A VARIATIONAL TIME DISCRETIZATION FOR COMPRESSIBLE EULER EQUATIONS

FABIO CAVALLETTI, MARC SEDJRO, AND MICHAEL WESTDICKENBERG

Abstract. We introduce a variational time discretization for the multi-dimensional gas dynamics equations, in the spirit of minimizing movements for curves of maximal slope. Each timestep requires the minimization of a functional measuring the acceleration of fluid elements, over the cone of monotone transport maps. We prove convergence to measure-valued solutions for the pressureless gas dynamics and the compressible Euler equations. For one space dimension, we obtain sticky particle solutions for the pressureless case.

1. Introduction

The compressible Euler equations model the dynamics of compressible fluids like gases. They form a system of hyperbolic conservation laws

$$\left.\begin{aligned} \partial_t\varrho + \nabla\cdot(\varrho\boldsymbol{v}) &= 0 \\ \partial_t(\varrho\boldsymbol{v}) + \nabla\cdot(\varrho\boldsymbol{v}\otimes\boldsymbol{v}) + \nabla p &= 0 \\ \partial_t\varepsilon + \nabla\cdot\big((\varepsilon+p)\boldsymbol{v}\big) &= 0 \end{aligned}\right\} \quad \text{in } [0,\infty)\times\mathbb{R}^d. \tag{1.1}$$

The unknowns $(\varrho,\boldsymbol{v},\varepsilon)$ depend on time $t\in[0,\infty)$ and space $x\in\mathbb{R}^d$ and we assume that suitable initial data (to be specified later) is given:

$$(\varrho,\boldsymbol{v},\varepsilon)(t=0,\cdot) =: (\bar\varrho,\bar{\boldsymbol{v}},\bar\varepsilon).$$

We will think of ϱ as a map from $[0,\infty)$ into the space of non-negative, finite Borel measures, which we denote by $\mathscr{M}_+(\mathbb{R}^d)$. The quantity ϱ is called the density and it represents the distribution of mass in time and space. The first equation in (1.1) (the continuity equation) expresses the local conservation of mass, where

$$\boldsymbol{v}(t,\cdot)\in\mathscr{L}^2\big(\mathbb{R}^d,\varrho(t,\cdot)\big) \quad \text{for all } t\in[0,\infty) \tag{1.2}$$

is the Eulerian velocity field taking values in $\mathbb{R}^d$. The second equation in (1.1) (the momentum equation) expresses the local conservation of momentum $\boldsymbol{m}:=\varrho\boldsymbol{v}$. The pressure p will be discussed below. Notice that $\boldsymbol{m}(t,\cdot)$ is a finite $\mathbb{R}^d$-valued Borel measure absolutely continuous with respect to $\varrho(t,\cdot)$ for all $t\in[0,\infty)$, because of (1.2). The quantity ε is the total energy of the fluid and $\varepsilon(t,\cdot)$ is again a measure in $\mathscr{M}_+(\mathbb{R}^d)$ for all times $t\in[0,\infty)$. It is reasonable to assume $\varepsilon(t,\cdot)$ to be absolutely continuous with respect to the density $\varrho(t,\cdot)$ (no energy in vacuum). The third (the energy) equation in (1.1) expresses the local conservation of energy.

Formally, the equations (1.1) imply that the total mass and energy are preserved over time. Therefore, if the fluid has finite mass and total energy initially, then this

---

Date: December 15, 2018.
2000 *Mathematics Subject Classification.* 35L65, 49J40, 82C40.
Key words and phrases. Compressible Gas Dynamics, Optimal Transport.

will be the case for all positive times. We will make this assumption in the following. Without loss of generality, we will also assume that the mass is equal to one, which implies that $\varrho(t,\cdot) \in \mathscr{P}(\mathbb{R}^d)$, the space of Borel probability measures.

To obtain a closed system (1.1) it is necessary to prescribe an equation of state, which relates the pressure p to the density ϱ and the total energy ε. It is provided by thermodynamics. The following three distinct situations are important:

1.1. **Pressureless gases.** The pressure p vanishes and so the total energy reduces to just the kinetic energy: $\varepsilon = \frac{1}{2}\varrho|\boldsymbol{v}|^2$. The equations (1.1) take the form

$$\left.\begin{aligned} \partial_t \varrho + \nabla\cdot(\varrho\boldsymbol{v}) = 0 \\ \partial_t(\varrho\boldsymbol{v}) + \nabla\cdot(\varrho\boldsymbol{v}\otimes\boldsymbol{v}) = 0 \end{aligned}\right\} \quad \text{in } [0,\infty)\times\mathbb{R}^d, \tag{1.3}$$

and the energy equation in (1.1) follows formally from the continuity and momentum equations. The system (1.3) has been proposed as a simple model describing the formation of galaxies in the early stage of the universe. Its one-dimensional version is a building block for semiconductor models. Since fluid elements do not interact with each other because there is no pressure, the density $\varrho(t,\cdot)$ may become singular with respect to the d-dimensional Lebesgue measure $\mathcal{L}^d$. For adhesion (or: sticky particle) dynamics this concentration effect is actually a desired feature; see [76]: If fluid elements meet at the same location, then they stick together to form larger compounds and so $\varrho(t,\cdot)$ can have singular parts (in particular, Dirac measures). Consequently (1.1) must be understood in the sense of distributions. While mass and momentum are conserved, kinetic energy may be destroyed since the collisions are inelastic. In particular, the energy equation in (1.1) will typically be an inequality only. We will call the assumption of adhesion dynamics an entropy condition.

There are now numerous articles studying the pressureless gas dynamics equations (1.3) in one space dimension and establishing global existence of solutions. Frequently, a sequence of approximate solutions is constructed by considering discrete particles, where the initial mass distribution is approximated by a finite sum of Dirac measures. The dynamics of these particles are described by a finite dimensional system of ordinary differential equations between collision times. Whenever multiple particles collide, the new velocity of the bigger particle is determined from the conservation of mass and momentum, and the choice of impact law. The general existence result is obtained by letting the number of discrete particles go to infinity. In order to pass to the limit, several approaches are feasible. We only mention two: One approach relies on the observation that the cumulative distribution function associated to the density ϱ satisfies a certain scalar conservation law (see [13]) so the theory of entropy solutions of scalar conservation laws can be applied. Another approach makes use of the well-known theory of first-order differential inclusions, applied to the cone of monotone transport maps from a reference measure space to $\mathbb{R}$; see [65]. We refer the reader to $[9,10,12,39,45,49,52,62,66,68,74]$ for more information.

For the multi-dimensional pressureless gas dynamics equations, global existence of solutions to (1.3) has been considered in [34]. The global existence proof in [71] for sticky particle solutions seems to be incomplete, as the authors in [14] show that for certain choices of initial data, sticky particle solutions cannot exist. This raises the question of the correct solution concept for the equations (1.3).

1.2. **Isentropic gases.** In this regime, the thermodynamical entropy of the fluid is assumed to be constant in space and time. Consequently, the pressure is a function of the density only. We introduce the internal energy

$$\mathcal{U}[\varrho] := \begin{cases} \displaystyle\int_{\mathbb{R}^d} U\big(r(x)\big)\,dx & \text{if } \varrho = r\mathcal{L}^d, \\ \infty & \text{otherwise,} \end{cases}$$

where $U(r) := \kappa r^\gamma$ for $r \geqslant 0$. The constant $\gamma > 1$ is called the adiabatic coefficient, and $\kappa > 0$ is another constant. The total energy is the sum of the kinetic energy introduced above and the internal energy. Since we are only interested in solutions of (1.1) with finite total energy, the density $\varrho(t,\cdot)$ must be absolutely continuous with respect to the Lebesgue measure for all $t \in [0,\infty)$. Let $r(t,\cdot)$ be its Radon-Nikodým derivative. Then $p(t,\cdot) = P\big(r(t,\cdot)\big)\,\mathcal{L}^d$ for all $t \in [0,\infty)$, where

$$P(r) = U'(r)r - U(r) \quad \text{for } r \geqslant 0.$$

This setup describes polytropic gases. Other choices of U are possible, for example $U(r) = \kappa r \log r$ for isothermal gases (then $P(r) = \kappa r$). We consider

$$\left.\begin{aligned} \partial_t\varrho + \nabla\cdot(\varrho\boldsymbol{v}) &= 0 \\ \partial_t(\varrho\boldsymbol{v}) + \nabla\cdot(\varrho\boldsymbol{v}\otimes\boldsymbol{v}) + \nabla P(\varrho) &= 0 \end{aligned}\right\} \quad \text{in } [0,\infty)\times\mathbb{R}^d \tag{1.4}$$

(with slight abuse of notation). As in the pressureless case, the energy equation in (1.1) follows formally from the continuity and the momentum equation.

It is well-known that a generic solution to the isentropic Euler equations will not remain smooth, even for regular initial data. Instead the solution will have jump discontinuities along codimension-one submanifolds in space-time, which are called shocks. Then the continuity and the momentum equation must be considered in the sense of distributions, and the energy equation does no longer follow automatically. A physically reasonable relaxation is to assume that no energy can be created by the fluid: The energy equality in (1.1) must be replaced by the inequality

$$\partial_t\Big(\tfrac{1}{2}\varrho|\boldsymbol{v}|^2 + U(\varrho)\Big) + \nabla\cdot\bigg(\Big(\tfrac{1}{2}\varrho|\boldsymbol{v}|^2 + U'(\varrho)\varrho\Big)\boldsymbol{v}\bigg) \leqslant 0 \tag{1.5}$$

in distributional sense. Physically, strict inequality in (1.5) means that mechanical energy is transformed into heat, a form of energy that is not accounted for by the model. Notice that a differential inequality like (1.5) contains some information on the regularity of solutions: The space-time divergence of a certain non-linear function of $(\varrho,\boldsymbol{v})$ is a non-positive distribution, and thus a measure. In the one-dimensional case, it is even reasonable to look for weak solutions of (1.4) that satisfy differential inequalities like (1.5) simultaneously for a large class of non-linear functions of $(\varrho,\boldsymbol{v})$ that are called entopy-entropy flux pairs. Such an assumption on the solutions is again an entropy condition. Using the method of compensated compactness, it is then possible the establish the global existence of weak (entropy) solutions of (1.4). We refer the reader to [21–23, 35–37, 56, 59, 60] for more information.

In several space dimensions the only available entropy-entropy flux pair is the total energy-energy flux. Using non-linear iteration schemes like the ones introduced by Nash [63, 64] to construct isometric imbeddings of Riemannian manifolds, one can establish the existence of a large class of initial data for which weak solutions of (1.4) exist globally in time. We refer the reader to the ground-breaking results by De Lellis and Székelyhidi [28, 29] and subsequent work [24–26, 40] by various authors. These results give, in fact, much more precise information: One can show that for suitable initial data there exist *infinitely many* weak solutions of (1.4), even if one requires that solutions satisfy an entropy condition in the form (1.5). This is related to the fact that there is—in addition to energy dissipation through shocks—an additional dissipation mechanism due to very high oscillations of the velocity field, which is reminiscent of anomalous dissipation in turbulence. Moreover, there is a precise threshold of Hölder regularity 1/3 between the energy conserving and the energy dissipating regimes. For incompressible flows, this has been conjectured based on physical considerations by Onsager [67]. A mathematical proof of this conjecture has been provided in a series of recent articles; see [16, 17, 53] and references therein. For related results for the compressible Euler equations see [41]. The Cauchy problem for (1.4) in several space dimensions, however, has not been solved yet: In order to apply the above methods for *any* given initial data, it currently seems necessary to allow a small increase in energy initially, which is in violation of (1.5).

1.3. **Full Euler equations.** We consider a polytropic gas with adiabatic coefficient $\gamma > 1$. Then the pressure is given in terms of $(\varrho, \boldsymbol{v}, \varepsilon)$ by the formula

$$p(t,\cdot) = (\gamma - 1)\left(\varepsilon - \frac{1}{2}\varrho|\boldsymbol{v}|^2\right)(t,\cdot) \quad \text{for all } t \in [0,\infty). \tag{1.6}$$

Density and pressure define the specific thermodynamical entropy, given as

$$S := \log\left(\frac{p}{c\varrho^\gamma}\right) \quad \text{with } c := \kappa(\gamma-1) > 0 \text{ and } \gamma > 1$$

in the case of polytropic gases. We assume that

$$S(t,\cdot) \in \mathscr{L}^1\big(\mathbb{R}^d, \varrho(t,\cdot)\big) \quad \text{for all } t \in [0,\infty),$$

so that the entropy density $\sigma = \varrho S$ is well-defined as a measure.

Definition 1.1 (Internal Energy)**.** Let $U(r,S) := \kappa e^S r^\gamma$ for all $r \geqslant 0$ and $S \in \mathbb{R}$, where $\kappa > 0$ and $\gamma > 1$ are constants. Then we define the internal energy

$$\mathcal{U}[\varrho,\sigma] := \begin{cases} \displaystyle\int_{\mathbb{R}^d} U\big(r(x), S(x)\big)\,dx & \text{if } \varrho = r\mathcal{L}^d \text{ and } \sigma = \varrho S, \\ \infty & \text{otherwise,} \end{cases}$$

for all pairs of measures $(\varrho,\sigma) \in \mathscr{P}(\mathbb{R}^d) \times \mathscr{M}_+(\mathbb{R}^d)$.

Since we are only interested in solutions with finite energy, the density must be absolutely continuous with respect to the Lebesgue measure, thus $\varrho(t,\cdot) = r(t,\cdot)\,\mathcal{L}^d$. In this case, we define $P(r,S) = U'(r,S)r - U(r,S)$ (here $'$ denotes differentiation with respect to r), and the pressure term in (1.13) takes the form

$$p(t,\cdot) = P\big(r(t,\cdot), S(t,\cdot)\big)\,\mathcal{L}^d \quad \text{for all } t \in [0,\infty). \tag{1.7}$$

Moreover, combining (1.6) and (1.7) with (1.1), we obtain that

$$\partial_t \sigma + \nabla\cdot(\sigma\boldsymbol{v}) = 0 \quad \text{in } [0,\infty)\times\mathbb{R}^d. \tag{1.8}$$

Equivalently, the specific entropy S must be constant along characteristics:

$$\partial_t S + \boldsymbol{v}\cdot\nabla S = 0 \quad \text{in } [0,\infty)\times\mathbb{R}^d. \tag{1.9}$$

Formally, system (1.1) is equivalent to the one where the energy equation is replaced by (1.8) (or even (1.9)). But since the solutions to the compressible Euler equations may become discontinuous in finite time, the physically reasonable relaxation is that the specific entropy should be non-decreasing forward in time, which expresses the second law of thermodynamics. It follows that

$$\inf_{x\in\mathbb{R}^d} S(t,x) \geqslant \inf_{x\in\mathbb{R}^d} \bar{S}(x) \quad \text{for all } t\in[0,\infty), \tag{1.10}$$

where $\bar{S}$ is the initial specific entropy. An Eulerian argument in support of (1.10), based on entropy inequalities, was given in [72]. We will assume that

$$\inf_{x\in\mathbb{R}^d} \bar{S}(x) \geqslant \alpha$$

for some $\alpha\in\mathbb{R}$. Since the shift of S by α can be absorbed into the constant $\kappa>0$, we may assume without loss of generality that $\alpha=0$, thus S is non-negative. As for the isentropic Euler equations, global existence and uniqueness of weak solutions to the full system (1.1) are open problems, even in one space dimension.

In this paper, we will consider a variational time discretization for the compressible gas dynamics equations that is motivated by minimizing movements for curves of maximal slope on metric spaces; see [4, 33, 54]. For any given initial data with finite mass and total energy, we prove that sequences of approximate solutions generated by this scheme converge to a measure-valued solution of (1.1).

Definition 1.2. We denote by $\mathscr{P}_2(\mathbb{R}^d)$ the space of Borel probability measures with finite second moment, endowed with the 2-Wasserstein distance; see Definition 2.1 below. For a map $t\mapsto\varrho_t\in\mathscr{P}_2(\mathbb{R}^d)$, $t\in[0,T]$, we denote by

$$\|\varrho\|_{\mathrm{Lip}([0,T];\mathscr{P}_2(\mathbb{R}^d))} := \sup_{\substack{t_1,t_2\in[0,T]\\ t_1\neq t_2}} \frac{\mathrm{W}_2(\varrho_{t_1},\varrho_{t_2})}{|t_2-t_1|}$$

the Lipschitz seminorm, with W_2 the Wasserstein distance; see (2.1).

Definition 1.3. We denote by $\mathscr{M}_{\mathrm{Ent}}(\mathbb{R}^d)$ the space of non-negative Borel measures with finite second moments and total variation equal to $\mathrm{Ent}\in[0,\infty)$, endowed with as suitably rescaled Wasserstein distance; see Definition 2.1.

We will assume that the total momentum vanishes initially, which implies that the total momentum vanishes for all $t>0$. This is not a restriction as the hyperbolic conservation law (1.1) is invariant under transformations to a moving reference frame in the absence of boundaries. The momentum map $t\mapsto\boldsymbol{m}_t=\varrho_t\boldsymbol{v}_t$ takes values in a convex set of $\mathbb{R}^d$-valued Borel measures whose total variations are uniformly bounded, as a consequence of a bound on the total energy. On this set, the narrow convergence of measures is metrized by the Monge-Kantorovich norm:

Definition 1.4. We denote by $\mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^N)$ the vector space of Lipschitz continuous maps $\zeta\colon\mathbb{R}^d\longrightarrow\mathbb{R}^N$. The Lipschitz constant of $\zeta\in\mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^N)$ is

$$\|\zeta\|_{\mathrm{Lip}(\mathbb{R}^d)} := \sup_{x_1\neq x_2}\frac{|\zeta(x_1)-\zeta(x_2)|}{|x_1-x_2|}.$$

We denote by $\mathrm{BL}(\mathbb{R}^d;\mathbb{R}^N)$ the subspace of bounded functions in $\mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^N)$. It is a Banach space when equipped with the bounded Lipschitz norm

$$\|\zeta\|_{\mathrm{BL}(\mathbb{R}^d)} := \max\Big\{\|\zeta\|_{\mathscr{L}^\infty(\mathbb{R}^d)}, \|\zeta\|_{\mathrm{Lip}(\mathbb{R}^d)}\Big\}.$$

Let $\mathrm{BL}_1(\mathbb{R}^d;\mathbb{R}^N)$ be the space of all $\zeta\in\mathrm{BL}(\mathbb{R}^d;\mathbb{R}^N)$ with $\|\zeta\|_{\mathrm{BL}(\mathbb{R}^d)}\leqslant 1$.

We denote by $\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d;\mathbb{R}^N)$ the space of $\mathbb{R}^N$-valued Borel measures $\boldsymbol{m}$ with zero mean and finite first moment, equipped with the Monge-Kantorovich norm

$$\|\boldsymbol{m}\|_{\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d)} := \sup\bigg\{\int_{\mathbb{R}^d}\zeta(x)\cdot\boldsymbol{m}(dx)\colon \zeta\in\mathrm{BL}_1(\mathbb{R}^d;\mathbb{R}^N)\bigg\}. \tag{1.11}$$

The Monge-Kantorovich norm is bounded above by the total variation.

We refer the reader to [18, 27, 38, 51] for additional information. Notice that the integral in (1.11) is well-defined because $\boldsymbol{m}$ has finite first moment, by assumption. By Cauchy-Schwarz inequality, this holds true whenever $\boldsymbol{m}=\varrho\boldsymbol{v}$ with

$$\varrho\in\mathscr{P}_2(\mathbb{R}^d) \quad\text{and}\quad \boldsymbol{v}\in\mathscr{L}^2(\mathbb{R}^d,\varrho).$$

Remark 1.5. We think of weak solutions of (1.1) as maps $t\mapsto(\varrho_t,\boldsymbol{m}_t,\sigma_t)$ taking values in a convex set of vector measures with uniformly bounded total variations, equipped with the Wasserstein distance/Monge-Kantorovich norm. Since the maps are Lipschitz continuous, they are strongly differentiable almost everywhere (a.e.) in time. In particular, the time derivative of the momentum exists as an element in the closure of the space of vector measures with respect to the Monge-Kantorovich norm, which is a proper subset of the dual space $\mathrm{BL}(\mathbb{R}^d;\mathbb{R}^d)^*$; see [8] for additional properties. The usage of the Monge-Kantorovich norm is thus very well adapted to the structure of the equations (1.1), with the time derivative of the momentum given as the divergence of a measure field taking values in the symmetric, positive semidefinite matrices. Testing against $\mathrm{BL}_1(\mathbb{R}^d;\mathbb{R}^d)$-functions, we can (mollify and) integrate by parts. Since the derivative of the test function is bounded in norm, we must control the size of the matrix field, for example with respect to the 1-Schatten norm (the sum over the absolute values of the singular values). For symmetric, positive semidefinite matrices, this is simply the trace of the matrix.

Definition 1.6. Let $E\subset\mathbb{R}$ be some subset and (X,d) a metric space. We denote by $\mathrm{BV}(E,X)$ the space of maps $f\colon I\longrightarrow E$ with finite variation

$$V_E(f) := \sup\sum_{i=1}^m d\big(f(t_{i-1}),f(t_i)\big),$$

where the sup is taken over all $t_0\leqslant t_1\leqslant\ldots\leqslant t_m$ contained in E.

We refer the reader to [43] for further information on metric space-valued functions of bounded variation. In particular, a version of Helly's compactness theorem for sequences of X-valued maps is proved there. A function $f\colon E\longrightarrow X$ has bounded variation if and only if it factors as $g\circ\phi$ where $\phi\colon E\longrightarrow\mathbb{R}$, defined as

$$\phi(t) := V_{(-\infty,t]\cap E}(f) \quad\text{for all } t\in E,$$

is its total variation and $g\colon\phi(E)\longrightarrow X$ is Lipschitz continuous. We will consider spaces $\mathrm{BV}(E,X)$ with $E=[0,T]$ and $X=\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d;\mathbb{R}^d)$.

