nabla \psi)) = 0, \ \alpha(0, \cdot) = \alpha_0, \det(D^2 \psi_t)\mu_0(\nabla \psi_t) = \alpha_t, \ \psi_t \text{ convex.} \tag{2.2}$$

Of course, for $d = 3$, $\mu_0$ the uniform probability measure on $\Omega$ and $A = J$, we recover the initial problem (1.1)-(1.2) after normalizing all measures by dividing them by $|\Omega|$. In view of the previous discussion, it is natural to define weak solutions of (2.2) as follows. A weak solution of (2.2) on $[0, T] \times \mathbb{R}^d$, is a continuous for $W_2$ curve of probability measures $t \in [0, T] \mapsto \alpha_t \in \mathscr{P}_c(\mathbb{R}^d)$ such that:

- $\operatorname{spt}(\alpha_t) \subset B_{2R_0 e^{\|A\|T}}$ for every $t \in [0, T]$ (which by Grönwall's lemma, is consistent with the linear growth of the velocity field: $|A(x - \nabla \psi(x))| \leq \|A\|(|x| + R_0)$ and the fact that the initial condition $\alpha_0$ is supported in $B_{R_0}$),

- for every $f \in C^1([0, T] \times \mathbb{R}^d)$, one has

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t f + Ax \cdot \nabla f]\alpha_t(\mathrm{d}x)\mathrm{d}t - \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} Ay \cdot \nabla f(t, x)\gamma_t(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t$$

$$= \int_{\mathbb{R}^d} f(T, x)\alpha_T(\mathrm{d}x) - \int_{\mathbb{R}^d} f(0, x)\alpha_0(\mathrm{d}x), \quad (2.3)$$

where $\gamma_t$ is an optimal plan between $\alpha_t$ and $\mu_0$ i.e. $\gamma_t \in \Pi(\alpha_t, \mu_0)$ and

$$W_2^2(\alpha_t, \mu_0) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma_t(\mathrm{d}x, \mathrm{d}y), \text{ for a.e. } t \in [0, T]. \tag{2.4}$$

Note that we have not imposed that $\mu_0$ is absolutely continuous with respect to the Lebesgue measure here (so that there may not exist optimal transport maps between $\mu_0$ and $\alpha_t$ which is the reason the weak formulation above has been written in terms of-possibly non unique-optimal plans).

**Entropic OT and entropic semi-geostrophic equations** Given $\varepsilon > 0$, $\alpha$ and $\mu$ in $\mathscr{P}_c(\mathbb{R}^d)$, the entropic regularization of the quadratic OT problem (2.1) reads

$$\operatorname{OT}_\varepsilon(\alpha, \mu) := \inf_{\gamma \in \Pi(\alpha, \mu)} \left\{ \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma(\mathrm{d}x, \mathrm{d}y) + \varepsilon H(\gamma | \alpha \otimes \mu) \right\} \tag{2.5}$$

where $H(p|q)$ denotes the relative entropy of $p$ with respect to $q$:

$$H(p|q) = \begin{cases} \int_{\mathbb{R}^d} \log\left(\frac{\mathrm{d}p}{\mathrm{d}q}(x)\right)p(\mathrm{d}x) & \text{if } p \ll q, \\ +\infty & \text{otherwise.} \end{cases}$$

There exists a unique optimal plan $\gamma^\varepsilon$ for $\mathrm{OT}_\varepsilon(\alpha, \mu)$ and it is well-known that $\gamma^\varepsilon$ has the Gibbs form

$$\gamma^\varepsilon(\mathrm{d}x, \mathrm{d}y) = \exp\left(-\frac{|x-y|^2}{2\varepsilon} + \frac{u^\varepsilon(y) + v^\varepsilon(x)}{\varepsilon}\right)\alpha(\mathrm{d}x)\mu(\mathrm{d}y) \qquad (2.6)$$

where the potentials $u^\varepsilon$ and $v^\varepsilon$ are such that $\gamma^\varepsilon \in \Pi(\alpha, \mu)$ i.e. satisfy the so-called Schrödinger system

$$v^\varepsilon(x) = -\varepsilon \log\left(\int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2\varepsilon} + \frac{u^\varepsilon(y)}{\varepsilon}} \mu(\mathrm{d}y)\right) \text{ for } \alpha\text{-a.e. } x, \qquad (2.7)$$

and

$$u^\varepsilon(y) = -\varepsilon \log\left(\int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2\varepsilon} + \frac{v^\varepsilon(x)}{\varepsilon}} \alpha(\mathrm{d}x)\right) \text{ for } \mu\text{-a.e. } y. \qquad (2.8)$$

These potentials are called Schrödinger potentials for the entropic optimal transport problem between $\alpha$ and $\mu$. For an overview of entropic optimal transport, we refer the reader to [23], the lecture notes [26] and the rerences therein. Note that formulas (2.7) and (2.8) can be extended to define smooth maps on the whole of $\mathbb{R}^d$ (which are defined uniquely up to additive constants) and that disintegrating the optimal entropic plan $\gamma^\varepsilon$ as

$$\gamma^\varepsilon(\mathrm{d}x, \mathrm{d}y) = \alpha(\mathrm{d}x)\gamma^\varepsilon(\mathrm{d}y|x)$$

one has for every $x \in \mathbb{R}^d$

$$\nabla v^\varepsilon(x) = \frac{\int_{\mathbb{R}^d}(x-y)e^{-\frac{|x-y|^2}{2\varepsilon} + \frac{u^\varepsilon(y)}{\varepsilon}}\mu(\mathrm{d}y)}{\int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2\varepsilon} + \frac{u^\varepsilon(y)}{\varepsilon}}\mu(\mathrm{d}y)} = x - \int_{\mathbb{R}^d} y\gamma^\varepsilon(\mathrm{d}y|x) \qquad (2.9)$$

and

$$D^2 v^\varepsilon(x) = \mathrm{id} - \varepsilon^{-1}\left(\int_{\mathbb{R}^d} y \otimes y\gamma^\varepsilon(\mathrm{d}y|x) - \int_{\mathbb{R}^d} y\gamma^\varepsilon(\mathrm{d}y|x) \otimes \int_{\mathbb{R}^d} y\gamma^\varepsilon(\mathrm{d}y|x)\right). \tag{2.10}$$

Before defining the entropic regularization of the semi-geostrophic system, let us recall two results that will be useful in the following. The first result follows from Proposition 3.1 of [10] which gives an error estimate of the form $\mathrm{OT}_\varepsilon - \frac{1}{2}W_2^2 = O(\varepsilon|\log(\varepsilon)|)$ or the (more precise) results from [25] where an $O(\varepsilon)$ suboptimality estimate for entropic plans is established. A non quantitative consequence of these results can be stated as:

**Proposition 2.1.** *Let $R > 0$ and for $\alpha, \beta \in \mathscr{P}(B_R)^2$, denote by $\gamma^\varepsilon_{\alpha,\beta}$ the solution of the entropic optimal transport problem between $\alpha$ and $\beta$. Then $\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \gamma^\varepsilon_{\alpha,\beta}(\mathrm{d}x, \mathrm{d}y)$ converges to $W_2^2(\alpha, \beta)$ uniformly in $\alpha, \beta \in \mathscr{P}(B_R)^2$ as $\varepsilon \to 0$.*