Remark 1.7. Since we consider densities with finite second moment (thus momenta with finite first moment), it is possible to consider test functions with non-compact support. With V any Banach space, let $\mathscr{C}_*(\mathbb{R}^d;V)$ be the space of all continuous functions $\zeta\colon\mathbb{R}^d\longrightarrow V$ with the property that $\lim_{|x|\to\infty}\zeta(x)\in V$ exists. Then

$$\mathscr{C}_*(\mathbb{R}^d;V) = V+\mathscr{C}_0(\mathbb{R}^d;V),$$

where $\mathscr{C}_0(\mathbb{R}^d;V)$ is the closure of the space of compactly supported continuous V-valued maps with respect to the the sup-norm. We define

$$\mathfrak{A} := \big\{u \in \mathscr{C}^1(\mathbb{R}^d;\mathbb{R}^D)\colon \nabla u \in \mathscr{C}_*\big(\mathbb{R}^d;\mathrm{Mat}_{D\times d}(\mathbb{R})\big)\big\}, \tag{1.12}$$

with $\mathrm{Mat}_{D\times d}(\mathbb{R})$ the space of real $(D\times d)$-matrices. We will not explicitly indicate the dimension D as it will be clear from the context. Functions in $\mathfrak{A}$ grow at most linearly at infinity. Let $\mathscr{C}^1_c([0,\infty))\otimes\mathfrak{A}$ be the space of tensor products

$$\eta\otimes\zeta(t,x) := \eta(t)\zeta(x) \quad \text{with } \eta\in\mathscr{C}^1_c\big([0,\infty)\big) \text{ and } \zeta\in\mathfrak{A}.$$

We will assume (1.13) to hold in duality with this space, testing against all

$$\eta\otimes\zeta\in\mathscr{C}^1_c\big([0,\infty)\big)\otimes\mathfrak{A}.$$

Notice that all products of conserved quantities $(\varrho,\boldsymbol{m},\sigma)$ with $\eta\otimes\zeta$ are integrable since these measures have finite first moments. On the other hand, the derivative $\nabla_x\zeta$ is bounded and so the integrals involving the fluxes are well-defined as well. For all $T>0$, the tensor product $\mathscr{C}([0,T])\otimes V$ is dense in $\mathscr{C}([0,T];V)$ with respect to the sup-norm, with V any locally convex topoligical vector space.

We can now state our main existence result.

Theorem 1.8 (Global Existence)**.** *Suppose that initial data*

$$\bar\varrho\in\mathscr{P}_2(\mathbb{R}^d), \quad \bar{\boldsymbol{v}}\in\mathscr{L}^2(\mathbb{R}^d,\bar\varrho), \quad \bar S\in\mathscr{L}^\infty_+(\mathbb{R}^d,\bar\varrho)$$

is given with vanishing total momentum and finite internal energy:

$$\int_{\mathbb{R}^d}\bar{\boldsymbol{v}}(x)\,\bar\varrho(dx)=0, \quad \mathcal{U}[\bar\varrho,\bar\sigma]<\infty,$$

where $\bar\sigma:=\bar\varrho\bar S$. Let $\mathrm{Ent}:=\int_{\mathbb{R}^d}\bar\sigma(dx)$ be the initial total entropy.

For any $T>0$ there exist curves

$$\varrho\in\mathrm{Lip}\big([0,T];\mathscr{P}_2(\mathbb{R}^d)\big), \quad \sigma\in\mathrm{Lip}\big([0,T];\mathscr{M}_{\mathrm{Ent}}(\mathbb{R}^d)\big),$$
$$\boldsymbol{m}\in\mathrm{Lip}\big([0,T];\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d;\mathbb{R}^d)\big)$$

with the following properties:

(1) *The initial data is attained:*

$$\varrho(0,\cdot)=\bar\varrho, \quad \boldsymbol{m}(0,\cdot)=\bar\varrho\bar{\boldsymbol{v}}, \quad \sigma(0,\cdot)=\bar\sigma.$$

(2) *We have $\boldsymbol{m}=:\varrho\boldsymbol{v}$ and $\sigma=:\varrho S$ with*

$$\boldsymbol{v}(t,\cdot)\in\mathscr{L}^2\big(\mathbb{R}^d,\varrho(t,\cdot)\big), \quad S(t,\cdot)\in\mathscr{L}^\infty_+\big(\mathbb{R}^d,\varrho(t,\cdot)\big)$$

for all $t\in[0,T]$.

(3) *There exist two Young measures*

$$\nu^1,\nu^2\in\mathscr{L}^\infty_w\big([0,T];\mathscr{M}_+(\dot{\mathbb{R}}^d\times\mathfrak{X})\big),$$

where $\mathfrak{X}$ is a suitable compactification of the set

$$X:=[0,\infty)\times\mathbb{R}^d\times[0,S_{\max}], \quad S_{\max}:=\|S\|_{\mathscr{L}^\infty(\mathbb{R}^d,\varrho)},$$

of admissible $(\varrho,\boldsymbol{v},S)$, such that

$$\left.\begin{aligned} \partial_t \varrho + \nabla\cdot[\varrho\boldsymbol{v}] = 0 \\ \partial_t(\varrho\boldsymbol{v}) + \nabla\cdot[\varrho\boldsymbol{v}\otimes\boldsymbol{v}] + \nabla[\![P(\varrho,S)]\!] = 0 \\ \partial_t(\varrho S) + \nabla\cdot[\varrho\boldsymbol{v}S] = 0 \end{aligned}\right\} \quad \textit{in } \Big(\mathscr{C}_c^1\big([0,T)\big)\otimes\mathfrak{A}\Big)^*. \tag{1.13}$$

Here the brackets $[\cdot]$ and $[\![\cdot]\!]$ denote the integration of ν^1 and ν^2, respectively, against suitable functions of $(\varrho,\boldsymbol{v},S)$. We refer the reader to Section 6.5 for details.

Remark 1.9. The two Young measures from Theorem 1.8 play slightly different roles: One is used to describe the kinematic aspects of the flow (transport). It is generated by a sequence of approximate solutions that are interpolated piecewise linearly in time. The other one is used to describe the dynamical aspects (acceleration due to pressure gradient). It is constructed using piecewise constant interpolants. Since the approximations of the maps $t\mapsto(\varrho_t,\boldsymbol{m}_t,\sigma_t)$ are sufficiently regular in time, both Young measures generate the same conserved quantities. In order to have equality for the non-linear terms $\varrho\boldsymbol{v}\otimes\boldsymbol{v}$ and $P(\varrho,S)$, however, one needs to control the time regularity of the total energy. This will be considered elsewhere. It requires a more refined time interpolation, like De Giorgi's variational interpolation for minimizing movements; see Section 3.2 in [4]. Notice that while the derivative of $t\mapsto\boldsymbol{m}_t$ is uniquely determined a.e. in time in the closure of the space of measures with respect to the Monge-Kantorovich norm, the matrix measure field representing it is not.

We conclude this section by highlighting some aspects of our method.

Maximization of Entropy Production.

The recent results by De Lellis, Székelyhidi, and others suggest that non-uniqueness of weak solutions of compressible Euler equations in several space dimensions is a fact of life. The accepted entropy conditions, in the form (1.5) for the isentropic case, for example, are insufficient to select a unique solution. It is therefore natural to at least try to identify the "extreme" solutions among all possible weak solutions. Since the entropy condition (1.5) already implies that the total energy is non-increasing in time, it appears promising to strengthen this condition by requiring that total energy be dissipated *at maximal rate*, as suggested by Dafermos [30]. It was shown in [25], however, that this entropy condition seems to favor the highly oscillatory solutions of the isentropic Euler equations, which are non-unique.

Instead of decreasing the total energy, we will balance the dissipation of internal energy and minimizing the work done by the system (defined in terms of acceleration of the fluid elements), which amounts to changing the velocity a little as possible. We partition a given time interval into subintervals of length $\tau>0$. The updates in each timestep are obtained as the solutions of the above minimization problem. A similar approach has been studied for polyconvex elasticity in [31, 32].

Recall that by the first law of thermodynamics $T\,dS = dU - W$, where T denotes temperature, dS and dU are the infinitesimal changes in thermodynamical entropy and internal energy, and W is the work done *on* the system by its surroundings. Classically, the work is given by the formula $W = -p\,dV$, with p the pressure and dV the infinitesimal change of volume. Instead, we will utilize the *minimal work* functional, which we will introduce in the next paragraph. Our method boils down to minimizing the sum $W+U$, depending on some timestep $\tau>0$. Denoting by W_τ, U_τ the corresponding minimizers, we obtain the inequality

$$U_\tau + W_\tau \leqslant U_0 + 0,$$

where the index 0 refers to the fluid state obtained by not doing any work. Defining $dU := U_0 - U_\tau$, we observe that we are trying to *maximize* $dU - W$, which can formally be interpreted as maximizing $T\,dS$, thus maximizing the entropy production. Similar ideas have been explored in the recent paper [15].

Minimal Work Functional

Given $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$, we denote by $\mathscr{P}_\varrho(\mathbb{R}^{2d})$ the space of Borel probability measures $\boldsymbol{\mu} \in \mathscr{P}_2(\mathbb{R}^{2d})$ whose first marginal $\mathbb{p}^1\#\boldsymbol{\mu} = \varrho$. Here $\mathbb{p}^1(x,\xi) := x$ for all $(x,\xi) \in \mathbb{R}^{2d}$, and $\#$ is the push-forward. Any measure $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ describes the state of some fluid with density $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and velocity distribution μ_x for ϱ-a.e. $x \in \mathbb{R}^d$, where $\boldsymbol{\mu}(dx,d\xi) =: \mu_x(d\xi)\,\varrho(dx)$ denotes the disintegration of $\boldsymbol{\mu}$ with respect to ϱ. The special case $\boldsymbol{\mu}(dx,d\xi) = \delta_{\boldsymbol{u}(x)}(d\xi)\,\varrho(dx)$ represents a monokinetic state where all fluid elements located at the position $x \in \mathbb{R}^d$ have the same velocity $\boldsymbol{u}(x) \in \mathbb{R}^d$ and are therefore indistinguishable. The velocity field $\boldsymbol{u} \in \mathscr{L}^2(\mathbb{R}^d,\varrho)$, by construction. We will occasionally use bold letters to denote elements in $\mathbb{R}^{2d}$ such as

$$\boldsymbol{x} = (x,\xi), \quad \boldsymbol{y} = (y,\upsilon), \quad \text{and} \quad \boldsymbol{z} = (z,\zeta),$$

where $x,y,z \in \mathbb{R}^d$ represent positions and $\xi,\upsilon,\zeta \in \mathbb{R}^d$ velocities.

In order to measure the "distance" between two state measures $\boldsymbol{\mu}^1,\boldsymbol{\mu}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, we will use the minimal acceleration cost introduced in [46]. It is defined as follows: For a given timestep $\tau > 0$, consider a fluid element with initial position/velocity $\boldsymbol{x} \in \mathbb{R}^{2d}$. Assume that the fluid element transitions into a new state $\boldsymbol{z} \in \mathbb{R}^{2d}$. The transition is described by a smooth curve $X(\cdot|\boldsymbol{x},\boldsymbol{z})\colon [0,\tau] \longrightarrow \mathbb{R}^d$ such that

$$(X,\dot{X})(0) = (x,\xi) \quad \text{and} \quad (X,\dot{X})(\tau) = (z,\zeta)$$

(with $X := X(\cdot|\boldsymbol{x},\boldsymbol{z})$). Among all such curves there are the ones that minimize the acceleration $\int_0^\tau |\ddot{X}(t)|^2\,dt$. They are uniquely determined and given by

$$X(t|\boldsymbol{x},\boldsymbol{z}) = x + t\xi + \Big(3(z-x) - \tau(\zeta+2\xi)\Big)\frac{t^2}{\tau^2} - \Big(2(z-x) - \tau(\zeta+\xi)\Big)\frac{t^3}{\tau^3}$$

for all $t \in [0,\tau]$. The minimal acceleration can be computed explicitly, which allows us to define a cost measuring the "distance" between the two end states:

$$a_\tau(\boldsymbol{x},\boldsymbol{z})^2 := 3\left|\frac{z-x}{\tau} - \frac{\zeta+\xi}{2}\right|^2 + \frac{1}{4}|\zeta-\xi|^2 \tag{1.14}$$

for all $\boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^{2d}$. Note that $a_\tau(\boldsymbol{x},\boldsymbol{z}) = 0$ if and only if $z = x + \tau\xi$ and $\zeta = \xi$.

The cost function (1.14) can be rewritten in the following form:

$$a_\tau(\boldsymbol{x},\boldsymbol{z})^2 = \frac{3}{4\tau^2}|(x+\tau\xi) - z|^2 + \left|\zeta - \left(\xi - \frac{3}{2\tau}\Big((x+\tau\xi) - z\Big)\right)\right|^2 \tag{1.15}$$

for every $\boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^{2d}$. The first term measures how much the final position z differs from $x + \tau\xi$, which would be the position of the fluid element after a *free transport*. The second term measures the difference between ζ and the velocity

$$V_\tau(\boldsymbol{x},z) := \xi - \frac{3}{2\tau}\Big((x+\tau\xi) - z\Big), \tag{1.16}$$

which is the velocity that minimizes the acceleration among all curves that connect the initial position/velocity $\boldsymbol{x} \in \mathbb{R}^{2d}$ to the final position $z \in \mathbb{R}^d$. Notice that by minimizing the final velocity ζ for fixed $(\boldsymbol{x},z)$, setting ζ equal to (1.16), we closely link velocity and transport. The minimal work functional is then defined as an optimal transport problem with cost function (1.14) and for pairs of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathscr{P}_2(\mathbb{R}^{2d})$, in analogy to the Wasserstein distance; see Definition 2.1.

Non-Coercive Energy Functional and Monotone Maps.

Our variational method amounts to minimizing the sum of minimal work functional plus internal energy of the *final* state (the one after transport). We will be particularly interested in transport *maps*, where a fluid element at location $x \in \mathbb{R}^d$ is transported wholly to a new position $\boldsymbol{t}(x) \in \mathbb{R}^d$, without being split up. More precisely, we consider $\mathbb{R}^d$-valued transport maps $\boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$. Taking into account Definition 1.1 and assuming that the specific thermodynamical entropy S is simply transported along with the flow (recall (1.9)), we can formally write

$$\mathcal{U}[\boldsymbol{t}\#\varrho, \boldsymbol{t}\#\sigma] = \int_{\mathbb{R}^d} U\big(\varrho(x), S(x)\big) \,\det\big(\nabla\boldsymbol{t}(x)\big)^{1-\gamma}\,dx \tag{1.17}$$

for the internal energy of the final state, using the change of variables formula. Here we have identified ϱ with its Lebesgue density. Moreover, we have assumed that the transport map $\boldsymbol{t}$ is sufficiently regular and invertible. Our minimization scheme will ensure that the internal energy of the new fluid state is finite. But since the map $\boldsymbol{t} \mapsto \mathcal{U}[\boldsymbol{t}\#\varrho, \boldsymbol{t}\#\sigma]$ is not coercive, it does not suggest any natural function space setting to formulate the minimization problem. In order to be able to use the direct method of the calculus of variations, we need compactness of sublevel sets of the internal energy functional in a suitable topology, and its lower semicontinuity with respect to this topology. To achieve this, we make two choices:

- we assume that the transport maps are *monotone*, and
- in (1.17) we replace the gradient $\nabla\boldsymbol{t}(x)$ by its *symmetric part*;

see Proposition 5.4. A monotone map is locally of bounded variation. In particular, its variation (the total variation of its derivative) over any convex set can be controled in terms of its oscillation (the size of its range). This provides us with the necessary compactness of sublevel sets. On the other hand, by using only the symmetric part of the derivative $\nabla\boldsymbol{t}(x)$, the resulting energy functional becomes convex and lower semicontinuous with respect to weak* convergenve in the space of functions with bounded variation. We refer the reader to Section 5 for details.

The Geometry of Monotone Maps.

In continuum mechanics, a configuration is a function that assigns to each point of the body manifold (called the reference configuration) its position in physical space $\mathbb{R}^d$, at any given time. These maps are required to be injective because matter must not interpenetrate. The space of configurations therefore cannot be a vector space (subtracting a configuration from itself, we obtain the zero map, which is not injective). Our assumption of monotonicity of the transport maps is consistent with these considerations, but a bit stronger than mere injectivity. Note, however, that we require monotonicity of the transport maps in the limit of small timesteps $\tau > 0$ only, so that the minimizing $\boldsymbol{t}$ will be a perturbation of the identity map, which is monotone. Recall also that in the theory of generalized gradient flows on the space of probability measures, which utilizes the Wasserstein distance to determine the local geometry of the problem, the optimal transport maps are *cyclically monotone* (gradients of convex functions), which is a stronger condition than monotonicity. We do not wish to work with cyclically monotone maps since the induced velocity fields (obtained as limits of difference quotients between optimal transport maps and the identity) inherit the gradient property. For the compressible Euler equations, this would result in the regime of potential flows; see also [73].

Our variational methods is phrased as a convex minimization problem over the closed convex cone of monotone transport maps. As is usual in optimization, such a constraint may result in the appearance of Lagrange multipliers in the optimality conditions. For the monotonicity constraint under consideration here, it turns out that the representation of such Lagrange multipliers fits very neatly into the overall structure of our problem. In fact, elements in the closed convex cone that is polar to the cone of monotone maps can be represented as divergences of measure fields taking values in the positive semidefinite symmetric matrices; see Section 4.3. This is precisely the form of the flux terms in the gas dynamics equations (1.1).

In continuum mechanics, admissible velocities are elements of the tangent cone to the manifold of configurations. Consequently, if we consider the map $t \mapsto \varrho_t$ as a curve on the manifold of probability measures, then the corresponding velocity $\boldsymbol{v}$ should represent a curve in the tangent bundle. In our variational time discretization we update the velocity as follows: we first move the current velocity using the optimal transport map, then we project onto a suitably defined tangent cone to the cone of monotone maps at the new configuration. This projection will turn out to be trivial in the cases with pressure; in the pressureless case, the projection of velocity will be related to the sticky particle condition. This two-step update for the velocity is similar to the construction of the parallel transport of tangent vector fields along the space of probability measures, as developed in [3, 48].

Measure-Valued Solutions.

The Young measures of Theorem 1.8 are obtained as weak* limits

$$\nu \in \mathscr{L}^\infty_w\big([0,\infty); \mathscr{M}_+(\dot{\mathbb{R}}^d \times \mathfrak{X})\big)$$

of sequences of analogous maps constructed from approximate solutions $(\varrho_n, \boldsymbol{v}_n, S_n)$ of (1.1). The (t,x)-marginal of such ν is the weak* limit $t \mapsto \mu_t(dx)$ of

$$\mu_{n,t} := \varrho_{n,t} + \Big(\tfrac{1}{2}\varrho_{n,t}|\boldsymbol{v}_{n,t}|^2 + U(\varrho_{n,t}, S_{n,t})\Big)$$

(see (6.19)/(6.21)), which captures the space-time distribution of mass and total energy. The Young measure ν captures both oscillations and concentration in the approximating sequence. Notice that the concept of measure-valued solutions to hyperbolic balance laws is fairly weak. On the other hand, in view of the non-uniqueness results by De Lellis and Székelyhidi one may wonder whether a distinguished weak solution of (1.1) can be identified at all and what sets it apart from the other solutions. It has therefore been suggested by some researchers that the solution concept for (1.1) must be reconsidered, for example in favor of measure-valued or statistical solutions; see [42, 57, 58]. It would be interesting to investigate whether the non-linear iteration techniques introduces by De Lellis, Székelyhidi, and others can be used to promote measure-valued solutions to weak ones, at least in regions where the flow is expected to be laminar instead of turbulent/non-unique.

Our variational time discretization decreases the total energy, while preserving the entropy. This may seem backwards from the physical point of view. We would like to point out, however, that in turbulence it is standard to assume that solutions of the incompressible Navier-Stokes equations converge (in the high Reynolds number limit) to velocity fields that dissipate kinetic energy, even though they formally solve the incompressible Euler equations. Therefore the incompressible Euler equations seem to only give an incomplete description of the actual physical phenomena. It is natural to expect that similar effects occur in the compressible models.

2. Notation

In the following, we will always assume that $\mathbb{R}^D$ is equipped with the Euclidean inner product, for which we write $x \cdot y$ or $\langle x, y\rangle$ with $x, y \in \mathbb{R}^D$.

Let $\mathrm{Mat}_d(\mathbb{R})$ be the space of real $(d \times d)$-matrices and

$$\mathrm{Mat}_d(\mathbb{R}, \Box) := \Big\{A \in \mathrm{Mat}_d(\mathbb{R})\colon v \cdot (Av) \,\Box\, 0 \text{ for all } v \in \mathbb{R}^d\Big\}$$

where $\Box$ stands for either $\geqslant$ or $>$. We will refer to the elements of $\mathrm{Mat}_d(\mathbb{R}, \geqslant)$ (resp. $\mathrm{Mat}_d(\mathbb{R}, >)$) as positive semi-definite (resp. positive definite) matrices. Notice that these matrices are not assumed to be symmetric. The analogous spaces of symmetric matrices will be denoted by $\mathrm{Sym}_d(\mathbb{R})$ and $\mathrm{Sym}_d(\mathbb{R}, \Box)$. We have $A \in \mathrm{Mat}_d(\mathbb{R}, \Box)$ if and only if $A^{\mathrm{sym}} \in \mathrm{Sym}_d(\mathbb{R}, \Box)$ where $A^{\mathrm{sym}} := (A + A^{\mathrm{T}})/2$ is the symmetric part of A. The antisymmetric part of A is defined as $A^{\mathrm{anti}} := (A - A^{\mathrm{T}})/2$ and we will denote by $\mathrm{Skew}_d(\mathbb{R})$ the space of antisymmetric real $(d \times d)$-matrices. Recall that the Frobenius inner product of matrices is defined as

$$A : B := \mathrm{tr}(A^{\mathrm{T}} B) \quad \text{for all } A, B \in \mathrm{Mat}_d(\mathbb{R}).$$

The norms on these spaces will be the ones induced by the inner products.

We denote by $\mathscr{C}_{\mathrm{b}}(\mathbb{R}^D)$ the space of bounded continuous functions on $\mathbb{R}^D$ and by $\mathscr{P}(\mathbb{R}^D)$ the space of Borel probability measures. Weak convergence of sequences of probability measures is defined by testing against functions in $\mathscr{C}_{\mathrm{b}}(\mathbb{R}^D)$. For any $1 \leqslant p < \infty$ we denote by $\mathscr{P}_p(\mathbb{R}^D)$ the space of Borel probability measures with finite pth moment, so that $\int_{\mathbb{R}^D} |x|^p \,\varrho(dx) < \infty$ for every $\varrho \in \mathscr{P}_p(\mathbb{R}^D)$.

Definition 2.1 (p-Wasserstein Distance)**.** For any $\varrho^1, \varrho^2 \in \mathscr{P}(\mathbb{R}^D)$ let

$$\mathrm{Adm}(\varrho^1, \varrho^2) := \Big\{\boldsymbol{\gamma} \in \mathscr{P}(\mathbb{R}^{2D})\colon \mathbb{p}^k \# \boldsymbol{\gamma} = \varrho^k \text{ with } k = 1..2\Big\}$$

be the space of admissible transport plans connecting ϱ^1 and ϱ^2, where

$$\mathbb{p}^k(x^1, x^2) := x^k \quad \text{for all } (x^1, x^2) \in \mathbb{R}^{2D} = (\mathbb{R}^D)^2$$

and $k = 1..2$, and $\#$ denotes the push-forward of measures. For any $1 \leqslant p < \infty$ the p-Wasserstein distance $\mathrm{W}_p(\varrho^1, \varrho^2)$ between ϱ^1, ϱ^2 is defined by

$$\mathrm{W}_p(\varrho^1, \varrho^2)^p := \inf_{\boldsymbol{\gamma} \in \mathrm{Adm}(\varrho^1, \varrho^2)} \int_{\mathbb{R}^{2D}} |x^1 - x^2|^p \,\boldsymbol{\gamma}(dx^1, dx^2). \tag{2.1}$$

Remark 2.2. The inf in (2.1) is actually attained, so the set $\mathrm{Opt}(\varrho^1, \varrho^2)$ of transport plans $\boldsymbol{\gamma}$ that minimize (2.1) (called optimal transport plans) is non-empty. For $p = 2$ the support of each $\boldsymbol{\gamma} \in \mathrm{Opt}(\varrho^1, \varrho^2)$ is contained in the subdifferential of a lower semicontinuous, convex map (therefore it is cyclically monotone). If ϱ^1 is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^D$, then each optimal transport plan is induced by a map (its support lies on the graph of a function):

$$\boldsymbol{\gamma} = (\mathrm{id}, \boldsymbol{t}) \# \varrho^1 \quad \text{for suitable } \boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^D, \varrho^1).$$

We refer the reader to [4] for further details.

For any $n \in \mathbb{N}$ and $k = 1 \dots n$, we define projections

$$\mathbb{p}^k(x^1 \dots x^n) := x^k \quad \text{for all } (x^1 \dots x^n) \in \mathbb{R}^{nd} = (\mathbb{R}^d)^n.$$

We will also use projections $\mathbb{x}$ and $\mathbb{y}^k$ defined by

$$\mathbb{x}(x, y^1 \dots y^n) := x, \quad \mathbb{y}^k(x, y^1 \dots y^n) := y^k$$

for all $(x, y^1 \dots y^n) \in \mathbb{R}^{(n+1)d} = (\mathbb{R}^d)^{n+1}$ and $k = 1 \dots n$, with $n \in \mathbb{N}$. Sometimes it will be convenient to write $\mathbb{z}^k$ or $\mathbb{v}^k$ in place of $\mathbb{y}^k$ (same definition), depending on whether the symbols represent positions or velocities, which will be clear from the context. For $n = 1$ we will usually write $\mathbb{y} := \mathbb{y}^1$ etc.

Definition 2.3 (Distance)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ be given and

$$\mathscr{P}_\varrho(\mathbb{R}^{2d}) := \Big\{ \boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d}) \colon \mathbb{x}\#\boldsymbol{\gamma} = \varrho \Big\}.$$

We introduce a distance as follows: for any $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ we define

$$\mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)^2 := \int_{\mathbb{R}^d} \mathrm{W}(\gamma^1_x, \gamma^2_x)^2 \, \varrho(dx),$$

where $\boldsymbol{\gamma}^k(dx, dy) =: \gamma^k_x(dy)\, \varrho(dx)$ with $k = 1..2$ denotes the disintegration of $\boldsymbol{\gamma}^k$, and where W is the Wasserstein distance on $\mathscr{P}_2(\mathbb{R}^d)$; see [4, 48].

Definition 2.4 (Transport Plans)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ be given.

(i.) *Admissible Plans.* For any $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ we define

$$\mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) := \Big\{ \boldsymbol{\alpha} \in \mathscr{P}_2(\mathbb{R}^{3d}) \colon (\mathbb{x}, \mathbb{y}^k)\#\boldsymbol{\alpha} = \boldsymbol{\gamma}^k \text{ with } k = 1..2 \Big\}.$$

(ii.) *Optimal Plans.* For any $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ we define

$$\mathrm{Opt}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) := \Big\{ \boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \colon \\ \mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)^2 = \int_{\mathbb{R}^{3d}} |y^1 - y^2|^2 \, \boldsymbol{\alpha}(dx, dy^1, dy^2) \Big\}.$$

Theorem 2.5. *Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ be given.*

(i.) *The function W_ϱ is a distance on $\mathscr{P}_\varrho(\mathbb{R}^{2d})$ and lower semicontinuous with respect to weak convergence in $\mathscr{P}_2(\mathbb{R}^{2d})$. We have*

$$\mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)^2 = \min_{\boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)} \int_{\mathbb{R}^{3d}} |y^1 - y^2|^2 \, \boldsymbol{\alpha}(dx, dy^1, dy^2)$$

for all $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, and thus $\mathrm{Opt}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)$ is non-empty.

(ii.) *The set $(\mathscr{P}_\varrho(\mathbb{R}^{2d}), \mathrm{W}_\varrho)$ is a complete metric space.*

Proof. We refer the reader to Section 4.1 in [48]. □

Definition 2.6 (Barycentric Projection)**.** For any $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ the barycentric projection $\mathrm{b}(\boldsymbol{\gamma}) \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ is defined as

$$\mathrm{b}(\boldsymbol{\gamma})(x) := \int_{\mathbb{R}^d} y \, \gamma_x(dy) \quad \text{for } \varrho\text{-a.e. } x \in \mathbb{R}^d,$$

where $\boldsymbol{\gamma}(dx, dy) =: \gamma_x(dy)\, \varrho(dx)$ is the disintegration of $\boldsymbol{\gamma}$.

An important subset of $\mathscr{P}_\varrho(\mathbb{R}^{2d})$ consists of those measures $\boldsymbol{\gamma}$ that are induced by maps: there exists a $\boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ taking values in $\mathbb{R}^d$ such that

$$\boldsymbol{\gamma}(dx, dy) = \delta_{\boldsymbol{t}(x)}(dy)\, \varrho(dx).$$

In this case, the distance W_ϱ reduces to the $\mathscr{L}^2(\mathbb{R}^d, \varrho)$-distance of the corresponding maps. If $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $\boldsymbol{\gamma}^1 = (\mathrm{id}, \boldsymbol{t})\#\varrho$ with $\boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$, then

$$\mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)^2 = \int_{\mathbb{R}^{2d}} |\boldsymbol{t}(x) - y^2|^2\, \boldsymbol{\gamma}^2(dx, dy^2); \tag{2.2}$$

see Lemma 5.3.2 in [4]. If $\mathrm{W}_\varrho(\boldsymbol{\gamma}^n, \boldsymbol{\gamma}) \longrightarrow 0$ as $n \to \infty$, with $\boldsymbol{\gamma}^n, \boldsymbol{\gamma}^\infty \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $\boldsymbol{\gamma}^n = (\mathrm{id}, \boldsymbol{t}^n)\#\varrho$ for some $\boldsymbol{t}^n \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$, then $\boldsymbol{t}^n \longrightarrow \boldsymbol{t}$ strongly in $\mathscr{L}^2(\mathbb{R}^d, \varrho)$ and $\boldsymbol{\gamma} = (\mathrm{id}, \boldsymbol{t})\#\varrho$. Indeed, our assumption implies that the sequence $\{\boldsymbol{t}^n\}_n$ is Cauchy in $\mathscr{L}^2(\mathbb{R}^d, \varrho)$ and hence converges to a limit $\boldsymbol{t}^\infty$, by completeness. On the other hand, since $(\mathrm{id}, \boldsymbol{t}^n, \boldsymbol{t}^\infty)\#\varrho \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^n, \boldsymbol{\gamma}^\infty)$ with $\boldsymbol{\gamma}^\infty := (\mathrm{id}, \boldsymbol{t}^\infty)\#\varrho$, we have

$$\mathrm{W}_\varrho(\boldsymbol{\gamma}^n, \boldsymbol{\gamma}^\infty) \leqslant \|\boldsymbol{t}^n - \boldsymbol{t}^\infty\|_{\mathscr{L}^2(\mathbb{R}^d, \varrho)} \longrightarrow 0 \quad \text{as } n \to \infty.$$

Then we use that $\mathrm{W}_\varrho(\boldsymbol{\gamma}, \boldsymbol{\gamma}^\infty) \leqslant \mathrm{W}_\varrho(\boldsymbol{\gamma}^n, \boldsymbol{\gamma}) + \mathrm{W}_\varrho(\boldsymbol{\gamma}^n, \boldsymbol{\gamma}^\infty)$. We have the estimate

$$\|\mathrm{b}(\boldsymbol{\gamma}^1) - \mathrm{b}(\boldsymbol{\gamma}^2)\|_{\mathscr{L}^2(\mathbb{R}^d, \varrho)} \leqslant \mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2),$$

as follows easily from Theorem 2.5 (i.) and Jensen inequality.

Minimal Work. As outlined in the Introduction, our approach relies on a functional measuring the work done to the fluid, called minimal work functional.

Definition 2.7 (Minimal Work)**.** For any pair of measures $\boldsymbol{\mu}^1, \boldsymbol{\mu}^2 \in \mathscr{P}_2(\mathbb{R}^{2d})$ we denote by $\mathrm{Adm}(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2)$ the set of transport plans $\boldsymbol{\omega} \in \mathscr{P}(\mathbb{R}^{4d})$ with

$$(\mathbb{p}^1, \mathbb{p}^2)\#\boldsymbol{\omega} = \boldsymbol{\mu}^1 \quad \text{and} \quad (\mathbb{p}^3, \mathbb{p}^4)\#\boldsymbol{\omega} = \boldsymbol{\mu}^2.$$

The minimal work is the functional A_τ defined by

$$\mathrm{A}_\tau(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2)^2 := \inf \left\{ \int_{\mathbb{R}^{4d}} a_\tau(\boldsymbol{x}^1, \boldsymbol{x}^2)^2\, \boldsymbol{\omega}(d\boldsymbol{x}^1, d\boldsymbol{x}^2) \colon \boldsymbol{\omega} \in \mathrm{Adm}(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2) \right\}. \tag{2.3}$$

Note that A_τ is not a distance: It is not symmetric in its arguments $\boldsymbol{\mu}^1$ and $\boldsymbol{\mu}^2$, which follows from the asymmetry of the cost function (1.14). Moreover, it does not vanish if $\boldsymbol{\mu}^1 = \boldsymbol{\mu}^2$. Instead, we have the following relation:

$$\mathrm{A}_\tau(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2) = 0 \quad \Longleftrightarrow \quad \boldsymbol{\mu}^2 = F_\tau\#\boldsymbol{\mu}^1,$$

where $F_\tau \colon \mathbb{R}^{2d} \longrightarrow \mathbb{R}^{2d}$ is the *free transport map* defined by

$$F_\tau(\boldsymbol{x}) := (x + \tau\xi, \xi) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^{2d}.$$

The minimal work functional measures how much each fluid element deviates from the straight path determined by its initial velocity; see [46] for more details.

When minimizing the integral in (2.3) over all plans $\boldsymbol{\omega} \in \mathscr{P}_2(\mathbb{R}^{4d})$ with

$$(\mathbb{p}^1, \mathbb{p}^2, \mathbb{p}^3)\#\boldsymbol{\omega} =: \boldsymbol{\beta} \quad \text{for given } \boldsymbol{\beta} \in \mathscr{P}_2(\mathbb{R}^{3d}),$$

then there exists a unique such minimizer, which takes the form $\boldsymbol{\omega} = H_\tau\#\boldsymbol{\beta}$, with the map $H_\tau \colon \mathbb{R}^{3d} \longrightarrow \mathbb{R}^{4d}$ defined for all $\boldsymbol{x} \in \mathbb{R}^{2d}$ and $z \in \mathbb{R}^d$ as

$$H_\tau(\boldsymbol{x}, z) := \Big(\boldsymbol{x}, z, V_\tau(\boldsymbol{x}, z)\Big).$$

This determines the final velocity in terms of the data $\boldsymbol{x}$ and the new position z.

3. Energy Minimization: First Properties

In preparation of our variational time discretization for (1.1), we first consider the metric projection onto the cone of monotone transport plans.

3.1. **Monotone Transport Plans.** To every subset $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ we can associate a set-valued map $u_\Gamma \colon \mathbb{R}^d \longrightarrow P(\mathbb{R}^d)$ (where $P(\mathbb{R}^d)$ is the power set of $\mathbb{R}^d$) by

$$u_\Gamma(x) := \Big\{ y \in \mathbb{R}^d \colon (x,y) \in \Gamma \Big\} \quad \text{for all } x \in \mathbb{R}^d.$$

For any set-valued map $u \colon \mathbb{R}^d \longrightarrow P(\mathbb{R}^d)$, we denote by

$$\operatorname{dom}(u) := \Big\{ x \in \mathbb{R}^d \colon u(x) \neq \varnothing \Big\},$$
$$\operatorname{graph}(u) := \Big\{ (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \colon y \in u(x) \Big\}$$

its domain and graph. A subset $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is called monotone if

$$\langle x_1 - x_2, y_1 - y_2 \rangle \geqslant 0 \quad \text{for any pair of } (x_i, y_i) \in \Gamma.$$

Such a set is called maximal monotone if for any monotone set $\Gamma' \subset \mathbb{R}^d \times \mathbb{R}^d$ with $\Gamma \subset \Gamma'$ we have that $\Gamma = \Gamma'$. Equivalently, if it is not possible to enlarge Γ without destroying the monotonicity. We will call any set-valued map u as above (maximal) monotone if the set $\operatorname{graph}(u)$ is (maximal) monotone.

By Zorn's lemma, any monotone set (equivalently, any monotone set-valued map) can be extended to a maximal monotone set (map). Typically, this extension is not unique. A maximal monotone extension can be obtained constructively as follows: Let $\Gamma \in \mathbb{R}^d \times \mathbb{R}^d$ be monotone. Then (for all $(x,y), (x^*,y^*) \in \mathbb{R}^d \times \mathbb{R}^d$)

(1) define the Fitzpatrick function

$$F_\Gamma(x,y) := \sup \Big\{ \langle y', x \rangle + \langle y, x' \rangle - \langle y', x' \rangle \colon (x', y') \in \Gamma \Big\};$$

(2) compute its Fenchel conjugate

$$F_\Gamma^*(y^*,x^*) := \sup \Big\{ \langle y^*, x \rangle + \langle y, x^* \rangle - F_\Gamma(x,y) \colon (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \Big\};$$

(3) compute the proximal average

$$N_\Gamma(x,y) := \inf \bigg\{ \tfrac{1}{2} F_\Gamma(x_1,y_1) + \tfrac{1}{2} F_\Gamma^*(y_2,x_2) + \tfrac{1}{8}\|x_1 - x_2\|^2 + \tfrac{1}{8}\|y_1 - y_2\|^2 \colon$$
$$(x,y) = \tfrac{1}{2}(x_1,y_1) + \tfrac{1}{2}(x_2,y_2) \bigg\}.$$

The function N_Γ is lower semicontinuous, convex, and proper, and the set

$$\bar{\Gamma} := \Big\{ (x,y) \colon N_\Gamma(x,y) = \langle y, x \rangle \Big\} \tag{3.1}$$

is a maximal monotone extension of Γ. We refer the reader to [6, 47] for details.

Remark 3.1. For any maximal monotone set-valued function $u \colon \mathbb{R}^d \longrightarrow P(\mathbb{R}^d)$ the image $u(x)$ of any $x \in \mathbb{R}^d$ is closed and convex (possibly empty); see Proposition 1.2 of [1]. Therefore the dimension $\dim u(x)$ is well-defined. The singular sets

$$\Sigma^k(u) := \Big\{ x \in \mathbb{R}^d \colon \dim u(x) \geqslant k \Big\}, \quad \text{with } k = 1 \dots d,$$

are countably $\mathcal{H}^{d-k}$-rectifiable; see Theorem 2.2 of [1] for details. Here $\mathcal{H}^n$ denotes the n-dimensional Hausdorff measure. In particular, the set of points $x \in \operatorname{dom}(u)$ for which $u(x)$ contains more than one point (that is, the set $\Sigma^1(u)$) is negligible with respect to the Lebesgue measure $\mathcal{L}^d$. Outside $\Sigma^1(u)$ the function u is continuous. This observation will allow us to think of a maximal monotone map u as a Lebesgue measurable, *single-valued* function (just redefine u on the null set $\Sigma^1(u)$).

Definition 3.2 (Monotone Transport Plans)**.** For any $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$, we define

$$C_\varrho := \Big\{ \boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d}) \colon \operatorname{spt} \boldsymbol{\gamma} \text{ is a monotone subset of } \mathbb{R}^d \times \mathbb{R}^d \Big\}. \tag{3.2}$$

Our definition of monotonicity for measures in $\mathscr{P}_\varrho(\mathbb{R}^{2d})$ is motivated by the optimal transport plans of Definition 2.1: an *optimal* transport plan $\boldsymbol{\gamma}$ is characterized by the property that $\operatorname{spt} \boldsymbol{\gamma}$ must be a *cyclically monotone* set; see Section 6.2.3 in [4]. Then there exists a lower semicontinuous, convex, proper function φ with

$$\varphi(x) + \varphi^*(y) = \langle y, x \rangle \quad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d.$$

Here φ^* denotes the Fenchel conjugate to φ. In our setting, the cyclical monotonicity is replaced by monotonicity, and $N_\Gamma(x,y)$ of (3.1) plays the role of $\varphi(x) + \varphi^*(y)$. In the terminology of [47], the function N_Γ is called a self-dual Lagrangian. The cone C_ϱ contains the set of optimal transport plans defined above.

Since we do not make any assumptions on ϱ, its support may be a proper subset of $\mathbb{R}^d$ and have "holes." Fortunately, the monotonicity constraint enables us to work with objects that are defined on a fixed convex open subset of $\mathbb{R}^d$:

Definition 3.3 (Associated Maps)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ be given. For every $\boldsymbol{\gamma} \in C_\varrho$ we call u a maximal monotone map associated to $\boldsymbol{\gamma}$ if u is the maximal monotone set-valued map induced by a maximal monotone extension of $\Gamma := \operatorname{spt} \boldsymbol{\gamma}$.

Lemma 3.4. *For any $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\gamma} \in C_\varrho$, the domain of a maximal monotone map u associated to $\boldsymbol{\gamma}$ contains the convex open set $\Omega := \operatorname{int} \overline{\operatorname{conv}} \operatorname{spt} \varrho$, where* int *denotes the interior of a set,* conv *the convex hull, and* $\overline{\operatorname{conv}}$ *its closure.*

Proof. Let $\boldsymbol{\gamma} \in C_\varrho$ be given and consider any maximal monotone map u associated to $\boldsymbol{\gamma}$. Then graph(u) is a maximal monotone extension of $\Gamma := \operatorname{spt} \boldsymbol{\gamma}$, which implies that the projection $X := \mathbb{p}^1(\Gamma)$ of Γ onto $\mathbb{R}^d$ is contained in dom(u). Since

$$\operatorname{int} \overline{\operatorname{conv}} \operatorname{dom}(u) \subset \operatorname{dom}(u) \subset \overline{\operatorname{conv}} \operatorname{dom}(u)$$

(this is true for every maximal monotone set-valued function; see Corollary 1.3 of [1]) we conclude that the convex open set $\operatorname{int} \overline{\operatorname{conv}}(X) \subset \operatorname{dom}(u)$. It therefore suffices to show that $\operatorname{int} \overline{\operatorname{conv}}(X) = \Omega$. Note that Ω is independent of $\boldsymbol{\gamma}$ and u.