The second result concerns the stability of the Schrödinger potentials with respect to the marginals[1] and follows from Theorem 1.1 of [9]:

**Proposition 2.2.** *Let $R > 0$ and $\mu_0 \in \mathscr{P}(B_R)$. For $\alpha \in \mathscr{P}(B_R)$, let $v^\varepsilon[\alpha]$ be the Schrödinger potential for the entropic optimal transport problem between $\alpha$ and $\mu_0$. Then, there is a constant $\Delta(\varepsilon, R)$ depending on $\varepsilon$ and $R$ such that for every $\alpha_1, \alpha_2 \in \mathscr{P}(B_R)^2$, we have*

$$\|\nabla v^\varepsilon[\alpha_1] - \nabla v^\varepsilon[\alpha_2]\|_{L^\infty(B_R)} \leq \Delta(\varepsilon, R) W_2(\alpha_1, \alpha_2). \tag{2.11}$$

Given $T > 0$, $R_0 > 0$, $\mu_0$ and $\alpha_0$ in $\mathscr{P}(B_{R_0})$, we now consider the following entropic regularization with parameter $\varepsilon > 0$ of (2.2)

$$\partial_t \alpha^\varepsilon + \mathrm{div}(\alpha^\varepsilon A(\nabla v^\varepsilon)) = 0 \text{ in } [0, T] \times \mathbb{R}^d, \ \alpha^\varepsilon(0, \cdot) = \alpha_0, \tag{2.12}$$

where

$$\nabla v_t^\varepsilon(x) = x - \int_{\mathbb{R}^d} y \gamma_t^\varepsilon(\mathrm{d}y|x) \tag{2.13}$$

and $\gamma_t^\varepsilon$ is the solution of $\mathrm{OT}^\varepsilon(\alpha_t^\varepsilon, \mu_0)$:

$$\gamma_t^\varepsilon \in \Pi(\alpha_t^\varepsilon, \mu_0), \ \mathrm{OT}_\varepsilon(\alpha_t^\varepsilon, \mu_0) = \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) + \varepsilon H(\gamma_t^\varepsilon | \alpha_t^\varepsilon \otimes \mu_0). \tag{2.14}$$

Since $\gamma_t^\varepsilon$ has second marginal $\mu_0$ supported on $B_{R_0}$, the conditional probability $\gamma_t^\varepsilon(.|x)$ is supported on $B_{R_0}$, as well so that $\nabla v^\varepsilon$ has controlled linear growth

$$|\nabla v_t^\varepsilon(x)| \leq |x| + R_0, \ \forall x \in \mathbb{R}^d, \ \forall t. \tag{2.15}$$

Therefore one defines weak solutions of (2.12)-(2.13)-(2.14) in a similar way as we did for (2.2). A weak solution of (2.12)-(2.13)-(2.14) on $[0, T] \times \mathbb{R}^d$, is a continuous for $W_2$ curve of probability measures $t \in [0, T] \mapsto \alpha_t^\varepsilon \in \mathscr{P}_c(\mathbb{R}^d)$ such that:

- $\mathrm{spt}(\alpha_t^\varepsilon) \subset B_{2R_0 e^{\|A\|T}}$ for every $t \in [0, T]$,

- for every $f \in C^1([0, T] \times \mathbb{R}^d)$, one has

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t f + Ax \cdot \nabla f] \alpha_t^\varepsilon(\mathrm{d}x) \mathrm{d}t - \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} Ay \cdot \nabla f(t, x) \gamma_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) \mathrm{d}t$$

$$= \int_{\mathbb{R}^d} f(T, x) \alpha_T^\varepsilon(\mathrm{d}x) - \int_{\mathbb{R}^d} f(0, x) \alpha_0(\mathrm{d}x), \quad (2.16)$$

where for a.e. $t \in [0, T]$, $\gamma_t^\varepsilon$ is the solution of $\mathrm{OT}_\varepsilon(\alpha_t^\varepsilon, \mu_0)$ i.e. (2.14) holds.

---

[1]It is worth mentioning that the constant $\Delta(\varepsilon, R)$ in (2.11) derived from the analysis of [9] scales exponentially badly with $\varepsilon$. Interestingly, in the very recent paper [16] close (but not obviously applicable to our context) results are obtained with much better constants of the order of $R^2 \varepsilon^{-1}$.

# 3 Well-posedness for fixed $\varepsilon > 0$

To prove well-posedness of entropic semi-geostrophic equations, we will show that these equations enjoy nice Lipschitz properties (detailed in paragraph 3.1 below) for which an elementary fixed-point argument can be used.

## 3.1 A fixed-point argument

Consider a map $B : \alpha \in \mathscr{P}_c(\mathbb{R}^d) \mapsto B[\alpha] \in C(\mathbb{R}^d, \mathbb{R}^d)$ with the following properties:

- **(H1)** There exists $C > 0$ such that

$$|B[\alpha](x)| \leq C(1 + |x|), \ \forall (x, \alpha) \in \mathbb{R}^d \times \mathscr{P}_c(\mathbb{R}^d). \tag{3.1}$$

- **(H2)** For every $R > 0$

$$K_R := \sup\{\mathrm{Lip}(B[\alpha], B_R), \ \alpha \in \mathscr{P}(B_R)\} < +\infty. \tag{3.2}$$

- **(H3)** For every $R > 0$

$$M_R := \sup\left\{\frac{\|B[\alpha^1] - B[\alpha^2]\|_{L^\infty(B_R)}}{W_2(\alpha^1, \alpha^2)}, \ (\alpha^1, \alpha^2) \in \mathscr{P}(B_R)^2, \ \alpha^1 \neq \alpha^2\right\} < +\infty. \tag{3.3}$$

**Lemma 3.1.** *Let $B : \alpha \in \mathscr{P}_c(\mathbb{R}^d) \mapsto B[\alpha] \in C(\mathbb{R}^d, \mathbb{R}^d)$ satisfy **(H1)**-**(H2)**-**(H3)**, $R_0 > 0$, $\alpha_0 \in \mathscr{P}(B_{R_0})$ and $T > 0$. Setting $R_T := (R_0 + 1)e^{CT}$, there exists a unique curve $t \in [0, T] \mapsto \alpha_t \in \mathscr{P}(B_{R_T})$ which solves*

$$\partial_t \alpha + \mathrm{div}(\alpha B[\alpha]) = 0, \ \alpha(0, \cdot) = \alpha_0, \tag{3.4}$$