To prove the claim, choose any $x \in X$ and $r > 0$. Then we can estimate

$$\varrho\big(B_r(x)\big) = \boldsymbol{\gamma}\big(B_r(x) \times \mathbb{R}^d\big) \geqslant \boldsymbol{\gamma}\big(B_r(x) \times B_r(y)\big) > 0,$$

for suitable $y \in \mathbb{R}^d$ with $(x,y) \in \Gamma = \operatorname{spt} \boldsymbol{\gamma}$. Since $x \in X$ and $r > 0$ were arbitrary, we get that $X \subset \operatorname{spt} \varrho$, which implies that $\operatorname{int} \overline{\operatorname{conv}}(X) \subset \Omega$.

Conversely, for every $x \in \Omega$ there exists a ball $B_r(x) \subset \overline{\operatorname{conv}} \operatorname{spt} \varrho$ for some $r > 0$. Pick an open d-cube Q centered at x as large as possible with $Q \subset B_{r/2}(x)$. Then the closure $\overline{Q}$ is the convex hull of its corners $x_i \in \partial B_{r/2}(x)$, which satisfy

$$x_i \in \overline{\operatorname{conv}} \operatorname{spt} \varrho \quad \text{for } i = 1 \dots 2^d.$$

Let $\ell > 0$ denote the side length of Q and $0 < \varepsilon < \ell/8$. Then there exist

$$y_i \in B_\varepsilon(x_i) \cap \operatorname{conv} \operatorname{spt} \varrho \quad \text{for } i = 1 \dots 2^d.$$

Each y_i can be written as a convex combination

$$y_i = \sum_{k=1}^{N_i} \lambda_{i,k} z_{i,k} \quad \text{with } \lambda_{i,k} \in [0,1] \text{ and } \sum_{k=1}^{N_i} \lambda_{i,k} = 1,$$

for suitable $z_{i,k} \in \operatorname{spt} \varrho$ and $N_i \in \mathbb{N}$. We now claim that for any $z \in \operatorname{spt} \varrho$ and $\varepsilon > 0$ there exists $\bar{z} \in B_\varepsilon(z) \cap X$. Assume for the moment that the claim is true. Then for each $z_{i,k}$ we can find $\bar{z}_{i,k} \in B_\varepsilon(z_{i,k}) \cap X$. We define convex combinations

$$\bar{y}_i := \sum_{k=1}^{N_i} \lambda_{i,k} \bar{z}_{i,k} \quad \text{for all } i = 1 \dots 2^d,$$

which satisfy $\|y_i - \bar{y}_i\| \leqslant \varepsilon$ and thus $\bar{y}_i \in B_{2\varepsilon}(x_i)$ for all i. Consequently, the convex hull of these $\bar{y}_i$ contains a d-cube centered at x with side length $\ell/2$, which in turn contains a ball $B_\delta(x)$ for $\delta > 0$ small enough. By construction, this ball is a subset of the convex hull of the $\bar{z}_{i,k} \in X$ from above, so that $x \in \operatorname{int} \overline{\operatorname{conv}}(X)$. This proves the lemma. To establish the claim, assume that on the contrary, there exists $\varepsilon > 0$ with the property that for all $\bar{z} \in B_\varepsilon(z)$ we have $\bar{z} \notin X$. Then

$$\begin{aligned} \varrho\big(B_\varepsilon(z)\big) &= \boldsymbol{\gamma}\big(B_\varepsilon(z) \times \mathbb{R}^d\big) \\ &\leqslant \boldsymbol{\gamma}\Big((\mathbb{R}^d \setminus X) \times \mathbb{R}^d\Big) \leqslant \boldsymbol{\gamma}\Big((\mathbb{R}^d \times \mathbb{R}^d) \setminus \Gamma\Big) = 0. \end{aligned}$$

The second equality follows from the fact that $\boldsymbol{\gamma}$ (being a finite Borel measure on a locally compact Hausdorff space with countable basis) is inner regular; see [44]. We conclude that $z \notin \operatorname{spt} \varrho$, which is a contradiction. ☐

3.2. **Minimal Acceleration Cost.** Suppose that $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ are given. For any timestep $\tau > 0$ we would like to minimize the acceleration $\mathrm{A}_\tau(\boldsymbol{\mu}, \boldsymbol{\gamma})$ over all $\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ with

(1) the transport plan taking $\mathbb{p}^1\#\boldsymbol{\mu}$ to $\mathbb{p}^1\#\boldsymbol{\gamma}$ is monotone,
(2) the velocity distribution of $\boldsymbol{\gamma}$ is tangent to C_ϱ at the new configuration

(where the tangency to C_ϱ is yet to be specified). As mentioned above, this would be consistent with the usual setting in continuum mechanics. Unfortunately, tangent cones often do not possess good continuity properties, as can already be observed in convex polygons in $\mathbb{R}^2$: the tangent cone at any point on an edge of the polygon is a half-space. But the tangent cone collapses to a smaller set at a corner. Consequently, the distance of a fixed point in $\mathbb{R}^2$ to the tangent cone may jump upwards as the base point of the tangent cone approaches a corner of the polygon.

We will therefore use an operator splitting: We first search for the transport that minimizes the acceleration cost, not imposing any restrictions on the final velocity, which will be determined a posteriori by formula (1.16). Then we project this new velocity onto the tangent cone (to be defined) at the new configuration. The second term in (1.15) now measures the cost of realizing a feasible velocity.

As explained above, if the velocity distribution of the second measure in (2.3) is not fixed, then the minimal acceleration cost simplifies. We therefore consider the following minimization problem: find the minimizer $\boldsymbol{\beta}_\tau \in \mathscr{P}_2(\mathbb{R}^{3d})$ of

$$\boldsymbol{\beta} \mapsto \frac{3}{4\tau^2} \int_{\mathbb{R}^{3d}} |(x + \tau\xi) - z|^2 \, \boldsymbol{\beta}(d\boldsymbol{x}, dz) \tag{3.3}$$

among all $\boldsymbol{\beta} \in \mathscr{P}_2(\mathbb{R}^{3d})$ with the following two properties:

$$(1.) \quad (\mathbb{p}^1, \mathbb{p}^2)\#\boldsymbol{\beta} = \boldsymbol{\mu}, \qquad (2.) \quad (\mathbb{p}^1, \mathbb{p}^3)\#\boldsymbol{\beta} \in C_\varrho. \tag{3.4}$$

It will be convenient to define $\boldsymbol{\upsilon}_\tau := (\mathbb{x}, \mathbb{x} + \tau\mathbb{v})\#\boldsymbol{\mu}$ and instead to minimize

$$\boldsymbol{\alpha} \mapsto \frac{3}{4\tau^2} \int_{\mathbb{R}^{3d}} |y-z|^2 \, \boldsymbol{\alpha}(dx, dy, dz)$$

over all $\boldsymbol{\alpha} \in \mathscr{P}_2(\mathbb{R}^{3d})$ with the following two properties:

$$(1.) \quad (\mathbb{p}^1, \mathbb{p}^2)\#\boldsymbol{\alpha} = \boldsymbol{\upsilon}_\tau, \qquad (2.) \quad (\mathbb{p}^1, \mathbb{p}^3)\#\boldsymbol{\alpha} \in C_\varrho. \tag{3.5}$$

Notice that for every $\tau > 0$ the push-forward under the map $(x,\xi) \mapsto (x+\tau\xi, \xi)$ with $(x,\xi) \in \mathbb{R}^{2d}$ is an automorphism between the spaces of measures $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathscr{P}_2(\mathbb{R}^{3d})$ satisfying (3.5) and (3.4), respectively. We observe that (modulo the factor $3/4\tau^2$) we obtain exactly the minimization that defines the distance W_ϱ (see Theorem 2.5), where the second measure is allowed to range freely over the set C_ϱ. Therefore the minimization amounts to finding the element in C_ϱ closest to $\boldsymbol{\upsilon}_\tau$ with respect to the distance W_ϱ, i.e., to computing the metric projection onto C_ϱ.

3.3. **Metric Projection.** In order to study the minimization problem introduced in the previous section, we introduce on $\mathscr{P}_\varrho(\mathbb{R}^{2d})$ the analogues of scalar multiplication and vector addition in Hilbert spaces. This will allow us to define convexity of subsets of $\mathscr{P}_\varrho(\mathbb{R}^{2d})$ and metric projections onto such sets.

Definition 3.5 (Addition/Multiplication)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ be given.

(i.) *Scaling.* For any $\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $s \in \mathbb{R}$ let

$$s\boldsymbol{\gamma} := (\mathbb{x}, s\mathbb{y})\#\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d}).$$

(ii.) *Sum.* For any $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ let

$$\boldsymbol{\gamma}^1 \oplus \boldsymbol{\gamma}^2 := \Big\{ (\mathbb{x}, \mathbb{y}^1 + \mathbb{y}^2)\#\boldsymbol{\alpha} \colon \boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \Big\} \subset \mathscr{P}_\varrho(\mathbb{R}^{2d}).$$

If the plans are induced by functions, then the operations in Definition 3.5 reduce to the usual vector space structures on the Hilbert space $\mathscr{L}^2(\mathbb{R}^d, \varrho)$. Note also that for all $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $s \in \mathbb{R}$ we have the useful equality

$$\mathrm{W}_\varrho(s\boldsymbol{\gamma}^1, s\boldsymbol{\gamma}^2) = |s| \mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2).$$

We refer the reader to Section 4.1 in [48] for a proof.

Definition 3.6 (Closed Convex Cone)**.** A non-empty subset $C \subset \mathscr{P}_\varrho(\mathbb{R}^{2d})$ will be called a closed convex set if it has the following two properties:

(i.) **Closed.** Consider $\boldsymbol{\gamma}^k \in C$ and $\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ with

$$\mathrm{W}_\varrho(\boldsymbol{\gamma}^k, \boldsymbol{\gamma}) \longrightarrow 0 \quad \text{as } k \to \infty.$$

Then also $\boldsymbol{\gamma} \in C$.

(ii.) **Convex.** For all $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in C$ and $s \in [0,1]$ we have

$$(1-s)\boldsymbol{\gamma}^1 \oplus s\boldsymbol{\gamma}^2 \subset C. \tag{3.6}$$

The set C is a closed convex cone if it also has the following property:

(iii.) **Cone.** For all $\boldsymbol{\gamma} \in C$ and $s \geqslant 0$ we have $s\boldsymbol{\gamma} \in C$.

We consider metric projections onto closed convex sets in $\mathscr{P}_\varrho(\mathbb{R}^{2d})$. They have similar properties like projections in Hilbert spaces.

Proposition 3.7 (Metric Projection)**.** *Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ be given and $C \subset \mathscr{P}_\varrho(\mathbb{R}^{2d})$ a closed convex set. For any $\boldsymbol{\upsilon} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ there is a unique $\mathbb{p}_C(\boldsymbol{\upsilon}) \in C$ with*

$$\mathrm{W}_\varrho\big(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon})\big) \leqslant \mathrm{W}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}) \quad \textit{for all } \boldsymbol{\eta} \in C.$$

For every $\boldsymbol{\eta} \in C$ and all $\boldsymbol{\beta} \in \mathscr{P}_2(\mathbb{R}^{4d})$ with

$$\begin{aligned} (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^2)\#\boldsymbol{\beta} &\in \mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon})\big), \\ (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^3)\#\boldsymbol{\beta} &\in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}). \end{aligned} \tag{3.7}$$

we have the inequality

$$\int_{\mathbb{R}^{4d}} \langle y^1 - y^2, y^2 - y^3 \rangle \, \boldsymbol{\beta}(dx, dy^1, dy^2, dy^3) \geqslant 0. \tag{3.8}$$

Conversely, assume that there exists a $\boldsymbol{\zeta} \in \mathscr{C}$ with the following property: for all $\boldsymbol{\eta} \in C$ there exists $\boldsymbol{\beta} \in \mathscr{P}_2(\mathbb{R}^{4d})$ with

$$\begin{aligned} (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^2)\#\boldsymbol{\beta} &\in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\zeta}), \\ (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^3)\#\boldsymbol{\beta} &\in \mathrm{Opt}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}), \end{aligned} \tag{3.9}$$

such that inequality (3.8) *holds true. Then $\boldsymbol{\zeta} = \mathbb{p}_C(\boldsymbol{\upsilon})$.*

For any $\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and any $\boldsymbol{\omega} \in \mathscr{P}_2(\mathbb{R}^{5d})$ such that

$$\begin{aligned} (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^2)\#\boldsymbol{\omega} &\in \mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon}^1, \mathbb{p}_C(\boldsymbol{\upsilon}^1)\big), \\ (\mathbb{x}, \mathbb{y}^3, \mathbb{y}^4)\#\boldsymbol{\omega} &\in \mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon}^2, \mathbb{p}_C(\boldsymbol{\upsilon}^2)\big), \end{aligned} \tag{3.10}$$

we can estimate as follows:

$$\int_{\mathbb{R}^{5d}} |y^2 - y^4|^2 \, \boldsymbol{\omega}(dx, dy^1, \ldots, dy^4) \leqslant \int_{\mathbb{R}^{5d}} |y^1 - y^3|^2 \, \boldsymbol{\omega}(dx, dy^1, \ldots, dy^4). \tag{3.11}$$

In particular, we have the contraction $\mathrm{W}_\varrho(\mathbb{p}_C(\boldsymbol{\upsilon}^1), \mathbb{p}_C(\boldsymbol{\upsilon}^2)) \leqslant \mathrm{W}_\varrho(\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2)$.

Proof. The proof is similar to the one of Proposition 4.30 in [48].

Step 1. Let $d := \inf\{\mathrm{W}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}) \colon \boldsymbol{\eta} \in C\} \geqslant 0$ and consider a sequence of plans $\boldsymbol{\eta}^n \in C$ such that $\mathrm{W}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}^n) \longrightarrow d$ as $n \to \infty$. For any pair of indices $m, n \in \mathbb{N}$ choose $\boldsymbol{\beta}^{m,n} \in \mathscr{P}_2(\mathbb{R}^{4d})$ with the property that

$$\begin{aligned} (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^2)\#\boldsymbol{\beta}^{m,n} &\in \mathrm{Opt}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}^m), \\ (\mathbb{x}, \mathbb{y}^1, \mathbb{y}^3)\#\boldsymbol{\beta}^{m,n} &\in \mathrm{Opt}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}^n), \end{aligned}$$

and define the plans

$$\begin{aligned} \boldsymbol{\alpha}^{m,n} &:= (\mathbb{x}, \mathbb{y}^2, \mathbb{y}^3)\#\boldsymbol{\beta}^{m,n} \qquad \in \mathrm{Adm}_\varrho(\boldsymbol{\eta}^m, \boldsymbol{\eta}^n), \\ \boldsymbol{\eta}^{m,n} &:= (\mathbb{x}, \tfrac{1}{2}\mathbb{y}^2 + \tfrac{1}{2}\mathbb{y}^3)\#\boldsymbol{\beta}^{m,n} \in C. \end{aligned}$$

The last inclusion follows from convexity (3.6). We claim that the sequence $\{\boldsymbol{\eta}^n\}_n$ is a Cauchy sequence with respect to W_ϱ. Indeed we have

$$\begin{aligned}
&\frac{1}{2}\mathrm{W}_\varrho(\boldsymbol{\eta}^m,\boldsymbol{\eta}^n)^2\\
&\quad\leqslant \int_{\mathbb{R}^{4d}} \frac{1}{2}|y^2-y^3|^2\,\boldsymbol{\beta}^{m,n}(dx,dy^1\dots dy^3)\\
&\quad= \int_{\mathbb{R}^{4d}} \left(|y^1-y^2|^2+|y^1-y^3|^2-2\left|y^1-\frac{y^2+y^3}{2}\right|^2\right)\boldsymbol{\beta}^{m,n}(dx,dy^1\dots dy^3)\\
&\quad\leqslant \mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}^m)^2+\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}^n)^2-2\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}^{m,n})^2
\end{aligned}$$

for all $m,n\in\mathbb{N}$. Notice that $\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}^{m,n})\geqslant d$ because $\boldsymbol{\eta}^{m,n}\in C$. This yields

$$\frac{1}{2}\mathrm{W}_\varrho(\boldsymbol{\eta}^m,\boldsymbol{\eta}^n)^2\leqslant \mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}^m)^2+\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}^n)^2-2d^2. \tag{3.12}$$

Since by assumption the sequence $\{\boldsymbol{\eta}^n\}_n$ is minimizing, the right-hand side of (3.12) converges to zero as $m,n\to\infty$, which proves our claim. Recall that $(\mathscr{P}_\varrho(\mathbb{R}^{2d}),\mathrm{W}_\varrho)$ is a complete metric space. It follows that there is a $\mathbb{p}_C(\boldsymbol{\upsilon})\in C$ with the property that $\mathrm{W}_\varrho(\boldsymbol{\eta}^n,\mathbb{p}_C(\boldsymbol{\upsilon}))\longrightarrow 0$. By lower semicontinuity of the distance W_ϱ, we now have $\mathrm{W}_\varrho(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon}))=d$. This establishes the existence of a minimizer.

Step 2. To prove uniqueness, assume that there exists $\boldsymbol{\eta}\in C$ with $\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta})=d$. Now choose a plan $\boldsymbol{\beta}\in\mathscr{P}_2(\mathbb{R}^{4d})$ that satisfies

$$\begin{aligned}
&(\mathbb{x},\mathbb{y}^1,\mathbb{y}^2)\#\boldsymbol{\beta}\in\mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon})\big),\\
&(\mathbb{x},\mathbb{y}^1,\mathbb{y}^3)\#\boldsymbol{\beta}\in\mathrm{Opt}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}),
\end{aligned}$$

and define the plans

$$\begin{aligned}
\boldsymbol{\alpha}&:=(\mathbb{x},\mathbb{y}^1,\tfrac12\mathbb{y}^2+\tfrac12\mathbb{y}^3)\#\boldsymbol{\beta}\in\mathrm{Adm}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\eta}}),\\
\bar{\boldsymbol{\eta}}&:=(\mathbb{x},\tfrac12\mathbb{y}^2+\tfrac12\mathbb{y}^3)\#\boldsymbol{\beta}\qquad\in C.
\end{aligned}$$

The last inclusion again follows from convexity (3.6). We can then estimate

$$\begin{aligned}
2\mathrm{W}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\eta}})^2&\leqslant\int_{\mathbb{R}^{4d}}2\left|y^1-\frac{y^2+y^3}{2}\right|^2\boldsymbol{\beta}(dx,dy^1\dots dy^3)\\
&=\int_{\mathbb{R}^{4d}}\left(|y^1-y^2|^2+|y^1-y^3|^2-\frac12|y^2-y^3|^2\right)\boldsymbol{\beta}(dx,dy^1\dots dy^3)\\
&\leqslant\mathrm{W}_\varrho\big(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon})\big)^2+\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta})^2-\frac12\mathrm{W}_\varrho\big(\mathbb{p}_C(\boldsymbol{\upsilon}),\boldsymbol{\eta}\big)^2.
\end{aligned}\tag{3.13}$$

By our choice of $\boldsymbol{\eta}$, we obtain $\mathrm{W}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\eta}})^2\leqslant d^2-\frac14\mathrm{W}_\varrho(\mathbb{p}_C(\boldsymbol{\upsilon}),\boldsymbol{\eta})^2$, which shows that if $\mathbb{p}_C(\boldsymbol{\upsilon})$ and $\boldsymbol{\eta}$ are different, then $\mathrm{W}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\eta}})<d$. This contradicts the definition of d because $\bar{\boldsymbol{\eta}}\in C$. Therefore the minimizer must be unique.

Step 3. For any $\boldsymbol{\eta}\in C$ consider now $\boldsymbol{\beta}\in\mathscr{P}_2(\mathbb{R}^{4d})$ with (3.7). For every $s>0$ we define $\boldsymbol{\eta}_s:=(\mathbb{x},(1-s)\mathbb{y}^2+s\mathbb{y}^3)\#\boldsymbol{\beta}\in C$; see (3.6). Then

$$\begin{aligned}
&\int_{\mathbb{R}^{4d}}|y^1-y^2|^2\,\boldsymbol{\beta}(dx,dy^1,dy^2,dy^3)=\mathrm{W}_\varrho\big(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon})\big)^2\\
&\qquad\leqslant\mathrm{W}_\varrho(\boldsymbol{\upsilon},\boldsymbol{\eta}_s)^2\leqslant\int_{\mathbb{R}^{4d}}\big|y^1-\big((1-s)y^2+sy^3\big)\big|^2\,\boldsymbol{\beta}(dx,dy^1,dy^2,dy^3),
\end{aligned}$$

which implies the estimate

$$0 \geqslant -2s \int_{\mathbb{R}^{4d}} \langle y^1 - y^2, y^2 - y^3 \rangle \, \boldsymbol{\beta}(dx, dy^1, dy^2, dy^3) \\ - s^2 \int_{\mathbb{R}^{4d}} |y^2 - y^3|^2 \, \boldsymbol{\beta}(dx, dy^1, dy^2, dy^3). \tag{3.14}$$

Notice that the second integral on the right-hand side of (3.14) is finite. Dividing the inequality (3.14) by $-2s < 0$ and letting $s \to 0$, we obtain (3.8).

Conversely, let $\boldsymbol{\zeta} \in C$. Assume that for every $\boldsymbol{\eta} \in C$ there exists $\boldsymbol{\beta} \in \mathscr{P}_2(\mathbb{R}^{4d})$ with (3.9) satisfying (3.8). Then we can estimate as follows:

$$\begin{aligned} \mathrm{W}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\zeta})^2 - \mathrm{W}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta})^2 &\leqslant \int_{\mathbb{R}^{4d}} \Big(|y^1 - y^2|^2 - |y^1 - y^3|^2 \Big) \, \boldsymbol{\beta}(dx, dy^1, dy^2, dy^3) \\ &= -2 \int_{\mathbb{R}^{4d}} \langle y^1 - y^2, y^2 - y^3 \rangle \, \boldsymbol{\beta}(dx, dy^1, dy^2, dy^3) \\ &\quad - \int_{\mathbb{R}^{4d}} |y^2 - y^3|^2 \, \boldsymbol{\beta}(dx, dy^1, dy^2, dy^3), \end{aligned}$$

which is non-positive, by assumption. Since $\boldsymbol{\eta} \in C$ was arbitrary, the plan $\boldsymbol{\zeta}$ must be equal to the uniquely determined metric projection $\mathbb{p}_C(\boldsymbol{\upsilon})$.