*in the weak sense on $[0, T] \times \mathbb{R}^d$.*

*Proof.* Set $K = K_{R_T}$, $M = M_{R_T}$ (see (3.2) and (3.3)) and

$$\lambda := \frac{5M}{2} + K. \tag{3.5}$$

Define

$$E_T := C([0, T], (\mathscr{P}(B_{R_T}), W_2))$$

which, equipped with the distance

$$\mathrm{dist}(\alpha^1, \alpha^2) := \sup_{t \in [0,T]} e^{-\lambda t} W_2(\alpha_t^1, \alpha_t^2), \ (\alpha^1, \alpha_2) = (\alpha_t^1, \alpha_t^2)_{t \in [0,T]} \in E_T^2$$

is a complete metric space. By well-known arguments on the continuity equation (see Chapter 8 in [4]), our assumptions on $B$ imply that $\alpha \in E_T$ solves (3.4) if and only if

$$\alpha = \Phi_{\alpha_0}(\alpha) \text{ with } \Phi_{\alpha_0}(\alpha)_t := X_t^\alpha {}_\# \alpha_0, \ t \in [0, T]$$

where $X_t^\alpha$ is the (globally well-defined) flow of $B[\alpha]$:

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t^\alpha(x) = B[\alpha_t](X_t^\alpha(x)), X_0^\alpha(x) = x, \ (t, x) \in [0, T] \times \mathbb{R}^d. \qquad (3.6)$$

By Grönwall's Lemma and (3.1) $\Phi_{\alpha_0}$ is a self map of $E_T$, we shall now prove that it is a contraction. Let $(\alpha^1, \alpha^2) \in E_T^2$, to shorten notation set $B^i(t, x) = B[\alpha_t^i](x)$, $i = 1, 2$ and let $X_t^1$, $X_t^2$ denote the flows of $B^1$ and $B^2$ respectively. Then, we first obviously have

$$W_2^2(\Phi_{\alpha_0}(\alpha^1)_t, \Phi_{\alpha_0}(\alpha^2)_t) = W_2^2(X_t^1 {}_\# \alpha_0, X_t^2 {}_\# \alpha_0) \le \int_{\mathbb{R}^d} |X_t^1(x) - X_t^2(x)|^2 \alpha_0(\mathrm{d}x)$$
$$(3.7)$$

Next we observe that for fixed $x \in B_{R_0}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} |X_t^1(x) - X_t^2(x)|^2 = 2(X_t^1(x) - X_t^2(x)) \cdot (B^1(t, X_t^1(x)) - B^2(t, X_t^2(x)))$$
$$= 2(X_t^1(x) - X_t^2(x)) \cdot (B^1(t, X_t^1(x)) - B^1(t, X_t^2(x)))$$
$$+ 2(X_t^1(x) - X_t^2(x)) \cdot (B^1(t, X_t^2(x)) - B^2(t, X_t^2(x)))$$
$$\le 2K|X_t^1(x) - X_t^2(x)|^2 + 2MW_2(\alpha_t^1, \alpha_t^2)|X_t^1(x) - X_t^2(x)|$$
$$\le (2K + M)|X_t^1(x) - X_t^2(x)|^2 + M \operatorname{dist}^2(\alpha^1, \alpha^2) e^{2\lambda t}$$

where we have used (3.2)-(3.3) and Cauchy-Schwarz for the first inequality and Young's inequality and the definition of dist in the last line. Thanks to Grönwall's Lemma, we get by our choice of $\lambda$ in (3.5):

$$|X_t^1(x) - X_t^2(x)|^2 e^{-2\lambda t} \le \frac{M \operatorname{dist}^2(\alpha^1, \alpha^2)}{2\lambda - 2K - M} = \frac{\operatorname{dist}^2(\alpha^1, \alpha^2)}{4}.$$

Together with (3.7), integrating the previous inequality with respect to $\alpha_0$ yields

$$\operatorname{dist}^2(\Phi_{\alpha_0}(\alpha^1), \Phi_{\alpha_0}(\alpha^2)) = \sup_{t \in [0, T]} e^{-2\lambda t} W_2^2(\Phi_{\alpha_0}(\alpha^1)_t, \Phi_{\alpha_0}(\alpha^2)_t) \le \frac{\operatorname{dist}^2(\alpha^1, \alpha^2)}{4}$$

which shows that $\Phi$ is $\frac{1}{2}$-Lipschitz hence a contraction for dist. The existence and uniqueness of a fixed-point follows from the Banach-Picard fixed-point Theorem which completes the proof. $\qquad \square$

*Remark* 3.2. The above proof also straighforwardly implies the Lipschitz dependence of the solution of (3.4) with respect to the initial condition. Indeed, let $\alpha_0$ and $\widetilde{\alpha}_0$ be in $\mathscr{P}(B_{R_0})$ $t \in [0, T] \mapsto \alpha_t$ and $t \in [0, T] \mapsto \widetilde{\alpha}_t$ denote the solution of (3.4) with respective initial conditions $\alpha_0$ and $\widetilde{\alpha}_0$. Then, defining $K$, $M$, $\lambda$, dist, $\Phi_{\alpha_0}$ and $\Phi_{\widetilde{\alpha}_0}$ as in the proof of Lemma 3.1, we have thanks to the triangle inequality and the fact that $\Phi_{\alpha_0}$ is $\frac{1}{2}$-Lipschitz

$$\text{dist}(\alpha, \widetilde{\alpha}) = \text{dist}(\Phi_{\alpha_0}(\alpha), \Phi_{\widetilde{\alpha}_0}(\widetilde{\alpha})) \leq \text{dist}(\Phi_{\alpha_0}(\alpha), \Phi_{\alpha_0}(\widetilde{\alpha})) + \text{dist}(\Phi_{\alpha_0}(\widetilde{\alpha}), \Phi_{\widetilde{\alpha}_0}(\widetilde{\alpha}))$$

$$\leq \frac{1}{2}\text{dist}(\alpha, \widetilde{\alpha}) + \text{dist}(\Phi_{\alpha_0}(\widetilde{\alpha}), \Phi_{\widetilde{\alpha}_0}(\widetilde{\alpha}))$$

Hence

$$\text{dist}(\alpha, \widetilde{\alpha}) \leq 2\,\text{dist}(\Phi_{\alpha_0}(\widetilde{\alpha}), \Phi_{\widetilde{\alpha}_0}(\widetilde{\alpha})) \leq 2W_2(\alpha_0, \widetilde{\alpha}_0) \sup_{t \in [0,T]} e^{-\lambda t}\text{Lip}(X_t^{\widetilde{\alpha}}, B_{R_T})$$

but it is straightforward to deduce from **(H2)** and Grönwall's Lemma, that $\text{Lip}(X_t^{\widetilde{\alpha}}, B_{R_T}) \leq e^{Kt}$ and since $\lambda - K = \frac{5M}{2} \geq 0$ we arrive at

$$\text{dist}(\alpha, \widetilde{\alpha}) \leq 2W_2(\alpha_0, \widetilde{\alpha}_0). \tag{3.8}$$

## 3.2 Well-posedness of entropic semi-geostrophic equations

The fact that (2.12)-(2.13)-(2.14) is well-posed then directly follows from Lemma 3.1 and regularity properties of entropic OT:

**Theorem 3.3.** *Given $T > 0$, $R_0 > 0$, $\alpha_0$ and $\mu_0$ in $\mathscr{P}(B_{R_0})$, for every $\varepsilon > 0$, the system (2.12)-(2.13)-(2.14) admits a unique weak solution on $[0, T] \times \mathbb{R}^d$.*

*Proof.* Given $\alpha \in \mathscr{P}_c(\mathbb{R}^d)$, denote by $\gamma^\varepsilon$ the solution of $\text{OT}_\varepsilon(\alpha, \mu_0)$ (equivalently solve the Schrödinger system (2.7)-(2.8) to obtain the potential $v^\varepsilon$) and set

$$B^\varepsilon[\alpha](x) := A(x - \int_{\mathbb{R}^d} y\gamma^\varepsilon(\mathrm{d}y|x)) = A(\nabla v^\varepsilon(x)), \ \forall x \in \mathbb{R}^d. \tag{3.9}$$

Let us show that the map $\alpha \mapsto B^\varepsilon[\alpha]$ satisfies **(H1)**-**(H2)**-**(H3)**. Concerning **(H1)**, the linear growth condition (3.1) follows from the fact that $\gamma^\varepsilon(.|x)$ has support in $B_{R_0}$, so that $|B^\varepsilon[\alpha](x)| \leq \|A\|(|x| + R_0)$ hence **(H1)** holds with the constant $C = \max(1, R_0)\|A\|$. As for **(H2)**, this also follows from the fact that $\gamma^\varepsilon(.|x)$ has support in $B_{R_0}$ and the expression of the hessian of $v^\varepsilon$ in (2.10) which yields the bound (exploding as $\varepsilon \to 0$)

$$\|DB[\alpha]\| \leq \|A\|\|D^2 v^\varepsilon\| \leq \|A\|(1 + 2\varepsilon^{-1}R_0^2).$$

10

Therefore **(H2)** holds with the constant $K_R = \|A\|(1 + 2\varepsilon^{-1}R_0^2)$ (depending on $\varepsilon$ but not $R$). Finally, it follows from Proposition 2.2 that $B^\varepsilon$ satisfies **(H3)** with the constant $M_R = \Lambda(\varepsilon, R)\|A\|$, $\Lambda(\varepsilon, R)$ being the constant in (2.11). Thanks to Lemma 3.1, we deduce that there exists a unique solution $\alpha^\varepsilon$ of the system (2.12)-(2.13)-(2.14) with $\alpha_t^\varepsilon$ supported for every $t \in [0, T]$ on the ball of radius $(R_0 + 1)e^{\max(1,R_0)\|A\|T}$ but a direct application of Grönwall's Lemma gives that it has to be supported on the ball of radius $2R_0 e^{\|A\|T}$ as in our initial definition in section 2. $\qquad\square$

## 3.3  On conservation of energy

In the physically relevant case $d = 3$, $A = J$, one can define the (entropically regularized) total energy

$$E_\varepsilon(\alpha) := \mathrm{OT}_\varepsilon(\alpha, \mu_0) + \int_{\mathbb{R}^3} x_3 \alpha.$$

Of course, the potential energy i.e. the second term is preserved along the flow $\alpha^\varepsilon$ of the entropic semi-geostrophic equation (obtained by Theorem 3.3), by taking a test function depending on $x_3$ only, we readily see that the marginal of $\alpha^\varepsilon$ in the $x_3$-variable is constant in time. As for the conservation of $t \mapsto \mathrm{OT}_\varepsilon(\alpha_t^\varepsilon, \mu_0)$, one can argue as follows.

Using a well-known dual formulation (see [26], [22], [23]) of entropic optimal transport (or by a direct computation using (2.7)-(2.8)), given $\beta \in \mathscr{P}_c(\mathbb{R}^3)$, one can also write $\mathrm{OT}_\varepsilon(\beta, \mu_0)$ as the supremum of an unconstrained concave maximization problem

$$\varepsilon + \sup_{u,v} \left\{ \int_{\mathbb{R}^3} u \mathrm{d}\mu_0 + \int_{\mathbb{R}^3} v \mathrm{d}\beta - \varepsilon \int_{\mathbb{R}^3 \times \mathbb{R}^3} e^{\frac{-|x-y|^2}{2\varepsilon}} e^{\frac{v(x)+u(y)}{\varepsilon}} \beta(\mathrm{d}x)\mu_0(\mathrm{d}y) \right\}$$

and this supremum is achieved when $u$ and $v$ are Schrödinger potentials between $\beta$ and $\mu_0$. In particular, if we denote by $v^\varepsilon = v^\varepsilon[\beta]$ and $u^\varepsilon = u^\varepsilon[\beta]$ these potentials, when $\eta$ is another compactly supported probability measure, we have

$$\mathrm{OT}_\varepsilon(\eta, \mu_0) - \mathrm{OT}_\varepsilon(\beta, \mu_0) \geq \int_{\mathbb{R}^3} v^\varepsilon(\eta - \beta)$$
$$- \varepsilon \int_{\mathbb{R}^3} \left( \int_{\mathbb{R}^3} e^{\frac{-|x-y|^2}{2\varepsilon}} e^{\frac{v^\varepsilon(x)+u^\varepsilon(y)}{\varepsilon}} \mu_0(\mathrm{d}y) \right)(\eta - \beta)(\mathrm{d}x)$$
$$= \int_{\mathbb{R}^3} v^\varepsilon(\eta - \beta)$$

$$(3.10)$$

where passing to the last line, we have used the fact that $v^\varepsilon$ and $u^\varepsilon$ are Schrödinger potentials between $\beta$ and $\mu_0$ (so that (2.7) holds for $\mu = \mu_0$) and that $\eta$ and $\beta$ have same total mass. For $t$ and $s$ in $[0, T]$, applying (3.10) to the measures $\alpha_s^\varepsilon$, $\alpha_t^\varepsilon$ and the potential $v_t^\varepsilon = v^\varepsilon[\alpha_t^\varepsilon]$, we get, using the fact that $\alpha^\varepsilon$ solves the semi entropic semi-geostrophic equation in the second line, the fact that $J$ is skew-symmetric in the third line, and finally the fact that $\beta \mapsto \nabla v^\varepsilon[\beta]$ satisfies **(H3)** (thanks to Proposition 2.2) and that $\tau \mapsto \alpha_\tau^\varepsilon$ is Lipschitz for $W_2$ in the last line

$$
\begin{aligned}
\mathrm{OT}_\varepsilon(\alpha_s^\varepsilon, \mu_0) - \mathrm{OT}_\varepsilon(\alpha_t^\varepsilon, \mu_0) &\geq \int_{\mathbb{R}^3} v_t^\varepsilon(\alpha_s^\varepsilon - \alpha_t^\varepsilon) \\
&= \int_t^s \int_{\mathbb{R}^3} \nabla v_t^\varepsilon(x) J(\nabla v_\tau^\varepsilon(x)) \alpha_\tau^\varepsilon(\mathrm{d}x)\mathrm{d}\tau \\
&= \int_t^s \int_{\mathbb{R}^3} \nabla v_t^\varepsilon(x) J(\nabla v_\tau^\varepsilon(x) - \nabla v_t^\varepsilon(x)) \alpha_\tau^\varepsilon(\mathrm{d}x)\mathrm{d}\tau \\
&\geq -|s-t| \|\nabla v^\varepsilon\|_{L^\infty(B_{2R_0 e^T})} \sup_{\tau \in [t,s]} \|\nabla v_\tau^\varepsilon - \nabla v_t^\varepsilon\|_{L^\infty(B_{2R_0 e^T})} \\
&\geq -C_\varepsilon(s-t)^2
\end{aligned}
$$

which, reversing the role of $s$ and $t$, yields $\mathrm{OT}_\varepsilon(\alpha_s^\varepsilon, \mu_0) - \mathrm{OT}_\varepsilon(\alpha_t^\varepsilon, \mu_0) = O((s-t)^2)$ hence the preservation of $\mathrm{OT}_\varepsilon(., \mu_0)$ and $E_\varepsilon$ along $\alpha^\varepsilon$.