Step 4. Consider now $\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and their metric projections onto C. For all $\boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2)$ there exists $\boldsymbol{\omega} \in \mathscr{P}_2(\mathbb{R}^{5d})$ with $(\mathbb{x}, \mathbb{y}^1, \mathbb{y}^3)\#\boldsymbol{\omega} = \boldsymbol{\alpha}$ and (3.10). Since $(\mathbb{x}, \mathbb{y}^1, \mathbb{y}^4)\#\boldsymbol{\omega} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}^1, \mathbb{p}_C(\boldsymbol{\upsilon}^2))$, we apply (3.8) and obtain

$$\int_{\mathbb{R}^{5d}} \langle y^1 - y^2, y^2 - y^4 \rangle \, \boldsymbol{\omega}(dx, dy^1, \ldots, dy^4) \geqslant 0. \tag{3.15}$$

Similarly, since $(\mathbb{x}, \mathbb{y}^3, \mathbb{y}^2)\#\boldsymbol{\omega} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}^2, \mathbb{p}_C(\boldsymbol{\upsilon}^1))$, we have

$$\int_{\mathbb{R}^{5d}} \langle y^3 - y^4, y^4 - y^2 \rangle \, \boldsymbol{\omega}(dx, dy^1, \ldots, dy^4) \geqslant 0. \tag{3.16}$$

Adding (3.15) and (3.16) and using the Cauchy-Schwarz inequality, we get (3.11). The left-hand side of (3.11) is always bigger than or equal to $\mathrm{W}_\varrho(\mathbb{p}_C(\boldsymbol{\upsilon}^1), \mathbb{p}_C(\boldsymbol{\upsilon}^2))^2$. The right-hand side equals $\mathrm{W}_\varrho(\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2)^2$ whenever $\boldsymbol{\alpha} \in \mathrm{Opt}_\varrho(\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2)$. □

3.4. **Non-Splitting Projections.** Under a suitable assumption on the closed convex set, the projections in Proposition 3.7 can be expressed in terms of maps.

Assumption 3.8. For any $\boldsymbol{\eta} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $\boldsymbol{\zeta} \in C$ we consider the disintegrations $\boldsymbol{\eta}(dx, dy) =: \eta_x(dy) \, \varrho(dx)$ and $\boldsymbol{\zeta}(dx, dy) =: \zeta_x(dy) \, \varrho(dx)$. We assume that if

$$\operatorname{spt} \eta_x \subset \overline{\operatorname{conv}}(\operatorname{spt} \zeta_x) \quad \text{for } \varrho\text{-a.e. } x \in \mathbb{R}^d, \tag{3.17}$$

then also $\boldsymbol{\eta} \in C$.

Proposition 3.9 (Properties of $\mathbb{p}_C(\boldsymbol{\upsilon})$)**.** *Let the closed convex set $C \subset \mathscr{P}_\varrho(\mathbb{R}^{2d})$ satisfy Assumption 3.8 and let $\mathbb{p}_C(\boldsymbol{\upsilon})$ be the metric projection of $\boldsymbol{\upsilon} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ onto C; see Proposition 3.7. Then there exists a unique $\mathbf{z}_{\boldsymbol{\upsilon}} \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\upsilon})$ with*

$$\mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon})\big) = \Big\{ (\mathbb{x}, \mathbb{y}, \mathbf{z}_{\boldsymbol{\upsilon}})\#\boldsymbol{\upsilon} \Big\}. \tag{3.18}$$

Proof. The proof is similar to the one of Propositions 4.32 in [48].

Step 1. Fix any $\boldsymbol{\alpha} \in \mathrm{Opt}_\varrho(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon}))$ and consider the disintegration

$$\boldsymbol{\alpha}(dx, dy, dz) =: \alpha_{(x,y)}(dz) \, \boldsymbol{\upsilon}(dx, dy). \tag{3.19}$$

Then we define the function

$$\mathbf{z}_{\boldsymbol{\upsilon}}(x,y) := \int_{\mathbb{R}^d} z\,\alpha_{(x,y)}(dz) \quad \text{for } \boldsymbol{\upsilon}\text{-a.e. } (x,y)\in\mathbb{R}^{2d}, \tag{3.20}$$

and the plans

$$\begin{aligned}\bar{\boldsymbol{\alpha}} &:= (\mathbb{x},\mathbb{y},\mathbf{z}_{\boldsymbol{\upsilon}})\#\boldsymbol{\upsilon} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\upsilon}}),\\ \bar{\boldsymbol{\upsilon}} &:= (\mathbb{x},\mathbf{z}_{\boldsymbol{\upsilon}})\#\boldsymbol{\upsilon}.\end{aligned}$$

We claim that $\bar{\boldsymbol{\upsilon}}\in C$. Notice first that clearly

$$\mathbf{z}_{\boldsymbol{\upsilon}}(x,y) \subset \overline{\mathrm{conv}}(\mathrm{spt}\,\alpha_{(x,y)}) \quad \text{for } \boldsymbol{\upsilon}\text{-a.e. } (x,y)\in\mathbb{R}^{2d}.$$

Now consider the disintegrations

$$\begin{aligned}\bar{\boldsymbol{\upsilon}}(dx,dz) &=: \bar{\upsilon}_x(dz)\,\varrho(dx),\\ \boldsymbol{\upsilon}(dx,dy) &=: \upsilon_x(dy)\,\varrho(dx),\\ \mathbb{p}_C(\boldsymbol{\upsilon})(dx,dz) &=: \mathbb{p}_C(\upsilon)_x(dz)\,\varrho(dx).\end{aligned}$$

It follows that

$$\mathbb{p}_C(\upsilon)_x(dz) = \int_{\mathbb{R}^d} \alpha_{(x,y)}(dz)\,\upsilon_x(dy) \quad \text{for } \varrho\text{-a.e. } x\in\mathbb{R}^d,$$

and so $\mathrm{spt}\,\alpha_{(x,y)} \subset \mathrm{spt}\,\mathbb{p}_C(\upsilon)_x$ for υ_x-a.e. $y\in\mathbb{R}^d$. This yields

$$\mathrm{spt}\,\bar{\upsilon}_x \subset \overline{\mathrm{conv}}\left(\mathrm{spt}\,\mathbb{p}_C(\upsilon)_x\right) \quad \text{for } \varrho\text{-a.e. } x\in\mathbb{R}^d.$$

By Asssumption 3.8, this implies that $\bar{\boldsymbol{\upsilon}}\in C$.

Using that $\bar{\boldsymbol{\alpha}}\in\mathrm{Adm}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\upsilon}})$, we now estimate

$$\begin{aligned}\mathrm{W}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\upsilon}})^2 &\leqslant \int_{\mathbb{R}^{2d}} |y-\mathbf{z}_{\boldsymbol{\upsilon}}(x,y)|^2\,\boldsymbol{\upsilon}(dx,dy)\\ &= \int_{\mathbb{R}^{2d}} \left|\int_{\mathbb{R}^d}(y-z)\,\alpha_{(x,y)}(dz)\right|^2 \boldsymbol{\upsilon}(dx,dy)\\ &\leqslant \int_{\mathbb{R}^{2d}}\int_{\mathbb{R}^d} |y-z|^2\,\alpha_{(x,y)}(dz)\,\boldsymbol{\upsilon}(dx,dy) = \mathrm{W}_\varrho(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon}))^2.\end{aligned}$$

The first equality follows from (3.20) and the second one from (3.19). For the second inequality we have used Jensen's inequality. Recall that Jensen's inequality is strict unless the probability measure is a Dirac measure, which implies that if $\alpha_{(x,y)}$ is not a Dirac measure for $\boldsymbol{\upsilon}$-a.e. $(x,y)\in\mathbb{R}^{2d}$, then $\mathrm{W}_\varrho(\boldsymbol{\upsilon},\bar{\boldsymbol{\upsilon}}) < \mathrm{W}_\varrho(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon}))$. This contradicts the definition of $\mathbb{p}_C(\boldsymbol{\upsilon})$ since $\bar{\boldsymbol{\upsilon}}\in C$. We conclude that

$$\alpha_{(x,y)}(dz) = \delta_{\mathbf{z}_{\boldsymbol{\upsilon}}(x,y)}(dz) \quad \text{for } \boldsymbol{\upsilon}\text{-a.e. } (x,y)\in\mathbb{R}^{2d},$$

and thus $\boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}}$. The same argument works for all $\boldsymbol{\alpha}\in\mathrm{Opt}_\varrho(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon}))$, and so all optimal transport plans between $\boldsymbol{\upsilon}$ and $\mathbb{p}_C(\boldsymbol{\upsilon})$ are induced by maps.

Step 2. To prove uniqueness, assume there exist two maps $\mathbf{z}^1,\mathbf{z}^2\in\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\upsilon})$ such that $\boldsymbol{\alpha}^k := (\mathbb{x},\mathbb{y},\mathbf{z}^k)\#\boldsymbol{\upsilon}\in\mathrm{Opt}_\varrho(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon}))$ for $k=1..2$. Let

$$\begin{aligned}\bar{\boldsymbol{\beta}} &:= (\mathbb{x},\mathbb{y},\mathbf{z}^1,\mathbf{z}^2)\#\boldsymbol{\upsilon},\\ \bar{\boldsymbol{\alpha}} &:= (\mathbb{x},\mathbb{y}^1,\tfrac{1}{2}\mathbb{y}^2+\tfrac{1}{2}\mathbb{y}^3)\#\bar{\boldsymbol{\beta}}.\end{aligned}$$

We claim that $\bar{\boldsymbol{\alpha}}\in\mathrm{Opt}_\varrho(\boldsymbol{\upsilon},\mathbb{p}_C(\boldsymbol{\upsilon}))$. If this is true, and if $\mathbf{z}^1$ and $\mathbf{z}^2$ are different, then $\bar{\boldsymbol{\alpha}}$ is not induced by a map, in contradiction to what we proved in Step 1. We can therefore define $\mathbf{z}_{\boldsymbol{\upsilon}}$ unambiguously by the property (3.18).

To prove the claim, let $\bar{\boldsymbol{\upsilon}} := (\mathbb{x}, \mathbb{y}^2)\#\bar{\boldsymbol{\alpha}} = (\mathbb{x}, \frac{1}{2}\mathbb{y}^2 + \frac{1}{2}\mathbb{y}^3)\#\bar{\boldsymbol{\beta}}$ and note that

$$(\mathbb{x}, \mathbb{y}^2, \mathbb{y}^3)\#\bar{\boldsymbol{\beta}} \in \mathrm{Adm}_\varrho\big(\mathbb{p}_C(\boldsymbol{\upsilon}), \mathbb{p}_C(\boldsymbol{\upsilon})\big).$$

Then $\bar{\boldsymbol{\upsilon}} \in C$ because of convexity (3.6). We have $\bar{\boldsymbol{\alpha}} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}, \bar{\boldsymbol{\upsilon}})$ and

$$(\mathbb{x}, \mathbb{y}^1, \mathbb{y}^2)\#\bar{\boldsymbol{\beta}} \in \mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon})\big),$$
$$(\mathbb{x}, \mathbb{y}^1, \mathbb{y}^3)\#\bar{\boldsymbol{\beta}} \in \mathrm{Opt}_\varrho\big(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon})\big).$$

Arguing as in estimate (3.13), we obtain that $\mathrm{W}_\varrho(\boldsymbol{\upsilon}, \bar{\boldsymbol{\upsilon}}) \leqslant \mathrm{W}_\varrho(\boldsymbol{\upsilon}, \mathbb{p}_C(\boldsymbol{\upsilon})) = d$, which shows that $\bar{\boldsymbol{\upsilon}} = \mathbb{p}_C(\boldsymbol{\upsilon})$, by uniqueness of the minimizer. □

Remark 3.10. Under Assumption 3.8, the third part of Proposition 3.7 simplifies as follows: for any plans $\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2 \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ let $\mathbf{z}^k \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\upsilon}^k)$ be defined as in (3.18) for $k = 1..2$. Then we have the following inequality:

$$\int_{\mathbb{R}^{3d}} |\mathbf{z}^1(x, y^1) - \mathbf{z}^2(x, y^2)|^2 \,\boldsymbol{\alpha}(dx, dy^1, dy^2)$$
$$\leqslant \int_{\mathbb{R}^{3d}} |y^1 - y^2|^2 \,\boldsymbol{\alpha}(dx, dy^1, dy^2) \quad \text{for all } \boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}^1, \boldsymbol{\upsilon}^2).$$

Remark 3.11. If C is a closed convex cone, then (3.8) implies the following statement: for every $\boldsymbol{\eta} \in C$ and all $\boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta})$ we have that

$$\int_{\mathbb{R}^{3d}} \langle y - \mathbf{z}_{\boldsymbol{\upsilon}}(x, y), \mathbf{z}_{\boldsymbol{\upsilon}}(x, y)\rangle \,\boldsymbol{\upsilon}(dx, dy) = 0, \tag{3.21}$$
$$\int_{\mathbb{R}^{3d}} \langle y - \mathbf{z}_{\boldsymbol{\upsilon}}(x, y), z\rangle \,\boldsymbol{\alpha}(dx, dy, dz) \leqslant 0. \tag{3.22}$$

Indeed note first that because of (3.18), the inequality (3.8) reads as follows:

$$\int_{\mathbb{R}^{3d}} \langle y - \mathbf{z}_{\boldsymbol{\upsilon}}(x, y), \mathbf{z}_{\boldsymbol{\upsilon}}(x, y) - z\rangle \,\boldsymbol{\alpha}(dx, dy, dz) \geqslant 0 \tag{3.23}$$

for all $\boldsymbol{\eta}, \boldsymbol{\alpha}$ as above. We have $\boldsymbol{\eta}_0 := (\mathrm{id}, 0)\#\varrho \in C$ since C is a cone. Then

$$\boldsymbol{\alpha}^1 := (\mathbb{x}, \mathbb{y}, 0)\#\boldsymbol{\upsilon} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}, \boldsymbol{\eta}_0).$$

Using $\boldsymbol{\alpha}^1$ in (3.23), we obtain the inequality

$$\int_{\mathbb{R}^{2d}} \langle y - \mathbf{z}_{\boldsymbol{\upsilon}}(x, y), \mathbf{z}_{\boldsymbol{\upsilon}}(x, y)\rangle \,\boldsymbol{\upsilon}(dx, dy) \geqslant 0. \tag{3.24}$$

On the other hand, we have $2\mathbb{p}_C(\boldsymbol{\upsilon}) \in C$ since C is a cone. Then

$$\boldsymbol{\alpha}^2 := (\mathbb{x}, \mathbb{y}, 2\mathbf{z}_{\boldsymbol{\upsilon}})\#\boldsymbol{\upsilon} \in \mathrm{Adm}_\varrho\big(\boldsymbol{\upsilon}, 2\mathbb{p}_C(\boldsymbol{\upsilon})\big).$$

Using $\boldsymbol{\alpha}^2$ in (3.23), we obtain the inequality

$$\int_{\mathbb{R}^{2d}} \langle y - \mathbf{z}_{\boldsymbol{\upsilon}}(x, y), -\mathbf{z}_{\boldsymbol{\upsilon}}(x, y)\rangle \,\boldsymbol{\upsilon}(dx, dy) \geqslant 0. \tag{3.25}$$

We now combine (3.24) and (3.25), and get (3.21) and thus (3.22).

3.5. **Monotone Transport Plans.** Propositions 3.7 and 3.9 apply to C_ϱ.

Proposition 3.12 (Monotone Transport Plans)**.** *Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$. Then*

$$C_\varrho := \Big\{\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})\colon \operatorname{spt}\boldsymbol{\gamma} \textit{ is a monotone subset of } \mathbb{R}^d \times \mathbb{R}^d\Big\}$$

(which is (3.2)*) is a closed convex cone and Assumption 3.8 is satisfied.*

Proof. We proceed in three steps.

Step 1. Consider first plans $\boldsymbol{\gamma}^k$ and $\boldsymbol{\gamma}$ as in Definition 3.6 (i.). Since

$$\mathrm{W}(\boldsymbol{\gamma}^k, \boldsymbol{\gamma}) \leqslant \mathrm{W}_\varrho(\boldsymbol{\gamma}^k, \boldsymbol{\gamma})$$

we have that $\boldsymbol{\gamma}^k \longrightarrow \boldsymbol{\gamma}$ with respect to the Wasserstein distance, and thus narrowly; see Proposition 7.1.5 in [4]. One can check that $\boldsymbol{\gamma} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$. Fix $(x_i, y_i) \in \operatorname{spt}\boldsymbol{\gamma}$ with $i = 1..2$. Since narrow convergence of probability measures implies Kuratowski convergence of their supports (see Proposition 5.1.8 in [4]), there exist

$$(x_i^k, y_i^k) \in \operatorname{spt}\boldsymbol{\gamma}^k \quad \text{such that} \quad \lim_{k\to\infty} (x_i^k, y_i^k) = (x_i, y_i)$$

for $i = 1..2$. Since $\operatorname{spt}\boldsymbol{\gamma}^k$ is monotone for all k, we obtain that

$$\langle x_1 - x_2, y_1 - y_2\rangle = \lim_{k\to\infty} \langle x_1^k - x_2^k, y_1^k - y_2^k\rangle \geqslant 0.$$

Since the (x^i, y^i) were arbitrary, we conclude that $\boldsymbol{\gamma} \in C_\varrho$.

Step 2. Let now $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in C_\varrho$ and $s \in [0,1]$ be given. For any $\boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)$ we define the interpolation transport plan $\boldsymbol{\gamma}_s := (\mathbb{x}, (1-s)\mathbb{y}^1 + s\mathbb{y}^2)\#\boldsymbol{\alpha} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$. Consider now any point $(x, y) \in \operatorname{spt}\boldsymbol{\gamma}_s$. By the definition of support of a measure, for all $\varepsilon > 0$ there exists $(\hat{x}, \hat{y}^1, \hat{y}^2) \in \operatorname{spt}\boldsymbol{\alpha}$ such that

$$\hat{x} \in B_\varepsilon(x) \quad \text{and} \quad (1-s)\hat{y}^1 + s\hat{y}^2 \in B_\varepsilon(y). \tag{3.26}$$

Indeed, assume there exists an $\varepsilon > 0$ with the property that for all $(\hat{x}, \hat{y}^1, \hat{y}^2) \in \operatorname{spt}\boldsymbol{\alpha}$ statement (3.26) is wrong. Then $\boldsymbol{\gamma}_s(B_\varepsilon(x) \times B_\varepsilon(y)) = 0$, which is a contradiction to our choice $(x, y) \in \operatorname{spt}\boldsymbol{\gamma}_s$. Now $(\hat{x}, \hat{y}^1, \hat{y}^2) \in \operatorname{spt}\boldsymbol{\alpha}$ implies that

$$0 < \boldsymbol{\alpha}\Big(B_r(\hat{x}) \times B_r(\hat{y}^1) \times B_r(\hat{y}^2)\Big) \leqslant \boldsymbol{\gamma}^k\Big(B_r(\hat{x}) \times B_r(\hat{y}^k)\Big)$$

for all $r > 0$, with $k = 1..2$. We conclude that $(\hat{x}, \hat{y}^k) \in \operatorname{spt}\boldsymbol{\gamma}^k$.

We can now apply the above argument to a pair of points $(x_i, y_i) \in \operatorname{spt}\boldsymbol{\gamma}_s$, with $i = 1..2$. For any $\varepsilon > 0$ we find $(\hat{x}_i, \hat{y}_i^k) \in \operatorname{spt}\boldsymbol{\gamma}^k$, $i = 1..2$, such that

$$\hat{x}_i \in B_\varepsilon(x_i) \quad \text{and} \quad (1-s)\hat{y}_i^1 + s\hat{y}_i^2 \in B_\varepsilon(y_i).$$

Since $\operatorname{spt}\boldsymbol{\gamma}^k$ is monotone, we obtain the estimate

$$\begin{aligned}\langle x_1 - x_2, y_1 - y_2\rangle &\geqslant (1-s)\langle \hat{x}_1 - \hat{x}_2, \hat{y}_1^1 - \hat{y}_2^1\rangle + s\langle \hat{x}_1 - \hat{x}_2, \hat{y}_1^2 - \hat{y}_2^2\rangle - 4(M+\varepsilon)\varepsilon \\ &\geqslant -4(M+\varepsilon)\varepsilon,\end{aligned}$$

with $M := \max_i\{|x_i|, |y_i|\}$. Since $\varepsilon > 0$ and $(x_i, y_i) \in \operatorname{spt}\boldsymbol{\gamma}_s$ were arbitrary, we get that $\operatorname{spt}\boldsymbol{\gamma}_s$ is monotone. Since $\boldsymbol{\alpha} \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)$ was arbitrary, we obtain (3.6). In a similar way, one proves that if $\boldsymbol{\gamma} \in C_\varrho$, then also $s\boldsymbol{\gamma} \in C_\varrho$ for all $s \geqslant 0$.