## 3.4 Discretization

The fixed-point argument of paragraph 3.1 is useful to derive well-posedness for entropic approximations of semi-geostrophic equations as we saw in paragraph 3.2, it is however useless in practice to design numerical simulations. In this paragraph, we therefore aim to discuss the convergence of an explicit scheme for (3.4) which is in line with [6] and where:

- the initial condition is approximated by some empirical measure (which makes pushforward operations explicit)

- the flow of $B[\alpha]$ is discretized in time by its explicit Euler counterpart (for entropic semi-geostrophic equations and for a discrete $\alpha$, computing $B[\alpha]$ amounts to compute the Schrödinger potential between $\alpha$ and $\mu_0$ which can be done efficiently by Sinkhorn's algorithm as in [6]).

More precisely, given $\alpha_0 \in \mathscr{P}(B_{R_0})$ and $T > 0$, we denote by $\alpha$ the solution of (3.4) obtained as in Lemma 3.1. Given $\widetilde{\alpha}_0 \in \mathscr{P}(B_{R_0})$ (which we can think of as a quantized version of $\alpha_0$), $N \in \mathbb{N}^*$ and $\tau = \frac{T}{N}$ a time step, we

consider the sequence of measures $\alpha_k^\tau$, $k = 0, \ldots, N - 1$ defined recursively by

$$\alpha_0^\tau = \widetilde{\alpha}_0, \ \alpha_{k+1}^\tau := (\mathrm{id} + \tau B[\alpha_k^\tau])_\# \alpha_k^\tau, \ k = 0, \ldots, N - 1. \tag{3.11}$$

Note that whenever $\widetilde{\alpha}_0$ is an empirical measure so is $\alpha_k^\tau$ and that $\alpha_k^\tau$ remains supported on $B_{R_T}$ with $R_T = (R_0 + 1)e^{CT}$ and $C$ is the constant appearing in (3.1). Let us finally denote by $\widetilde{\alpha}^\tau$ the piecewise constant interpolation $t \in [0, T) \mapsto \widetilde{\alpha}_t^\tau$:

$$\widetilde{\alpha}_t^\tau = \alpha_k^\tau, \ t \in [k\tau, (k+1)\tau), \ k = 0, \ldots, N - 1. \tag{3.12}$$

Then we have the following error estimate between the solution $\alpha$ of (3.4) and its discretized counterpart $\widetilde{\alpha}^\tau$:

**Lemma 3.4.** *Assuming $B$ satisfies* **(H1)-(H2)-(H3)**, *there is a positive constant $\Lambda$ depending on $T$, $R_T = (R_0 + 1)e^{CT}$, $C$, $K = K_{R_T}$ and $M = M_{R_T}$ such that, for $\alpha$ and $\widetilde{\alpha}^\tau$ as above, one has*

$$\sup_{t \in [0,T)} W_2(\alpha_t, \widetilde{\alpha}_t^\tau) \le \Lambda(\tau + W_2(\alpha_0, \widetilde{\alpha}_0)). \tag{3.13}$$

*Proof.* For notational simplicity, throughout this proof we just write $R = R_T$. Setting $L := \sup_{\beta \in \mathscr{P}(B_R)} \|B[\beta]\|_{L^\infty(B_R)}$, we first obviously have

$$W_2(\alpha_t, \alpha_s) \le L|t - s|, \ (s, t) \in [0, T]^2. \tag{3.14}$$

Hence

$$\sup_{t \in [0,T)} W_2(\alpha_t, \widetilde{\alpha}_t^\tau) \le \max_{k=0,\ldots,N} W_2(\alpha_{k\tau}, \alpha_k^\tau) + L\tau. \tag{3.15}$$

For $k = 0, \ldots, N$ set $d_k := W_2(\alpha_{k\tau}, \alpha_k^\tau)$. For $k = 0, \ldots, N - 1$, notice that $\alpha_{(k+1)\tau} = X_{k\#}^\tau \alpha_{k\tau}$ where $X_k^\tau := Y_{(k+1)\tau}$ and

$$\frac{d}{ds} Y_s(x) = B[\alpha_s](Y_s(x)), \ Y_{k\tau}(x) = x$$

so that

$$X_k^\tau(x) = x + \tau B[\alpha_{k\tau}](x) + \int_{k\tau}^{(k+1)\tau} (B[\alpha_s](Y_s(x)) - B[\alpha_{k\tau}](x)) \mathrm{d}s$$

using **(H2)-(H3)** and the fact that $|Y_s(x) - x| \le L(s - k\tau)$ for every $(x, s) \in B_R \times [k\tau, (k+1)\tau]$, we arrive at

$$\|X_k^\tau - (\mathrm{id} + \tau B[\alpha_{k\tau}])\|_{L^\infty(B_R)} \le (K + M)L\tau^2. \tag{3.16}$$

13

By the triangle inequality, we have

$$
\begin{aligned}
d_{k+1} &= W_2(X^\tau_{k\#}\alpha_{k\tau}, (\mathrm{id} + \tau B[\alpha^\tau_k])_{\#}\alpha^\tau_k) \\
&\leq W_2(X^\tau_{k\#}\alpha_{k\tau}, (\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha_{k\tau}) \\
&\quad + W_2((\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha_{k\tau}, (\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha^\tau_k) \\
&\quad + W_2((\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha^\tau_k, (\mathrm{id} + \tau B[\alpha^\tau_k])_{\#}\alpha^\tau_k).
\end{aligned}
$$

Firstly, (3.16) yields

$$
W_2(X^\tau_{k\#}\alpha_{k\tau}, (\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha_{k\tau}) \leq \|X^\tau_k - (\mathrm{id} + \tau B[\alpha_{k\tau}])\|_{L^\infty(B_R)} \leq (K+M)L\tau^2.
$$

Secondly, assumption **(H2)** entails

$$
\begin{aligned}
W_2((\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha_{k\tau}, (\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha^\tau_k) &\leq \mathrm{Lip}(\mathrm{id} + \tau B[\alpha_{k\tau}], B_R)W_2(\alpha_{k\tau}, \alpha^\tau_k) \\
&\leq (1 + K\tau)d_k.
\end{aligned}
$$

Finally, we deduce from **(H3)**

$$
W_2((\mathrm{id} + \tau B[\alpha_{k\tau}])_{\#}\alpha^\tau_k, (\mathrm{id} + \tau B[\alpha^\tau_k])_{\#}\alpha^\tau_k) \leq \tau \|B[\alpha_{k\tau}] - B[\alpha^\tau_k]\|_{L^\infty(B_R)} \leq M\tau d_k
$$

which enables us to conclude that

$$
d_{k+1} \leq (K+M)L\tau^2 + (1 + (M+K)\tau)d_k
$$

which, by the discrete Grönwall's Lemma, yields that for $k = 0, \ldots, N$

$$
d_k \leq (1 + (K+M)\tau)^k(L\tau + d_0) \leq e^{(K+M)T}(L\tau + W_2(\alpha_0, \widetilde{\alpha}_0))
$$

together with (3.15), this proves the desired inequality (3.13). $\qquad\square$

*Remark* 3.5. If we take $\widetilde{\alpha}_0 = \alpha_0$ in Lemma 3.4, the convergence of the theoretical Euler scheme is a well known fact, even in frameworks with much less regularity (see [11]). However, the interest of lemma 3.4 is to give a convergence rate that is linear in $\tau$ which is natural to expect in the case of the first-order Euler scheme with enough regularity. Since [9] provides bounds for higher derivatives of the entropic potentials, the proof could also be adapted to higher order explicit schemes such as Runge-Kutta and provide better convergence rates.

# 4 Convergence to a weak solution as $\varepsilon \to 0^+$

## 4.1 Convergence of continuous trajectories

For fixed $T > 0$ and $\varepsilon > 0$, let $\alpha^\varepsilon \in C([0,T], (\mathscr{P}(B_{2R_0 e^{\|A\|T}}), W_2))$ be the unique weak solution of (2.12)-(2.13)-(2.14) obtained as in Theorem 3.3. To simplify notation, let us set $R := 2R_0 e^{\|A\|T}$.

Since the velocity $A(\nabla v^\varepsilon)$ can be bounded in $L^\infty(B_R)$ uniformly in $\varepsilon$, there is a constant $\kappa > 0$ such that

$$W_2(\alpha_t^\varepsilon, \alpha_s^\varepsilon) \le \kappa|t - s|, \ \forall \varepsilon > 0, \ \forall (s,t) \in [0,T]^2.$$

Thanks to Arzèla-Ascoli Theorem, passing to a vanishing sequence $\varepsilon_n \to 0$ if necessary, one may assume that, for some $\alpha = (\alpha_t)_{t \in [0,T]} \in C([0,T], (\mathscr{P}(B_R), W_2))$, one has

$$\sup_{t \in [0,T]} W_2(\alpha_t^\varepsilon, \alpha_t) \to 0 \text{ as } \varepsilon \to 0^+. \tag{4.1}$$

Our aim is to show that this cluster point $\alpha$ is a weak solution of the un-regularized semi-geostrophic like system (2.2). By the very definition of $\alpha^\varepsilon$, for $f \in C^1([0,T] \times \mathbb{R}^d)$, one has

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t f + Ax \cdot \nabla f]\alpha_t^\varepsilon(\mathrm{d}x)\mathrm{d}t - \int_{\mathbb{R}^d} f(T,x)\alpha_T^\varepsilon(\mathrm{d}x) + \int_{\mathbb{R}^d} f(0,x)\alpha_0(\mathrm{d}x)$$
$$= \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} Ay \cdot \nabla f(t,x)\gamma_t^\varepsilon(\mathrm{d}x,\mathrm{d}y)\mathrm{d}t \quad (4.2)$$

where for a.e. $t \in [0,T]$, $\gamma_t^\varepsilon$ is the solution of $\mathrm{OT}_\varepsilon(\alpha_t^\varepsilon, \mu_0)$. Obviously the left-hand side of (4.2) converges as $\varepsilon \to 0^+$ to

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t f + Ax \cdot \nabla f]\alpha_t(\mathrm{d}x)\mathrm{d}t - \int_{\mathbb{R}^d} f(T,x)\alpha_T(\mathrm{d}x) + \int_{\mathbb{R}^d} f(0,x)\alpha_0(\mathrm{d}x)$$

As for the right-hand side, by Banach-Alaoglu's Theorem, one may assume that the family of measures $\gamma_t^\varepsilon(\mathrm{d}x,\mathrm{d}y) \otimes \mathrm{d}t$ weakly $*$ converges as $\varepsilon \to 0^+$ to some measure $\theta$ on $B_R^2 \times [0,T]$, since weak $*$ convergence implies weak $*$ convergence of marginals and $\gamma_t^\varepsilon \in \Pi(\alpha_t^\varepsilon, \mu_0)$ for a.e. $t \in [0,T]$, one has

$$\theta(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t) = \gamma_t(\mathrm{d}x,\mathrm{d}y) \otimes \mathrm{d}t \text{ with } \gamma_t \in \Pi(\alpha_t, \mu_0) \text{ for a.e. } t \in [0,T]$$

and then

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t f + Ax \cdot \nabla f]\alpha_t(\mathrm{d}x)\mathrm{d}t - \int_{\mathbb{R}^d} f(T,x)\alpha_T(\mathrm{d}x) + \int_{\mathbb{R}^d} f(0,x)\alpha_0(\mathrm{d}x)$$
$$= \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} Ay \cdot \nabla f(t,x)\gamma_t(\mathrm{d}x,\mathrm{d}y)\mathrm{d}t.$$

So, to establish that $\alpha$ is a weak solution of (2.2), it remains to show that $\gamma_t$ is an optimal transport plan between $\alpha_t$ and $\mu_0$ for a.e. $t \in [0, T]$. Since $\alpha_t^\varepsilon$ is supported on $B_R$ for all $t \in [0, T]$, by Proposition 2.1, $\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) - W_2^2(\alpha_t^\varepsilon, \mu_0)$ converges to 0 uniformly in $t \in [0, T]$. Thanks to (4.1), $W_2^2(\alpha_t^\varepsilon, \mu_0) \to W_2^2(\alpha_t, \mu_0)$ also uniformly in $t$, so that

$$\int_0^T W_2^2(\alpha_t, \mu_0) \mathrm{d}t = \lim_{\varepsilon \to 0} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) \mathrm{d}t$$
$$= \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma_t(\mathrm{d}x, \mathrm{d}y) \mathrm{d}t$$

where the second line follows from the weak $*$ convergence of $\gamma_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) \otimes \mathrm{d}t$ to $\gamma_t(\mathrm{d}x, \mathrm{d}y) \otimes \mathrm{d}t$, and then since $\gamma_t \in \Pi(\alpha_t, \mu_0)$, for a.e. $t \in [0, T]$, this shows that (2.4) holds so that $\alpha$ is a weak solution of (2.2). In other words, we have shown the following:

**Theorem 4.1.** *Cluster points as $\varepsilon \to 0$ of weak solutions of (2.12)-(2.13)-(2.14) are weak solutions of (2.2).*

*Remark* 4.2. Note that this gives yet another proof of existence of weak solutions of semi-geostrophic equations for compactly supported initial data by entropic approximation. When $d = 3$ and $A = J$, combining the considerations of paragraph 3.3 with the convergence of $\mathrm{OT}_\varepsilon$ to $W_2$, one can readily check that a cluster point $\alpha$ obtained as above, conserves the energy $W_2(\alpha_t, \mu_0)$ (as well as the potential energy of course). This property is actually already well-known for all weak solutions, see [3].