Step 3. In order to prove Assumption 3.8, note that if $\boldsymbol{\zeta} \in C_\varrho$, then its support is contained in the graph of a maximal monotone set-valued map u (we may consider a suitable extension if necessary). For ϱ-a.e. $x \in \mathbb{R}^d$ we have $\operatorname{spt}\zeta_x \subset u(x)$, which is a closed and convex set; see [1]. Then $\operatorname{spt}\eta_x \subset u(x)$ as well because of assumption (3.17), which implies that the support of $\boldsymbol{\eta}$ is monotone and hence $\boldsymbol{\eta} \in C_\varrho$. ☐

Remark 3.13. Proposition 3.9 implies that whenever $\boldsymbol{\upsilon} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ is induced by a map, i.e., there exists a $\boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ taking values in $\mathbb{R}^d$ such that

$$\boldsymbol{\upsilon}(dx, dy) = \delta_{\boldsymbol{t}(x)}(dy)\, \varrho(dx),$$

then the projection $\mathbb{p}_{C_\varrho}(\boldsymbol{\upsilon})$ is induced by a map as well:

$$\mathbb{p}_{C_\varrho}(\boldsymbol{\upsilon})(dx, dz) = \delta_{\mathbf{z}_{\boldsymbol{\upsilon}}(x, \boldsymbol{t}(x))}(dz)\, \varrho(dx).$$

Notice that if ϱ is absolutely continuous with respect to the Lebesgue measure, then all monotone transport plans in C_ϱ are in fact induced by maps. This follows in the same way as for optimal transport plans (which are contained in the subdifferentials of convex functions, thus monotone): the set of points where a (maximal) monotone set-valued map is multi-valued is a Lebesgue null set; see [1].

4. Energy Minimization: Pressureless Gases

For our variational time discretization of the pressureless gas dynamics equations (1.3), we divide the time interval $[0, T]$ into subintervals of length $\tau > 0$. For every timestep, we minimize the work 2.7 over the cone of monotone transport plans. As explained in Section 3.2, this reduces to a metric projection, which further simplifies to a minimization over a closed convex cone in a Hilbert space: we may consider monotone transport *maps* instead of plans because of Proposition 3.9.

4.1. **Configuration Manifold.** Going back to our original setup, we will consider monotone transport maps that are defined on measures $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ representing the distribution of mass and *velocity*, not representing transport plans.

Definition 4.1 (Configurations)**.** For any $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, let

$$C_{\boldsymbol{\mu}} := \Big\{ \mathbf{t} \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu}) \colon (\mathbb{x}, \mathbf{t})\#\boldsymbol{\mu} \in C_\varrho \Big\}.$$

Lemma 4.2 (Closed Convex Cone)**.** *$C_{\boldsymbol{\mu}}$ is a closed convex cone in $\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$.*

Proof. We observe first that for any $\mathbf{t}^1, \mathbf{t}^2 \in C_{\boldsymbol{\mu}}$, we have that

$$(\mathbb{x}, \mathbf{t}^1, \mathbf{t}^2)\#\boldsymbol{\mu} \in \mathrm{Adm}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \tag{4.1}$$

where $\boldsymbol{\gamma}^n := (\mathbb{x}, \mathbf{t}^n)\#\boldsymbol{\mu} \in C_\varrho$ with $n = 1..2$. This implies the estimate

$$\mathrm{W}_\varrho(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \leqslant \|\mathbf{t}^1 - \mathbf{t}^2\|_{\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})}. \tag{4.2}$$

Consider now a sequence $\mathbf{t}^k \longrightarrow \mathbf{t}$ in $\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$ and define $\boldsymbol{\gamma}^k := (\mathbb{x}, \mathbf{t}^k)\#\boldsymbol{\mu}$. Let $\boldsymbol{\gamma} := (\mathbb{x}, \mathbf{t})\#\boldsymbol{\mu}$ and notice that $\mathrm{W}_\varrho(\boldsymbol{\gamma}^k, \boldsymbol{\gamma}) \longrightarrow 0$ because of (4.2). If now $\mathbf{t}^k \in C_{\boldsymbol{\mu}}$ and thus $\boldsymbol{\gamma}^k \in C_\varrho$ for all k, then also $\boldsymbol{\gamma} \in C_\varrho$ since C_ϱ is closed with respect to W_ϱ; see Proposition 3.12. This proves that $\mathbf{t} \in C_{\boldsymbol{\mu}}$. For any $s \in [0, 1]$ we have

$$\boldsymbol{\gamma}_s := \big(\mathbb{x}, (1-s)\mathbf{t}^1 + s\mathbf{t}^2\big)\#\boldsymbol{\mu} = \big(\mathbb{x}, (1-s)\mathbb{y}^1 + s\mathbb{y}^2\big)\#\boldsymbol{\alpha}^{1,2},$$

where $\boldsymbol{\alpha}^{1,2} := (\mathbb{x}, \mathbf{t}^1, \mathbf{t}^2)\#\boldsymbol{\mu}$ with $\mathbf{t}^n \in C_{\boldsymbol{\mu}}$ and $n = 1..2$. Using (4.1), we conclude that $\boldsymbol{\gamma}_s \in (1-s)\boldsymbol{\gamma}^1 \oplus s\boldsymbol{\gamma}^2$ (see Definition 3.5), which is in C_ϱ, by Proposition 3.12. Hence $(1-s)\mathbf{t}^1 + s\mathbf{t}^2 \in C_{\boldsymbol{\mu}}$. The proof that $C_{\boldsymbol{\mu}}$ is a cone is analogous. □

Remark 4.3. For every $\tau > 0$ let $\mathbf{z}_\tau$ be the unique map defined in Proposition 3.9 representing the metric projection of $\boldsymbol{\upsilon}_\tau := (\mathbb{x}, \mathbb{x} + \tau\mathbb{v})\#\boldsymbol{\mu}$ onto the closed convex cone C_ϱ. We define $\mathbf{t}_\tau(x, \xi) := \mathbf{z}_\tau(x, x + \tau\xi)$ for $\boldsymbol{\mu}$-a.e. $(x, \xi) \in \mathbb{R}^{2d}$ so that

$$(\mathbb{x}, \mathbb{y}, \mathbf{z}_\tau)\#\boldsymbol{\upsilon}_\tau = (\mathbb{x}, \mathbb{x} + \tau\mathbb{v}, \mathbf{t}_\tau)\#\boldsymbol{\mu}. \tag{4.3}$$

Then $\mathbf{t}_\tau$ must be the uniquely determined metric projection of $\mathbb{x}+\tau\mathbb{v} \in \mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$ onto the cone $C_{\boldsymbol{\mu}}$ (see [75] for more information about metric projections in Hilbert spaces). Indeed for any $\mathbf{s} \in C_{\boldsymbol{\mu}}$ (for which $\boldsymbol{\gamma}_{\mathbf{s}} := (\mathbb{x},\mathbf{s})\#\boldsymbol{\mu} \in C_\varrho$) we have

$$\begin{aligned}\|(\mathbb{x}+\tau\mathbb{v})-\mathbf{t}_\tau\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})} &= \|\mathbb{y}-\mathbf{z}_\tau\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\upsilon}_\tau)} \\ &= \mathrm{W}_\varrho\big(\boldsymbol{\upsilon}_\tau, \mathbb{p}_{C_\varrho}(\boldsymbol{\upsilon}_\tau)\big) \\ &\leqslant \mathrm{W}_\varrho(\boldsymbol{\upsilon}_\tau,\boldsymbol{\gamma}_{\mathbf{s}}) \leqslant \|(\mathbb{x}+\tau\mathbb{v})-\mathbf{s}\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})}.\end{aligned}$$

The first equality follows from definition (4.3) and the second one from (3.18). The subsequent inequality is true because $\mathbb{p}_{C_\varrho}(\boldsymbol{\upsilon}_\tau)$ is closest to $\boldsymbol{\upsilon}_\tau$ in C_ϱ with respect to W_ϱ, by definition. Finally, we have used that $(\mathbb{x},\mathbb{x}+\tau\mathbb{v},\mathbf{s})\#\boldsymbol{\mu} \in \mathrm{Adm}_\varrho(\boldsymbol{\upsilon}_\tau,\boldsymbol{\gamma}_{\mathbf{s}})$. We will write $\mathbf{t}_\tau = \mathbb{p}_{C_{\boldsymbol{\mu}}}(\mathbb{x}+\tau\mathbb{v})$. The map $\mathbf{t}_\tau$ is uniquely determined by

$$\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi)-\mathbf{t}_\tau(x,\xi), \mathbf{t}_\tau(x,\xi)\rangle\, \boldsymbol{\mu}(dx,d\xi) = 0, \tag{4.4}$$

$$\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi)-\mathbf{t}_\tau(x,\xi), \mathbf{s}(x,\xi)\rangle\, \boldsymbol{\mu}(dx,d\xi) \leqslant 0 \quad \text{for all } \mathbf{s} \in C_{\boldsymbol{\mu}}. \tag{4.5}$$

We just need to combine Lemma 1.1 in [75] with Remark 3.11.

Remark 4.4. A map $\mathbf{t} \in \mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$ is in $C_{\boldsymbol{\mu}}$ if and only if the following statement is true: There exists a Borel set $N_{\mathbf{t}} \subset \mathbb{R}^{2d}$ with $\boldsymbol{\mu}(N_{\mathbf{t}}) = 0$ such that

$$\langle \mathbf{t}(x_1,\xi_1)-\mathbf{t}(x_2,\xi_2), x_1-x_2\rangle \geqslant 0 \quad \text{for all } (x_i,\xi_i) \in \mathbb{R}^{2d}\setminus N_{\mathbf{t}} \tag{4.6}$$

with $i = 1..2$. Indeed consider any $\mathbf{t} \in C_{\boldsymbol{\mu}}$ and $\boldsymbol{\gamma}_{\mathbf{t}} := (\mathbb{x},\mathbf{t})\#\boldsymbol{\mu} \in C_\varrho$. Then

$$N_{\mathbf{t}} := \Big\{(x,\xi)\in\mathbb{R}^{2d} \colon \big(x,\mathbf{t}(x,\xi)\big) \notin \operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}}\Big\} \quad \text{satisfies}$$

$$\boldsymbol{\mu}(N_{\mathbf{t}}) = \boldsymbol{\mu}\Big((\mathbb{x},\mathbf{t})^{-1}(\mathbb{R}^{2d}\setminus\operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}})\Big) = \boldsymbol{\gamma}_{\mathbf{t}}(\mathbb{R}^{2d}\setminus\operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}}) = 0$$

since $\boldsymbol{\gamma}_{\mathbf{t}}$ is inner regular (being a finite Borel measure on a locally compact Hausdorff space with countable basis; see [44]). As $\operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}}$ is monotone, (4.6) follows.

Conversely, suppose $\mathbf{t} \in \mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$ satisfies (4.6). Let $\boldsymbol{\gamma}_{\mathbf{t}} := (\mathbb{x},\mathbf{t})\#\boldsymbol{\mu}$. For any $(x_i,y_i) \in \operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}}$ with $i=1..2$ and any $\varepsilon > 0$ we have

$$\begin{aligned} 0 &< \boldsymbol{\gamma}_{\mathbf{t}}\big(B_\varepsilon(x_i)\times B_\varepsilon(y_i)\big) \\ &= \boldsymbol{\mu}\bigg(\Big\{(x,\xi)\in\mathbb{R}^{2d}\colon \big(x,\mathbf{t}(x,\xi)\big)\in B_\varepsilon(x_i)\times B_\varepsilon(y_i)\Big\}\bigg),\end{aligned}$$

by definition of support. Therefore there exist $(\hat{x}_i,\hat{\xi}_i) \in \mathbb{R}^{2d}$ such that

$$\big(\hat{x}_i, \mathbf{t}(\hat{x}_i,\hat{\xi}_i)\big) \in B_\varepsilon(x_i)\times B_\varepsilon(y_i) \quad \text{for } i = 1..2.$$

We may assume that $(\hat{x}_i,\hat{\xi}_i) \notin N_{\mathbf{t}}$, where $N_{\mathbf{t}}$ is the null set in (4.6). Then

$$\langle y_1-y_2, x_1-x_2\rangle \geqslant \langle \mathbf{t}(\hat{x}_1,\hat{\xi}_1)-\mathbf{t}(\hat{x}_2,\hat{\xi}_2), \hat{x}_1-\hat{x}_2\rangle - M\varepsilon,$$

with $M := 4\max_{i=1..2}\{|x_i|,|y_i|\}+2\varepsilon$. Since $(x_i,y_i)\in\operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}}$ and $\varepsilon>0$ are arbitrary, we conclude that $\operatorname{spt}\boldsymbol{\gamma}_{\mathbf{t}}$ is monotone, and therefore $\mathbf{t}\in C_{\boldsymbol{\mu}}$.

The cone $C_{\boldsymbol{\mu}}$ is the set of all possible configurations, with a reference configuration determined by $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$. Configurations do not permit any interpenetration of matter since the maps are monotone. They do admit, however, the concentration of mass if the transport is not strictly monotone. The fluid element at location/velocity $(x,\xi)\in\mathbb{R}^d\times\mathbb{R}^d$ will never split because its final position is a *function* of (x,ξ).

Definition 4.5 (Tangent Cone)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ be given. The tangent cone of $C_{\boldsymbol{\mu}}$ at the configuration $\mathbf{t} \in C_{\boldsymbol{\mu}}$ is defined as

$$\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}} := \overline{\Big\{ \mathbf{v} \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu}) \colon \text{there exists } \varepsilon > 0 \text{ with } \mathbf{t} + \varepsilon \mathbf{v} \in C_{\boldsymbol{\mu}} \Big\}}^{\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})}.$$

The set $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ is a closed convex cone with vertex at the origin (that is, the zero map) containing $C_{\boldsymbol{\mu}} - \mathbf{t}$. We refer the reader to §2 in [75] for additional information on tangent cones (also: support cones) to closed convex sets in Hilbert spaces.

For any $\mathbf{w} \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$ there exists a unique metric projection onto the closed convex cone $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$, which we will denote by $\mathbb{p}_{\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}}(\mathbf{w})$. It can be characterized by the following property: for all $\mathbf{u} \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$ we have

$$\begin{gathered} \mathbf{u} = \mathbb{p}_{\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}}(\mathbf{w}) \quad \text{if and only if} \\ \int_{\mathbb{R}^{2d}} \langle \mathbf{w}(x,\xi) - \mathbf{u}(x,\xi), \mathbf{u}(x,\xi) - \mathbf{v}(x,\xi) \rangle \, \boldsymbol{\mu}(dx, d\xi) \geqslant 0 \quad \text{for all } \mathbf{v} \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}. \end{gathered}$$

Since $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ is a cone, the latter condition is equivalent to

$$\begin{aligned} &\int_{\mathbb{R}^{2d}} \langle \mathbf{w}(x,\xi) - \mathbf{u}(x,\xi), \mathbf{u}(x,\xi) \rangle \, \boldsymbol{\mu}(dx, d\xi) = 0 \\ &\int_{\mathbb{R}^{2d}} \langle \mathbf{w}(x,\xi) - \mathbf{u}(x,\xi), \mathbf{v}(x,\xi) \rangle \, \boldsymbol{\mu}(dx, d\xi) \leqslant 0 \quad \text{for all } \mathbf{v} \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}. \end{aligned}$$

We also recall the following fact: for any $\mathbf{t} \in C_{\boldsymbol{\mu}}$ we have

$$\|\mathbf{u}\|^{-1}_{\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})} \Big\| \Big(\mathbb{p}_{C_{\boldsymbol{\mu}}}(\mathbf{t} + \mathbf{u}) - \mathbf{t} \Big) - \mathbb{p}_{\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}}(\mathbf{u}) \Big\|_{\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})} \longrightarrow 0 \tag{4.7}$$

as $\mathbf{u} \to 0$ over any locally compact cone of increments; see Lemma 4.6 of [75]. The metric projection $\mathbb{p}_{\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}}$ is therefore the differential of $\mathbb{p}_{C_{\boldsymbol{\mu}}}$ at $\mathbf{t} \in C_{\boldsymbol{\mu}}$.

Let us collect some additional properties of the tangent cone.

Proposition 4.6 (Tangent Cone)**.** *With* $\mathbf{t} \in C_{\boldsymbol{\mu}}$ *and* $\boldsymbol{\gamma}_{\mathbf{t}} := (\mathbb{x}, \mathbf{t})\#\boldsymbol{\mu}$*, assume*

$$\langle y_1 - y_2, x_1 - x_2 \rangle \geqslant \alpha |x_1 - x_2|^2 \quad \text{for all } (x_i, y_i) \in \operatorname{spt} \boldsymbol{\gamma}_{\mathbf{t}} \tag{4.8}$$

and $i = 1..2$*, where* $\alpha > 0$ *is some constant. Then we have:*

(i.) *Every* $\boldsymbol{u} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ *is contained in* $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$*.*
(ii.) *For all* $\mathbf{v} \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ *we also have* $-\mathbf{v} \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$*.*

In particular, the tangent cone $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ *is a closed subspace of* $\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$*.*

Proof. We divide the proof into three steps.

Step 1. Note first that any finite Borel measure ν on a locally compact Hausdorff space Ω with countable base is inner regular. Therefore the space of all continuous functions with compact support is dense in $\mathscr{L}^2(\Omega, \nu)$. We refer the reader to [44] for further details. If Ω is also a vector space, then the same statement is true for *smooth* functions with compact support. For every $\boldsymbol{u} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ there exists thus a sequence of smooth functions $\boldsymbol{u}^m$ with compact support with $\boldsymbol{u}^m \longrightarrow \boldsymbol{u}$ strongly in $\mathscr{L}^2(\mathbb{R}^d, \varrho)$ and therefore in $\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$. We claim that $\mathbf{t} + \varepsilon \boldsymbol{u}^m \in C_{\boldsymbol{\mu}}$ for $\varepsilon > 0$

sufficiently small. To prove this, let $N_{\mathbf{t}}$ be the null set of Remark 4.4. Then

$$\begin{aligned}&\Big\langle \big(\mathbf{t}(x_1,\xi_1)+\varepsilon\boldsymbol{u}^m(x_1)\big)-\big(\mathbf{t}(x_2,\xi_2)+\varepsilon\boldsymbol{u}^m(x_2)\big), x_1-x_2\Big\rangle\\ &\qquad=\langle \mathbf{t}(x_1,\xi_1)-\mathbf{t}(x_2,\xi_2), x_1-x_2\rangle-\varepsilon\langle \boldsymbol{u}^m(x_1)-\boldsymbol{u}^m(x_2), x_1-x_2\rangle\\ &\qquad\geqslant \Big(\alpha-\varepsilon\|D\boldsymbol{u}^m\|_{\mathscr{L}^\infty(\mathbb{R}^d)}\Big)|x_1-x_2|^2\geqslant 0\end{aligned}$$

for all $(x_i,\xi_i)\in\mathbb{R}^{2d}\setminus N_{\mathbf{t}}$ with $i=1..2$, for $\varepsilon>0$ small. This proves part (i.).

Step 2. We now show that for every $\mathbf{s}\in C_{\boldsymbol{\mu}}$ we also have $-\mathbf{s}\in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$. The argument is a modification of the proof of Proposition 4.28 of [48]. We first define the plan $\boldsymbol{\gamma}:=(\mathbb{x},\mathbf{s})\#\boldsymbol{\mu}\in C_\varrho$. Since $\operatorname{spt}\boldsymbol{\gamma}$ is a monotone subset of $\mathbb{R}^d\times\mathbb{R}^d$, there exists a maximal monotone extension of it, which we denote by Γ. Let u be the corresponding maximal monotone set-valued map, defined as

$$u(x):=\{y\in\mathbb{R}^d\colon (x,y)\in\Gamma\}\quad\text{for all } x\in\mathbb{R}^d.$$

It is well-known that for every $x\in\mathbb{R}^d$ the image $u(x)$ is a closed and convex subset of $\mathbb{R}^d$; see [1]. Consider the disintegration of the transport plan

$$\boldsymbol{\gamma}(dx,dy)=:\gamma_x(dy)\,\varrho(dx).$$

Then we have that $\gamma_x=\mathbf{s}(x,\cdot)\#\mu_x$ for ϱ-a.e. $x\in\mathbb{R}^d$, with $\boldsymbol{\mu}(dx,d\xi)=\mu_x(d\xi)\,\varrho(dx)$ the disintegration of $\boldsymbol{\mu}$. Let $\check\gamma_x:=(-\mathbf{s}(x,\cdot))\#\mu_x$ and $-\boldsymbol{\gamma}=(\mathbb{x},-\mathbf{s})\#\boldsymbol{\mu}$ so that

$$(-\boldsymbol{\gamma})(dx,dy)=\check\gamma_x(dy)\,\varrho(dx).$$

For ϱ-a.e. $x\in\mathbb{R}^d$ we denote by $A_x\subset u(x)$ the closed convex hull of $\operatorname{spt}\gamma_x$. For such an x there are two possibilities: either γ_x is a Dirac measure and $A_x=\{\mathrm{b}(\boldsymbol{\gamma})(x)\}$ (recall Definition 2.6), or A_x (and therefore $u(x)$) contains $\mathrm{b}(\boldsymbol{\gamma})(x)$ as an interior point with respect to the relative topology. In the latter case, the subspace

$$L_x:=\bigcup_{n\in\mathbb{N}} n\big(-\mathrm{b}(\boldsymbol{\gamma})(x)+u(x)\big)$$

has the property that $\operatorname{spt}\gamma_x\subset \mathrm{b}(\boldsymbol{\gamma})(x)+L_x$, and hence $\operatorname{spt}\check\gamma_x\subset -\mathrm{b}(\boldsymbol{\gamma})(x)+L_x$. Let $\mathbb{p}_x^n$ be the metric projection of $\mathbb{R}^d$ onto the closed convex set

$$-(n+1)\mathrm{b}(\boldsymbol{\gamma})(x)+nu(x) \tag{4.9}$$

for ϱ-a.e. $x\in\mathbb{R}^d$ and $n\in\mathbb{N}$. Since projections are contractions, we have

$$\begin{aligned}|\mathbb{p}_x^n(y)-y|&\leqslant \big|\mathbb{p}_x^n(y)-\mathbb{p}_x^n\big(-\mathrm{b}(\boldsymbol{\gamma})(x)\big)\big|+\big|\big(-\mathrm{b}(\boldsymbol{\gamma})(x)\big)-y\big|\\ &\leqslant 2\big|y-\big(-\mathrm{b}(\boldsymbol{\gamma})(x)\big)\big|\end{aligned}$$

for all $y\in\mathbb{R}^d$. We used that $-\mathrm{b}(\boldsymbol{\gamma})(x)$ is contained in (4.9). We have

$$\begin{aligned}&\int_{\mathbb{R}^d}\bigg(\int_{\mathbb{R}^d}\big|y-\big(-\mathrm{b}(\boldsymbol{\gamma})(x)\big)\big|^2\,\check\gamma_x(dy)\bigg)\,\varrho(dx)\\ &\qquad\leqslant 4\int_{\mathbb{R}^d}\bigg(\int_{\mathbb{R}^d}|y|^2\,\check\gamma_x(dy)\bigg)\,\varrho(dx)=4\int_{\mathbb{R}^{2d}}|\mathbf{s}(x,\xi)|^2\,\boldsymbol{\mu}(dx,d\xi),\end{aligned}$$

by definition of $\mathrm{b}(\boldsymbol{\gamma})(x)$ and Jensen's inequality. We now define the maps

$$\mathbf{s}^n(x,\xi):=\mathbb{p}_x^n\big(-\mathbf{s}(x,\xi)\big)\quad\text{for } \boldsymbol{\mu}\text{-a.e. } (x,\xi)\in\mathbb{R}^{2d}.$$

Using dominated convergence, we get for $n\to\infty$ that

$$\|\mathbf{s}^n-(-\mathbf{s})\|^2_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})}=\int_{\mathbb{R}^d}\bigg(\int_{\mathbb{R}^d}|\mathbb{p}_x^n(y)-y|^2\,\check\gamma_x(dy)\bigg)\,\varrho(dx)\longrightarrow 0, \tag{4.10}$$

because $\mathbb{p}_x^n(y) \longrightarrow y$ for all $y \in -\mathrm{b}(\boldsymbol{\gamma})(x) + L_x$ and for ϱ-a.e. $x \in \mathbb{R}^d$ such that γ_x is not a Dirac measure. We used again that $\mathbb{p}_x^n(-\mathrm{b}(\boldsymbol{\gamma})(x)) = -\mathrm{b}(\boldsymbol{\gamma})(x)$.

As discussed in Step 1, there exists a sequence of smooth, compactly supported functions $\boldsymbol{t}^m$ such that $\boldsymbol{t}^m \longrightarrow \mathrm{b}(\boldsymbol{\gamma})$ in $\mathscr{L}^2(\mathbb{R}^d, \varrho)$. We now define

$$\mathbf{s}^{n,m}(x,\xi) := \mathbf{s}^n(x,\xi) + (n+1)\big(\mathrm{b}(\boldsymbol{\gamma})(x) - \boldsymbol{t}^m(x)\big). \tag{4.11}$$

for $\boldsymbol{\mu}$-a.e. $(x,\xi) \in \mathbb{R}^{2d}$. We get for $m \to \infty$ (with n fixed) that

$$\|\mathbf{s}^{n,m} - \mathbf{s}^n\|^2_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})} = (n+1)^2 \int_{\mathbb{R}^d} |\mathrm{b}(\boldsymbol{\gamma})(x) - \boldsymbol{t}^m(x)|^2 \,\varrho(dx) \longrightarrow 0. \tag{4.12}$$

Combining (4.10) and (4.12), we find $\|\mathbf{s}^{n,m} - (-\mathbf{s})\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})} \longrightarrow 0$. We claim that $\mathbf{t} + \varepsilon\mathbf{s}^{n,m} \in C_{\boldsymbol{\mu}}$ for $\varepsilon > 0$ small. To prove this, we observe first that

$$\mathbf{s}^{n,m}(x,\xi) \subset -(n+1)\boldsymbol{t}^m(x) + nu(x) \quad \text{for } \boldsymbol{\mu}\text{-a.e. } (x,\xi) \in \mathbb{R}^{2d},$$

by definition of $\mathbb{p}_x^n$ and (4.11). With $N_{\mathbf{t}}$ the null set of Remark 4.4, we have

$$\begin{aligned}
&\Big\langle \big(\mathbf{t}(x_1,\xi_1) - \varepsilon(n+1)\boldsymbol{t}^m(x_1)\big) - \big(\mathbf{t}(x_2,\xi_2) - \varepsilon(n+1)\boldsymbol{t}^m(x_2)\big), x_1 - x_2 \Big\rangle \\
&\quad = \langle \mathbf{t}(x_1,\xi_1) - \mathbf{t}(x_1,\xi_1), x_1 - x_2 \rangle - \varepsilon(n+1)\langle \boldsymbol{t}^m(x_1) - \boldsymbol{t}^m(x_2), x_1 - x_2 \rangle \\
&\quad \geqslant \Big(\alpha - \varepsilon(n+1)\|D\boldsymbol{t}^m\|_{\mathscr{L}^\infty(\mathbb{R}^d)}\Big)|x_1 - x_2|^2 \geqslant 0
\end{aligned}$$

for all $(x_i,\xi_i) \in \mathbb{R}^{2d} \setminus N_{\mathbf{t}}$ with $i = 1..2$, for $\varepsilon > 0$ small. Since u is monotone, the support of $(\mathbb{x}, \mathbf{t} + \varepsilon\mathbf{s}^{n,m})\#\boldsymbol{\mu}$ is contained in a monotone subset of $\mathbb{R}^d \times \mathbb{R}^d$.

Step 3. We prove that if $\mathbf{v} \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ then also $-\mathbf{v} \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$. There exists a sequence of $\mathbf{v}^n \in \mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$ with $\|\mathbf{v}^n - \mathbf{v}\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})} \longrightarrow 0$ as $n \to \infty$, and such that $\mathbf{t} + \varepsilon^n\mathbf{v}^n \in C_{\boldsymbol{\mu}}$ for $\varepsilon^n > 0$ small. We have the following identity:

$$-\mathbf{v}^n = -\frac{1}{\varepsilon^n}(\mathbf{t} + \varepsilon^n\mathbf{v}^n) + \frac{1}{\varepsilon^n}\mathbf{t}.$$

The first term on the right-hand side is in $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ because of Step 2; the second one is in $C_{\boldsymbol{\mu}} \subset \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$. Since the tangent cone is a closed convex cone, we conclude that $-\mathbf{v}^n \in \operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$. Then we use that $\|(-\mathbf{v}^n) - (-\mathbf{v})\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})} \longrightarrow 0$. □

Remark 4.7. We emphasize that, unlike the tangent cone built from optimal transport maps/plans, which basically consists of *gradient* vector fields (see [4,48]), the tangent cone derived from monotone maps contains all of $\mathscr{L}^2(\mathbb{R}^d, \varrho)$ if $\mathbf{t}$ is strictly monotone in the sense of inequality(4.8). This condition is satisfied when $\mathbf{t} = \mathbb{x}$, for example. Generically, it can happen that the tangent cone is a proper subset of $\mathscr{L}^2(\mathbb{R}^d, \varrho)$: If $d = 1$ and $\mathbf{t} \in C_{\boldsymbol{\mu}}$ depends only on the spatial variable $x \in \mathbb{R}$ (so that $\mathbf{t} \in \mathscr{L}^2(\mathbb{R}, \varrho)$), then a velocity $\boldsymbol{v} \in \mathscr{L}^2(\mathbb{R}, \varrho)$ belongs to $\operatorname{Tan}_{\mathbf{t}} C_{\boldsymbol{\mu}}$ only if $\boldsymbol{v}$ is non-decreasing on each open interval on which $\mathbf{t}$ is constant; see Lemma 3.6 in [19] for more details.

4.2. **Minimization Problem.** We now introduce the main minimization problem for (1.3). Both mass and momentum will be conserved, by construction. But since transport maps $\mathbf{t} \in C_{\boldsymbol{\mu}}$ are not required to be strictly monotone (hence injective), it may happen that fluid elements with distinct velocities are transported to the same location. We will then use the barycentric projection to select an admissible velocity that are consistent with the monotonicity constraint. This results in fluid elements sticking together to form larger compounds.

Definition 4.8. For any $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, and any $\mathbf{t} \in C_{\boldsymbol{\mu}}$ let

$$\mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t}) := \Big\{ \boldsymbol{u} \circ \mathbf{t} \colon \boldsymbol{u} \in \mathscr{L}^2(\mathbb{R}^d, \varrho_{\mathbf{t}}) \Big\}, \quad \varrho_{\mathbf{t}} := \mathbf{t}\#\boldsymbol{\mu}.$$

One can check that $\mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t})$ is a closed subspace of $\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$ because

$$\|\boldsymbol{u} \circ \mathbf{t}\|_{\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})} = \|\boldsymbol{u}\|_{\mathscr{L}^2(\mathbb{R}^d,\varrho_{\mathbf{t}})}$$

for all $\boldsymbol{u} \circ \mathbf{t} \in \mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t})$; see Section 5.2 in [4]. Consequently, there exists an orthogonal projection onto this subspace, which we will denote by $\mathbb{p}_{\mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t})}$.

Definition 4.9 (Energy Minimization)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$, $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, and $\tau > 0$ be given. Then we consider the following three-step scheme:

(1) Compute the metric projection $\mathbf{t}_\tau := \mathbb{p}_{C_{\boldsymbol{\mu}}}(\mathbb{x} + \tau\mathbb{v})$ and define

$$\mathbf{w}_\tau(x,\xi) := V_\tau\big(x, \xi, \mathbf{t}_\tau(x,\xi)\big) \quad \text{for } \boldsymbol{\mu}\text{-a.e. } (x,\xi) \in \mathbb{R}^{2d}; \tag{4.13}$$

see Remark 4.3 and (1.16) for the definition of V_τ.

(2) Compute the orthogonal projection $\mathbf{u}_\tau := \mathbb{p}_{\mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t}_\tau)}(\mathbf{w}_\tau)$.

(3) Define the updated fluid state

$$\varrho_\tau := \mathbf{t}_\tau\#\boldsymbol{\mu}, \quad \boldsymbol{\mu}_\tau := (\mathbf{t}_\tau, \mathbf{u}_\tau)\#\boldsymbol{\mu}.$$

Notice that $\boldsymbol{\mu}_\tau$ is well-defined for any choice of $(\varrho, \boldsymbol{\mu}, \tau)$, and that $\mathbf{u}_\tau$ determines an Eulerian velocity field $\boldsymbol{u}_\tau \in \mathscr{L}^2(\mathbb{R}^d, \varrho_\tau)$ via $\mathbf{u}_\tau =: \boldsymbol{u}_\tau \circ \mathbf{t}_\tau$. We observe that $\boldsymbol{u}_\tau$ is just the barycentric projection of $\boldsymbol{\mu}_* := (\mathbf{t}_\tau, \mathbf{w}_\tau)\#\boldsymbol{\mu}$: Indeed we have

$$\int_{\mathbb{R}^{2d}} \big|\mathbf{w}_\tau(x,\xi) - \boldsymbol{u}\big(\mathbf{t}_\tau(x,\xi)\big)\big|^2 \,\boldsymbol{\mu}(dx,d\xi) = \int_{R^{2d}} |\zeta - \boldsymbol{u}(z,\zeta)|^2 \,\boldsymbol{\mu}_*(dz,d\zeta)$$

for all $\boldsymbol{u} \in \mathscr{L}^2(\mathbb{R}^d, \varrho_\tau)$, and the barycentric projection $\mathrm{b}(\boldsymbol{\mu}_*)$ is the unique element in $\mathscr{L}^2(\mathbb{R}^d, \varrho_\tau)$ closest to $\boldsymbol{\mu}_*$ with respect to $\mathscr{W}_{\varrho_\tau}$ (recall (2.2)). From Proposition 4.6, we deduce that $\mathscr{L}^2(\mathbb{R}^d, \varrho_\tau) \subset \mathrm{Tan}_{\mathbb{x}}\, \mathscr{C}_{\boldsymbol{\mu}_*}$. Step (2) of Definition 4.9 can therefore be interpreted as the projection of the updated state $\boldsymbol{\mu}_*$ onto (a subspace of) the tangent cone at the new configuration; see also Remark 4.11. A similar combination of transporting the vector field, then projecting it onto the tangent cone was used in [3] to construct the parallel transport along curves in $\mathscr{P}_2(\mathbb{R}^D)$; see also [11].

When ϱ_τ is absolutely continuous with respect to the Lebesgue measure so that there is no concentration (no sticking together of fluid elements), then the tangent cone at ϱ_τ consists only of monokinetic states, as follows from Remark 3.13.

Remark 4.10. We emphasize that the minimization of work links the transport $\mathbf{t}_\tau$ to the intermediate velocity $\mathbf{w}_\tau$ through the optimal velocity (4.13). This makes it possible to express the work equivalently in different form: Defining

$$\mathbf{v}_\tau(x,\xi) := \frac{\mathbf{t}_\tau(x,\xi) - x}{\tau} \quad \text{for } \boldsymbol{\mu}\text{-a.e. } (x,\xi) \in \mathbb{R}^{2d},$$

we have the following identities, which will be used frequently:

$$(x + \tau\xi) - \mathbf{t}_\tau(x,\xi) = \tau\big(\xi - \mathbf{v}_\tau(x,\xi)\big) = \frac{2\tau}{3}\big(\xi - \mathbf{w}_\tau(x,\xi)\big). \tag{4.14}$$

In particular, the transport velocity can be written as a convex combination

$$\mathbf{v}_\tau(x,\xi) = \frac{2}{3}\mathbf{w}_\tau(x,\xi) + \frac{1}{3}\xi \quad \Longleftrightarrow \quad \mathbf{w}_\tau(x,\xi) = \frac{3}{2}\mathbf{v}_\tau(x,\xi) - \frac{1}{2}\xi \tag{4.15}$$

for $\boldsymbol{\mu}$-a.e. $(x,\xi)\in\mathbb{R}^{2d}$. Inserting (4.14) into the work functional (3.3) with $z=\mathbf{t}(x,\xi)$ (recall that projections are non-splitting; see Section 3.4), we observe that the minimization over $C_{\boldsymbol{\mu}}$ can be reformulated as a minimization over velocities.

Remark 4.11. We will be mostly interested in situations where the initial state is *monokinetic*, for which $\boldsymbol{\mu}(dx,d\xi)=\delta_{\boldsymbol{u}(x)}(d\xi)\,\varrho(dx)$ for some velocity $\boldsymbol{u}\in\mathscr{L}^2(\mathbb{R}^d,\varrho)$. In this case, the transport maps in $C_{\boldsymbol{\mu}}$ are functions of $x\in\mathbb{R}^d$ alone because

$$\mathbf{t}(x,\xi):=\mathbf{t}\big(x,\boldsymbol{u}(x)\big)\quad\text{for }\boldsymbol{\mu}\text{-a.e. }(x,\xi)\in\mathbb{R}^{2d}.\tag{4.16}$$

For $d=1$, it was shown in [19, 65] that the family of transport maps

$$\mathbf{t}_s:=\mathbb{p}_{C_{\boldsymbol{\mu}}}(\mathbb{x}+s\mathbb{v})\quad\text{for all }s\geqslant 0,$$

determines a weak solution of the pressureless gas dynamics system (1.3) with initial data $(\varrho,\boldsymbol{u})$ in the following way: The density at time s is given by the push-forward $\varrho_s:=\mathbf{t}_s\#\boldsymbol{\mu}$. The (Lagrangian) velocity is defined by the formula

$$\mathbf{v}_s(x,\xi):=\lim_{h\to 0+}\frac{\mathbf{t}_{s+h}(x,\xi)-\mathbf{t}_s(x,\xi)}{h}\quad\text{for }\boldsymbol{\mu}\text{-a.e. }(x,\xi)\in\mathbb{R}^{2d}.$$

Again this is a function of $x\in\mathbb{R}^d$ alone because of (4.16). Using (4.7), we observe that $\mathbf{v}_s$ is, in fact, the metric projection of the *initial velocity* $\boldsymbol{u}$ onto the tangent cone $\operatorname{Tan}_{\mathbf{t}_s}C_{\boldsymbol{\mu}}$. Here we used that $\xi=\boldsymbol{u}(x)$ for $\boldsymbol{\mu}$-a.e. $(x,\xi)\in\mathbb{R}^{2d}$. Moreover, since the map $s\mapsto\mathbf{t}_s$ is Lipschitz continuous in $\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$ and therefore differentiable for a.e. $s\in\mathbb{R}$, we conclude for such s that $\mathbf{v}_s$ can also be obtained as

$$\mathbf{v}_s(x,\xi)=\lim_{h\to 0+}\frac{\mathbf{t}_{s-h}(x,\xi)-\mathbf{t}_s(x,\xi)}{-h}\quad\text{for }\boldsymbol{\mu}\text{-a.e. }(x,\xi)\in\mathbb{R}^{2d}.$$

and so $\mathbf{v}_s\in-\operatorname{Tan}_{\mathbf{t}_s}C_{\boldsymbol{\mu}}$ as well. This implies that

$$\mathbf{v}_s\in\mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t}_s)\quad\text{for a.e. }s\in\mathbb{R},$$

and, in fact, $\mathbf{v}_s$ is the orthogonal projection of the initial velocity onto this subspace; see [19, 65]. Our minimization problem preserves this structure: Note that

$$\begin{aligned}&\int_{\mathbb{R}^2}\Big\langle\xi-\boldsymbol{u}_\tau\big(\mathbf{t}_\tau(x,\xi)\big),\boldsymbol{v}\big(\mathbf{t}_\tau(x,\xi)\big)\Big\rangle\,\boldsymbol{\mu}(dx,d\xi)\\&\qquad=\int_{\mathbb{R}^2}\Big\langle\xi-\mathbf{w}_\tau(x,\xi),\boldsymbol{v}\big(\mathbf{t}_\tau(x,\xi)\big)\Big\rangle\,\boldsymbol{\mu}(dx,d\xi)\end{aligned}$$

for all $\boldsymbol{v}\in\mathscr{D}(\mathbb{R})$. Using the identity (4.15) for $\mathbf{w}_\tau$, for any $\alpha>0$ we can write

$$\begin{aligned}&\int_{\mathbb{R}^2}\Big\langle\xi-\mathbf{w}_\tau(x,\xi),\boldsymbol{v}\big(\mathbf{t}_\tau(x,\xi)\big)\Big\rangle\,\boldsymbol{\mu}(dx,d\xi)\\&\qquad=\frac{3}{2\tau}\int_{\mathbb{R}^2}\Big\langle(x+\tau\xi)-\mathbf{t}_\tau(x,\xi),(\boldsymbol{v}+\alpha\,\mathrm{id})\circ\mathbf{t}_\tau(x,\xi)\Big\rangle\,\boldsymbol{\mu}(dx,d\xi)\\&\qquad\quad-\frac{3\alpha}{2\tau}\int_{\mathbb{R}^2}\big\langle(x+\tau\xi)-\mathbf{t}_\tau(x,\xi),\mathbf{t}_\tau(x,\xi)\big\rangle\,\boldsymbol{\mu}(dx,d\xi).\end{aligned}$$

The last integral vanishes because of (4.4) in Remark 4.3. Since for large enough α the map $\boldsymbol{v}+\alpha\,\mathrm{id}$ is strictly increasing, one can check that $(\mathrm{id},(\boldsymbol{v}+\alpha\,\mathrm{id})\circ\mathbf{t}_\tau)\#\boldsymbol{\mu}\in C_{\boldsymbol{\mu}}$. Here we used the assumption that $d=1$ (since compositions of monotone maps are again monotone in one space dimension). Using inequality (4.5), we have

$$\int_{\mathbb{R}^2}\Big\langle\xi-\mathbf{w}_\tau(x,\xi),\boldsymbol{v}\big(\mathbf{t}_\tau(x,\xi)\big)\Big\rangle\,\boldsymbol{\mu}(dx,d\xi)\leqslant 0\quad\text{for all }\boldsymbol{v}\in\mathscr{D}(\mathbb{R}),$$

which in particular implies equality. Since $\mathscr{D}(\mathbb{R})$ is dense in $\mathscr{L}^2(\mathbb{R}, \varrho_\tau)$ we get

$$\int_{\mathbb{R}} \big\langle \boldsymbol{u}(x) - \boldsymbol{u}_\tau\big(\boldsymbol{t}_\tau(x)\big), \boldsymbol{v}\big(\boldsymbol{t}_\tau(x)\big)\big\rangle \,\varrho(dx) = 0 \quad \text{for all } \boldsymbol{v} \in \mathscr{L}^2(\mathbb{R}, \varrho_\tau),$$

writing $\boldsymbol{t}_\tau := \mathbf{t}_\tau(x, \boldsymbol{u}(x))$ for ϱ-a.e. $x \in \mathbb{R}^d$. As explained above, this orthogonality, in combination with the definition of $\boldsymbol{t}_\tau$ as the metric projection of $\mathrm{id} + \tau\boldsymbol{u}$ onto monotone maps, characterizes solutions of (1.3) satisfying a stickyness condition, for a.e. $\tau > 0$. So our discretization already generates the *exact* solution, not just an approximations, for $d = 1$ and monokinetic initial state $\boldsymbol{\mu} = (\mathrm{id}, \boldsymbol{u})\#\varrho$.