## 4.2   Convergence of discretizations

We finally consider the case where we both regularize the semi-geostrophic equation with an entropy i.e. replace $W_2^2$ by $\mathrm{OT}_\varepsilon$ with small $\varepsilon$ and approximate the solution of (2.12)-(2.13)-(2.14) by using an explicit Euler scheme with time step $\tau$ and quantizing the initial conditon by replacing $\alpha_0$ by some discrete measure $\widetilde{\alpha}_0$ as in paragraph 3.4. Combining Lemma 3.13 and Theorem 4.1, one easily gets that when letting first $\tau + W_2(\alpha_0, \widetilde{\alpha}_0)$ tend to 0 and then $\varepsilon \to 0$, we obtain (along a subsequence) convergence to a weak solution of (2.2). This is of little use in practice to guarantee convergence of the scheme proposed in [6], so we consider the situation where we let all parameters vanish at the same time in an arbitrary manner, i.e. for $\varepsilon > 0$, we consider a time step $\tau_\varepsilon > 0$ with $T = N_\varepsilon \tau_\varepsilon$ and a quantized approximation

of the initial $\alpha_0 \in \mathscr{P}(B_{R_0})$ and of the reference measure $\mu_0 \in \mathscr{P}(B_{R_0})$

$$\widetilde{\alpha}_0^\varepsilon := \frac{1}{M_\varepsilon} \sum_{i=1}^{M_\varepsilon} \delta_{x_i^\varepsilon}, \ x_i^\varepsilon \in B_{R_0}; \quad \widetilde{\mu}_0^\varepsilon := \frac{1}{M_\varepsilon} \sum_{i=1}^{M_\varepsilon} \delta_{y_i^\varepsilon}, \ y_i^\varepsilon \in B_{R_0},$$

and assume that

$$\tau_\varepsilon + W_2(\alpha_0, \widetilde{\alpha}_0^\varepsilon) + W_2(\mu_0, \widetilde{\mu}_0^\varepsilon) \to 0, \ \text{as } \varepsilon \to 0^+. \tag{4.3}$$

We then construct a piecewise constant curve of measures $t \in [0,T] \mapsto \widetilde{\alpha}_t^\varepsilon$ by an explicit Euler scheme as in 3.4, i.e.:

$$\widetilde{\alpha}_t^\varepsilon = \alpha_k^\varepsilon, \ t \in [k\tau_\varepsilon, (k+1)\tau_\varepsilon), \ k = 0, \dots, N_\varepsilon - 1$$

with

$$\alpha_0^\varepsilon = \widetilde{\alpha}_0^\varepsilon, \ \alpha_{k+1}^\varepsilon = (\mathrm{id} + \tau_\varepsilon B^\varepsilon[\alpha_k^\varepsilon])_\# \alpha_k^\varepsilon, \ k = 0, \dots, N_\varepsilon - 1$$

where $B^\varepsilon$ is defined through the solution of $\mathrm{OT}_\varepsilon(\alpha, \widetilde{\mu}_0^\varepsilon)$ as in (3.9). Observing that $W_2(\widetilde{\alpha}_t^\varepsilon, \widetilde{\alpha}_s^\varepsilon) \le \kappa(|t - s| + \tau)$ for every $t, s$ in $[0,T]$ and a constant $\kappa$ independent of $\varepsilon$, passing along a (not relabeled) vanishing sequence $\varepsilon_n \to 0$, we may assume that for some $\alpha = (\alpha_t)_{t \in [0,T]} \in C([0,T], (\mathscr{P}(B_R), W_2))$ (where we recall that we have set $R := 2R_0 e^{\|A\|T}$) one has

$$\sup_{t \in [0,T)} W_2(\widetilde{\alpha}_t^\varepsilon, \alpha_t) \to 0 \text{ as } \varepsilon \to 0^+, \tag{4.4}$$

and $W_2(\alpha_t, \alpha_s) \le \kappa|t - s|$, for all $s$ and $t$ in $[0,T]$. We then have the following convergence result, which shows that cluster points of the previous approximations are weak solutions of (2.2):

**Theorem 4.3.** *If $\alpha$ is obtained as a cluster point of the discretized entropic regularization $(\widetilde{\alpha}_t^\varepsilon)_{t \in [0,T)}$ i.e. such that (4.4) holds, then $\alpha$ is a weak solution of (2.2).*

*Proof.* Let $f \in C^1([0,T] \times \mathbb{R}^d)$, observe that

$$\int_0^T \int_{\mathbb{R}^d} \partial_t f \widetilde{\alpha}_t^\varepsilon = \sum_{k=0}^{N_\varepsilon - 1} \int_{\mathbb{R}^d} \int_{k\tau_\varepsilon}^{(k+1)\tau_\varepsilon} \partial_t f \alpha_k^\varepsilon = \sum_{k=0}^{N_\varepsilon - 1} \int_{\mathbb{R}^d} (f((k+1)\tau_\varepsilon, \cdot) - f(k\tau_\varepsilon, \cdot)) \alpha_k^\varepsilon$$

$$= \sum_{k=1}^{N_\varepsilon - 1} \int_{\mathbb{R}^d} f(k\tau_\varepsilon, \cdot)(\alpha_{k-1}^\varepsilon - \alpha_k^\varepsilon) + \int_{\mathbb{R}^d} f(T, \cdot) \alpha_{N-1}^\varepsilon - \int_{\mathbb{R}^d} f(0, \cdot) \alpha_0^\varepsilon \tag{4.5}$$

note that the last two terms in (4.5) converge as $\varepsilon \to 0$ respectively to $\int_{\mathbb{R}^d} f(T, \cdot) \alpha_T$ and $-\int_{\mathbb{R}^d} f(0, \cdot) \alpha_0$. Setting $B_k^\varepsilon = B^\varepsilon[\alpha_k^\varepsilon]$ and denoting by $\gamma_{k-1}^\varepsilon$

the solution of $\mathrm{OT}_\varepsilon(\alpha_{k-1}^\varepsilon, \widetilde{\mu}_0^\varepsilon)$ and using the fact that $\alpha_k^\varepsilon = (\mathrm{id} + \tau_\varepsilon B_k^\varepsilon)_{\#}\alpha_{k-1}^\varepsilon$, enables to rewrite

$$
\begin{aligned}
\int_{\mathbb{R}^d} f(k\tau_\varepsilon, \cdot)(\alpha_{k-1}^\varepsilon - \alpha_k^\varepsilon) &= \int_{\mathbb{R}^d} (f(k\tau_\varepsilon, x) - f(k\tau_\varepsilon, x + \tau_\varepsilon B_{k-1}^\varepsilon(x)))\alpha_{k-1}^\varepsilon(\mathrm{d}x) \\
&= -\tau_\varepsilon \int_{\mathbb{R}^d} \nabla f(k\tau_\varepsilon, x) \cdot B_{k-1}^\varepsilon(x)\alpha_{k-1}^\varepsilon(\mathrm{d}x) + o(\tau_\varepsilon) \\
&= \int_{(k-1)\tau_\varepsilon}^{k\tau_\varepsilon} \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla f(t, x) \cdot A(y - x)\gamma_{k-1}^\varepsilon(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t + o(\tau_\varepsilon).
\end{aligned}
$$

Considering the piecewise constant curve of plans $t \mapsto \widetilde{\gamma}_t^\varepsilon$ defined by $\widetilde{\gamma}_t^\varepsilon = \gamma_k^\varepsilon$ for $t \in [k\tau_\varepsilon, (k+1)\tau_\varepsilon)$, recalling that $N_\varepsilon \tau_\varepsilon = T$, we thus have

$$
\begin{aligned}
\int_0^T \int_{\mathbb{R}^d} \partial_t f \widetilde{\alpha}_t^\varepsilon &= \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla f(t, x) \cdot A(y - x)\widetilde{\gamma}_t^\varepsilon(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t \\
&\quad + \int_{\mathbb{R}^d} f(T, x)\alpha_T(\mathrm{d}x) - \int_{\mathbb{R}^d} f(0, x)\alpha_0(\mathrm{d}x) + o(1).
\end{aligned}
\tag{4.6}
$$

As in paragraph 4.1, we may assume (possibly after an extraction) that $\widetilde{\gamma}_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) \otimes \mathrm{d}t$ weakly $*$ converge as $\varepsilon \to 0^+$ to some measure of the form $\gamma_t(\mathrm{d}x, \mathrm{d}y) \otimes \mathrm{d}t$. Then, exactly as in paragraph 4.1, Proposition 2.1 ensures that for a.e. $t$, $\gamma_t \in \Pi(\alpha_t, \mu_0)$ is an optimal transport plan between $\alpha_t$ and $\mu_0$ i.e. satisfies (2.4). Letting $\varepsilon \to 0^+$ in (4.6), thanks to (4.4) and $\gamma_t \in \Pi(\alpha_t, \mu_0)$, we get

$$
\begin{aligned}
\int_0^T \int_{\mathbb{R}^d} [\partial_t f + Ax \cdot \nabla f]\alpha_t &= \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla f(t, x) \cdot Ay\, \gamma_t(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t \\
&\quad + \int_{\mathbb{R}^d} f(T, x)\alpha_T(\mathrm{d}x) - \int_{\mathbb{R}^d} f(0, x)\alpha_0(\mathrm{d}x)
\end{aligned}
$$

and since $\gamma_t$ is an optimal transport plan between $\alpha_t$ and $\mu_0$, $\alpha$ is a weak solution of (2.2). $\square$

# References

[1] Luigi Ambrosio, Maria Colombo, Guido De Philippis, and Alessio Figalli. Existence of Eulerian solutions to the semigeostrophic equations

in physical space: the 2-dimensional periodic case. *Comm. Partial Differential Equations*, 37(12):2209–2227, 2012.

[2] Luigi Ambrosio, Maria Colombo, Guido De Philippis, and Alessio Figalli. A global existence result for the semigeostrophic equations in three dimensional convex domains. *Discrete Contin. Dyn. Syst.*, 34(4):1251–1268, 2014.

[3] Luigi Ambrosio and Wilfrid Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.*, 61(1):18–53, 2008.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[5] Jean-David Benamou and Yann Brenier. Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem. *SIAM J. Appl. Math.*, 58(5):1450–1461, 1998.

[6] Jean-David Benamou, Colin John Cotter, and Hugo Malamut. Entropic optimal transport solutions of the semigeostrophic equations. *J. Comput. Phys.*, 500:Paper No. 112745, 16, 2024.

[7] David P. Bourne, Charlie P. Egan, Beatrice Pelloni, and Mark Wilkinson. Semi-discrete optimal transport methods for the semi-geostrophic equations. *Calc. Var. Partial Differential Equations*, 61(1):Paper No. 39, 34, 2022.

[8] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[9] Guillaume Carlier, Lénaïc Chizat, and Maxime Laborde. Displacement smoothness of entropic optimal transport, 2024. to appear in ESAIM COCV.

[10] Guillaume Carlier, Paul Pegon, and Luca Tamanini. Convergence rate of general entropic optimal transport costs. *Calc. Var. Partial Differential Equations*, 62(4):Paper No. 116, 28, 2023.

[11] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Dissipative probability vector fields and generation of evolution semigroups in wasserstein spaces. *Probability Theory and Related Fields*, 185(3):1087–1182, 2023.

[12] Michael John Priestley Cullen. *A Mathematical Theory of Large-Scale Atmosphere/Ocean Flow.* London: Imperial College Press, 2006.

[13] Michael John Priestley Cullen. *The mathematics of large-scale atmosphere and ocean.* Singapore: World Scientific, 2021.

[14] Michael John Priestley Cullen and Robert James Purser. An extended Lagrangian theory of semigeostrophic frontogenesis. *J. Atmospheric Sci.*, 41(9):1477–1497, 1984.

[15] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26, 2013.

[16] Vincent Divol, Jonathan Niles-Weed, and Aram-Alexandre Pooladian. Tight stability bounds for entropic Brenier maps, 2024. arXiv:2404.02855.

[17] Charlie P. Egan, David P. Bourne, Colin J. Cotter, Michael J. P. Cullen, Beatrice Pelloni, Steven M. Roper, and Mark Wilkinson. A new implementation of the geometric method for solving the Eady slice equations. *J. Comput. Phys.*, 469:Paper No. 111542, 30, 2022.

[18] Arnt Eliassen. The quasi-static equations of motion with pressure as independent variable. *Geofys. Publ. Norske Vid.-Akad. Oslo*, 17(3):44, 1949.

[19] Mikhail Feldman and Adrian Tudorascu. On Lagrangian solutions for the semi-geostrophic system with singular initial data. *SIAM J. Math. Anal.*, 45(3):1616–1640, 2013.

[20] Mikhail Feldman and Adrian Tudorascu. On the semi-geostrophic system in physical space with general initial data. *Arch. Ration. Mech. Anal.*, 218(1):527–551, 2015.

[21] Brian J. Hoskins. The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmosph. Science*, 32:233–242, 1975.

[22] Christian Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(4):1879–1920, 2012.

[23] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.

[24] Grégoire Loeper. A fully nonlinear version of the incompressible Euler equations: the semigeostrophic system. *SIAM J. Math. Anal.*, 38(3):795–823, 2006.

[25] Hugo Malamut and Maxime Sylvestre. Convergence rates of the regularized optimal transport : Disentangling suboptimality and entropy, 2023. arXiv:2306.06940.

[26] Marcel Nutz. Introduction to Entropic Optimal Transport. Lecture notes, Columbia University, 2022.

[27] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[28] R. Tyrrell Rockafellar. *Convex analysis*, volume No. 28 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1970.