Remark 4.12. In [14] the authors prove the non-existence of sticky particle solutions to (1.3) for well-designed initial data. In their construction the number of collisions grows unboundedly the closer one gets to the initial time, and so the dynamics has arbitrarily small time scales. Using our discretization, we can construct a sequence of approximate solutions to (1.3) starting from the initial data in [14]. We will show below that this approximation converges to a measure-valued solution of (1.3). The timestep $\tau > 0$ in our discretization introduces a minimal time scale below which the dynamics is not completely resolved but is "smeared out." It would be interesting to know to which solution our discretization converges in the limit $\tau \to 0$.

Remark 4.13. The constant map $\mathbf{s}(x, \xi) = \pm b$ for all $(x, \xi) \in \mathbb{R}^{2d}$, where $b \in \mathbb{R}^d$ is some vector, is an element of $C_{\boldsymbol{\mu}}$. Using this function in (4.5), we get

$$\begin{aligned}\int_{\mathbb{R}^{2d}} \langle b, \zeta\rangle \,\boldsymbol{\mu}_\tau(dz, d\zeta) &= \int_{\mathbb{R}^{2d}} \langle b, \mathbf{u}_\tau(x, \xi)\rangle \,\boldsymbol{\mu}(dx, d\xi) \\ &= \int_{\mathbb{R}^{2d}} \langle b, \mathbf{w}_\tau(x, \xi)\rangle \,\boldsymbol{\mu}(dx, d\xi) = \int_{\mathbb{R}^{2d}} \langle b, \xi\rangle \,\boldsymbol{\mu}(dx, d\xi);\end{aligned}$$

see (4.15). Recall that $\mathbf{u}_\tau$ is the orthogonal projection of $\mathbf{w}_\tau$ onto $\mathscr{H}_{\boldsymbol{\mu}}(\mathbf{t})$, which contains $\mathbf{s}$. We conclude that the minimization preserves the total momentum.

Similarly, we can use the test functions $\mathbf{s}(x, \xi) = \pm Ax$ for all $(x, \xi) \in \mathbb{R}^{2d}$ with $A \in \mathrm{Skew}_d(\mathbb{R})$ in (4.5) because these functions are monotone. It follows that

$$\int_{\mathbb{R}^{2d}} \langle \mathbf{w}_\tau(x, \xi), Ax\rangle \,\boldsymbol{\mu}(dx, d\xi) = \int_{\mathbb{R}^{2d}} \langle \xi, Ax\rangle \boldsymbol{\mu}(dx, d\xi);$$

recall (4.14). We decompose the left-hand side using $x = \mathbf{t}_\tau(x, \xi) - \tau\mathbf{v}_\tau(x, \xi)$. The corresponding second integral can be estimated as

$$\left| -\tau \int_{\mathbb{R}^{2d}} \langle \mathbf{w}_\tau(x, \xi), A\mathbf{v}_\tau(x, \xi)\rangle \,\boldsymbol{\mu}(dx, d\xi)\right| \leqslant \tau\|A\| \|\mathbf{w}_\tau\|^{1/2}_{\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu}))} \|\mathbf{v}_\tau\|^{1/2}_{\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu}))}.$$

We will see below that both $\mathscr{L}^2(\mathbb{R}^{2d}, \boldsymbol{\mu})$-norms can be estimated against the kinetic energy $\int_{\mathbb{R}^{2d}} |\xi|^2 \,\boldsymbol{\mu}(dx, d\xi)$ of the initial state, uniformly in τ. Moreover, we get

$$\begin{aligned}\int_{\mathbb{R}^{2d}} \langle \mathbf{w}_\tau(x, \xi), A\mathbf{t}_\tau(x, \xi)\rangle \,\boldsymbol{\mu}(dx, d\xi) &= \int_{\mathbb{R}^{2d}} \langle \zeta, Az\rangle \,\boldsymbol{\mu}_*(dz, d\zeta) \\ &= \int_{\mathbb{R}^d} \langle \boldsymbol{u}_\tau(z), Az\rangle \,\varrho_\tau(dz),\end{aligned}$$

where we have used that $\boldsymbol{u}_\tau$ is the barycentric projection of $\boldsymbol{\mu}_*$. We conclude that our minimization preserves total angular momentum up to order τ.

4.3. **Polar Cone.** In this section, we will give a representation of the elements in the polar cone of $C_{\boldsymbol{\mu}}$. As we will see later, such elements appear as stress tensors. In Remark 1.7, we have defined the space $\mathscr{C}_*(\mathbb{R}^d;\mathbb{R}^D)$ of all continuous functions $f\colon \mathbb{R}^d \longrightarrow \mathbb{R}^D$ for which $\lim_{|x|\to\infty} f(x) \in \mathbb{R}^D$ exists. We identify $\mathscr{C}_*(\mathbb{R}^d;\mathbb{R}^D)$ with the space $\mathscr{C}(\dot{\mathbb{R}}^d;\mathbb{R}^D)$ of continuous functions on the one-point compactification $\dot{\mathbb{R}}^d$ of $\mathbb{R}^d$: We adjoin to $\mathbb{R}^d$ a point ∞ and define a distance (see [61])

$$d(x,y) := \begin{cases} \min\{|x-y|, h(x)+h(y)\} & \text{if } x,y \in \mathbb{R}^d, \\ h(x) & \text{if } x \in \mathbb{R}^d \text{ and } y = \infty, \\ 0 & \text{if } x,y = \infty, \end{cases}$$

where $h(x) := 1/(1+|x|)$ for all $x \in \mathbb{R}^d$. Then $|x| \to \infty$ is equivalent to $d(x,\infty) \to 0$. To any $g \in \mathscr{C}_*(\mathbb{R}^d;\mathbb{R}^D)$ we associate $\dot{g} \in \mathscr{C}(\dot{\mathbb{R}}^d;\mathbb{R}^D)$ defined as

$$\dot{g}(x) := \begin{cases} g(x) & \text{if } x \in \mathbb{R}^d, \\ \lim_{|x|\to\infty} g(x) & \text{if } x = \infty. \end{cases}$$

Conversely, the restriction of any function in $\mathscr{C}(\dot{\mathbb{R}}^d;\mathbb{R}^D)$ to $\mathbb{R}^d$ induces a function in $\mathscr{C}_*(\mathbb{R}^d;\mathbb{R}^D)$. We will hence not distinguish between the two spaces. Similarly, we define $\mathscr{C}_*(\mathbb{R}^d;\mathrm{Mat}_l(\mathbb{R}))$, $\mathscr{C}_*(\mathbb{R}^d;\mathrm{Sym}_l(\mathbb{R}))$, and $\mathscr{C}_*(\mathbb{R}^d;\mathrm{Sym}_l(\mathbb{R},\geqslant))$.

For any $u \in \mathscr{C}^1(\mathbb{R}^d;\mathbb{R}^d)$ we refer to the symmetric part $\nabla u(x)^{\mathrm{sym}}$ for all $x \in \mathbb{R}^d$ as its deformation tensor, which is an element of $\mathscr{C}(\mathbb{R}^d;\mathrm{Sym}_d(\mathbb{R}))$. Let

$$\mathscr{C}^1_*(\mathbb{R}^d;\mathbb{R}^d) := \big\{ u \in \mathscr{C}^1(\mathbb{R}^d;\mathbb{R}^d)\colon \nabla u \in \mathscr{C}_*\big(\mathbb{R}^d;\mathrm{Mat}_d(\mathbb{R})\big)\big\},$$
$$\mathrm{Mon}(\mathbb{R}^d) := \big\{ u \in \mathscr{C}^1_*(\mathbb{R}^d;\mathbb{R}^d)\colon u \text{ is monotone}\big\}.$$

The cone $\mathrm{Mon}(\mathbb{R}^d)$ contains, in particular, all linear maps $u(x) := Ax$ for all $x \in \mathbb{R}^d$, with $A \in \mathrm{Mat}_d(\mathbb{R},\geqslant)$. We will use the following result from [20]:

Theorem 4.14 (Stress Tensor)**.** *Assume that there exist a measure $\mathbf{F} \in \mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)$ with finite first moment and a measure $\mathbf{P} \in \mathscr{M}(\mathbb{R}^d;\mathrm{Sym}_d(\mathbb{R},\geqslant))$ with*

$$G(u) := -\int_{\mathbb{R}^d} \langle u(x), \mathbf{F}(dx)\rangle - \int_{\mathbb{R}^d} \mathrm{tr}\big(\nabla u(x)\mathbf{P}(dx)\big) \geqslant 0 \tag{4.17}$$

for all $u \in \mathrm{Mon}(\mathbb{R}^d)$. Then there exists $\mathbf{R} \in \mathscr{M}(\dot{\mathbb{R}}^d;\mathrm{Sym}_d(\mathbb{R},\geqslant))$ such that

$$G(u) = \int_{\dot{\mathbb{R}}^d} \mathrm{tr}\big(\nabla u(x)\mathbf{R}(dx)\big) \quad \textit{for all } u \in \mathscr{C}^1_*(\mathbb{R}^d;\mathbb{R}^d),$$
$$\int_{\dot{\mathbb{R}}^d} \mathrm{tr}\big(\mathbf{R}(dx)\big) = -\int_{\mathbb{R}^d} \langle x, \mathbf{F}(dx)\rangle - \int_{\mathbb{R}^d} \mathrm{tr}\big(\mathbf{P}(dx)\big). \tag{4.18}$$

Notice that the integral in (4.17) is finite for any choice of $u \in \mathscr{C}^1_*(\mathbb{R}^d;\mathbb{R}^d)$ since the first moment of $\mathbf{F}$ is finite, by assumption. Recall that the trace of a symmetric matrix equals the sum of its eigenvalues, which in the case of a positive semidefinite matrix are all non-negative. Therefore (4.18) controls the size of $\mathbf{R}$.

Remark 4.15. The stress tensor $\mathbf{R}$ does not actually assign any mass to the remainder $\dot{\mathbb{R}}^d \setminus \mathbb{R}^d$, so Theorem 4.14 remains true if the compactification $\dot{\mathbb{R}}^d$ is replaced by $\mathbb{R}^d$. In fact, recall that $\mathbb{R}^d$ (being a separable metric space) is a Radon space, so that any finite Borel measure is inner regular. Consider a non-negative, radially symmetric test function $\varphi \in \mathscr{D}(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} \varphi(x)\,dx = 1$ and define

$$u_R := \nabla(\phi_R \star \varphi), \quad \text{with} \quad \phi_R(x) := \tfrac{1}{2}\max\{|x|^2 - R^2, 0\}$$

for $x \in \mathbb{R}^d$ and $R > 0$. The map $\phi_R \star \varphi$ is convex and smooth (since the convolution preserves convexity), hence u_R is monotone and smooth. Notice that $u_R(x) = 0$ for all $|x| \leqslant R - c$, with c the (finite) diameter of $\operatorname{spt}\varphi$. Moreover, we have

$$\int_{\mathbb{R}^d} \varphi(x-y)|y|^2\,dy = |x|^2 + \bigg(\int_{\mathbb{R}^d} |z|^2 \varphi(z)\,dz \bigg)$$

for all $x \in \mathbb{R}^d$, which implies that $u_R(x) = x$ and $Du_R(x) = \mathbb{1}$ for $|x| \geqslant R + c$. In particular, we observe that $u_R \in \operatorname{Mon}(\mathbb{R}^d)$ for all $R > 0$. Then

$$\int_{|x|\geqslant R+c} \operatorname{tr}\big(\mathbf{R}(dx)\big) \leqslant C\bigg(\int_{|x|\geqslant R-c} |x|\,|\mathbf{F}(dx)| + \int_{|x|\geqslant R-c} \operatorname{tr}\big(\mathbf{P}(dx)\big)\bigg), \qquad (4.19)$$

with C some finite constant depending on φ. The right-hand side of (4.19) converges to zero as $R \to \infty$ since both measures $|\mathbf{F}|$ and $\operatorname{tr}(\mathbf{P})$ are inner regular and the first moment of $\mathbf{F}$ is finite. We conclude that $\operatorname{tr}(\mathbf{R})(\dot{\mathbb{R}}^d \setminus \mathbb{R}^d) = 0$.

For $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $\tau > 0$ let $\mathbf{t}_\tau$ be given by Definition 4.9. Then

$$-\frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \mathbf{t}_\tau(x,\xi), \mathbf{s}(x,\xi)\rangle\, \boldsymbol{\mu}(dx,d\xi) \geqslant 0 \quad \text{for all } \mathbf{s} \in C_{\boldsymbol{\mu}},$$

which is (4.5). In particular, this inequality is true for $\mathbf{s} = u$ with $u \in \operatorname{Mon}(\mathbb{R}^d)$. Functions in $\operatorname{Mon}(\mathbb{R}^d)$ have at most linear growth and are therefore in $\mathscr{L}^2(\mathbb{R}^d, \varrho)$. Applying Theorem 4.14 (with $\mathbf{P} \equiv 0$), we get $\mathbf{R}_\tau \in \mathscr{M}(\mathbb{R}^d; \operatorname{Sym}_d(\mathbb{R}, \geqslant))$ with

$$\int_{\mathbb{R}^d} \operatorname{tr}\big(\nabla u(x)\mathbf{R}_\tau(dx)\big) = -\frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \mathbf{t}_\tau(x,\xi), u(x)\rangle\, \boldsymbol{\mu}(dx,d\xi), \qquad (4.20)$$

$$\int_{\mathbb{R}^d} \operatorname{tr}\big(\mathbf{R}_\tau(dx)\big) = -\frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \mathbf{t}_\tau(x,\xi), x\rangle\, \boldsymbol{\mu}(dx,d\xi); \qquad (4.21)$$

see Remark 4.15. The representation in Theorem 4.14 generalizes a similar description of the polar cone of monotone maps obtained in [65] in one space dimension. Using the identity (4.15), we obtain the following identities:

$$\begin{aligned} \int_{\mathbb{R}^d} \operatorname{tr}\big(\nabla u(x)\mathbf{R}_\tau(dx)\big) &= -\frac{3}{2\tau} \int_{\mathbb{R}^{2d}} \langle \xi - \mathbf{v}_\tau(x,\xi), u(x)\rangle\, \boldsymbol{\mu}(dx,d\xi) \\ &= -\frac{1}{\tau} \int_{\mathbb{R}^{2d}} \langle \xi - \mathbf{w}_\tau(x,\xi), u(x)\rangle\, \boldsymbol{\mu}(dx,d\xi), \end{aligned}$$

with transport velocity $\mathbf{v}_\tau(x,\xi) := (\mathbf{t}_\tau(x,\xi) - x)/\tau$ for $\boldsymbol{\mu}$-a.e. $(x,\xi) \in \mathbb{R}^{2d}$, and

$$\begin{aligned} \int_{\mathbb{R}^d} \operatorname{tr}\big(\mathbf{R}_\tau(dx)\big) &= -\frac{3}{2\tau} \int_{\mathbb{R}^{2d}} \langle \xi - \mathbf{v}_\tau(x,\xi), x\rangle\, \boldsymbol{\mu}(dx,d\xi) \\ &= -\frac{1}{\tau} \int_{\mathbb{R}^{2d}} \langle \xi - \mathbf{w}_\tau(x,\xi), x\rangle\, \boldsymbol{\mu}(dx,d\xi). \end{aligned}$$

Remark 4.16. In order to explore the significance of $\mathbf{R}_\tau$, we consider

$$\boldsymbol{\mu}(dx,d\xi) = \tfrac{1}{4}\delta_0(d\xi)\,\mathcal{L}^1|_{(-1,1)}(dx) + \tfrac{1}{2}\delta_1(d\xi)\,\delta_0(dx).$$

For any $\tau > 0$ the support of the transport plan $(\mathbb{x}, \mathbb{x} + \tau\mathbb{v})\#\boldsymbol{\mu}$ is not monotone. Then $\boldsymbol{\gamma}_\tau := (\mathbb{x}, \mathbf{t}_\tau)\#\boldsymbol{\mu}$, with $\mathbf{t}_\tau$ given by Definition 4.9, can be computed as

$$\begin{aligned} \boldsymbol{\gamma}_\tau(dx,dy) = \tfrac{1}{4}\Big(&\delta_{\beta(\tau)\tau}(dy)\,\mathcal{L}^1|_{[0,\beta(\tau)\tau]}(dx) + \delta_x(dy)\,\mathcal{L}^1|_{(-1,1)\setminus[0,\beta(\tau)\tau]}(dx)\Big) \\ &+ \tfrac{1}{2}\delta_{\beta(\tau)\tau}(dy)\,\delta_0(dx), \end{aligned}$$

where $\beta(\tau) \in [0,1]$ is the minimizer of the following function:

$$\varphi_\tau(\beta) := \tfrac{1}{2}|1-\beta|^2\tau^2 + \tfrac{1}{4}\int_0^{\beta\tau} |\beta\tau - x|^2\,dx,$$

which represents the $\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$-distance of $\mathbb{x}+\tau\mathbb{v}$ to some map in $C_{\boldsymbol{\mu}}$ parameterized by β. One can check that $\beta(\tau) := \frac{2}{\tau}(\sqrt{1+\tau}-1)$ and $\beta(\tau) \longrightarrow 1$ as $\tau \to 0$. The induced velocity distribution $\boldsymbol{\mu}_\tau := (\mathbb{x},(\mathbb{y}-\mathbb{x})/\tau)\#\boldsymbol{\gamma}_\tau$ equals

$$\begin{aligned}\boldsymbol{\mu}_\tau(dx,d\xi) = \tfrac{1}{4}\Big(&\delta_{\beta(\tau)-x/\tau}(d\xi)\,\mathcal{L}^1|_{[0,\beta(\tau)\tau]}(dx) + \delta_0(d\xi)\,\mathcal{L}^1|_{(-1,1)\setminus[0,\beta(\tau)\tau]}(dx)\Big)\\ &+ \tfrac{1}{2}\delta_{\beta(\tau)}(d\xi)\,\delta_0(dx).\end{aligned}$$

The first ξ-moments of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_\tau$ determine the corresponding momenta:

$$\begin{aligned}\boldsymbol{m}(dx) &:= \tfrac{1}{2}\delta_0(dx),\\ \boldsymbol{m}_\tau(dx) &:= \tfrac{1}{4}\Big(\beta(\tau) - \frac{x}{\tau}\Big)\,\mathcal{L}^1|_{[0,\beta(\tau)\tau]}(dx) + \tfrac{1}{2}\beta(\tau)\,\delta_0(dx).\end{aligned}$$

Therefore the change in momentum (which represents an acceleration) has two parts: The velocity of the fluid element with mass $1/2$ located at $x=0$ decreases, so the momentum is getting smaller. This momentum is *transfered* to fluid elements in the interval $[0,\beta(\tau)\tau]$, which pick up speed. The transfer is described by the derivative of the non-negative measure from Theorem 4.14. Let $\mathbf{R}_\tau := R_\tau\mathcal{L}^1$ with

$$R_\tau(x) := \begin{cases}\tfrac{1}{2}\big(1-\beta(\tau)\big) - \tfrac{1}{4}\Big(\beta(\tau)x - \dfrac{x^2}{2\tau}\Big) & \text{if } x \in [0,\beta(\tau)\tau],\\ 0 & \text{otherwise.}\end{cases}$$

Since $\frac{1}{2}(1-\beta(\tau)) = \frac{1}{8}\beta(\tau)^2\tau$ the measure $\mathbf{R}_\tau$ is non-negative, supported in $[0,\beta(\tau)\tau]$, and it satisfies $\boldsymbol{m}-\boldsymbol{m}_\tau = \partial_x\mathbf{R}_\tau$ in $\mathscr{D}'(\mathbb{R})$. Note further that R_τ vanishes as $\tau\to 0$, in any $\mathscr{L}^p(\mathbb{R})$ with $1 \leqslant p < \infty$. Theorem 4.14 suggests that a similar structure can be found in higher space dimensions: the metric projection onto C_ϱ may cause the transfer of momentum to neighboring fluid elements, captured by the distributional divergence $\nabla\cdot\mathbf{R}_\tau$ of the stress tensor field $\mathbf{R}_\tau$. This transfer manifests itself also in the kinetic energy balance, which we will consider next.

Proposition 4.17 (Energy Balance)**.** *For any $(\varrho,\boldsymbol{\mu},\tau)$ as in Definition 4.9 consider the quantities $(\mathbf{t}_\tau,\mathbf{w}_\tau,\mathbf{u}_\tau,\boldsymbol{\mu}_\tau)$ specified there. Let $\mathbf{R}_\tau \in \mathscr{M}(\mathbb{R}^d;\mathrm{Sym}_d(\mathbb{R},\geqslant))$ be the stress tensor field satisfying* (4.20)/(4.21)*. Then we have*

$$\begin{aligned}\mathcal{E}[\boldsymbol{\mu}_\tau] + \int_{\mathbb{R}^{2d}} \Big(\tfrac{1}{6}|\mathbf{w}_\tau - \xi|^2 + \tfrac{1}{2}|\mathbf{u}_\tau - \mathbf{w}_\tau|^2\Big)\,\boldsymbol{\mu}(dx,d\xi)&\\ + \int_{\mathbb{R}^d} \mathrm{tr}\big(\mathbf{R}_\tau(dx)\big) = \mathcal{E}[\boldsymbol{\mu}]&,\end{aligned}$$

with total/kinetic energy

$$\mathcal{E}[\boldsymbol{\mu}] := \int_{\mathbb{R}^{2d}} \tfrac{1}{2}|\xi|^2\boldsymbol{\mu}(dx,d\xi).$$

Recall that $\boldsymbol{\mu}_\tau = (\mathrm{id},\boldsymbol{u}_\tau)\#\varrho_\tau$ for some $\boldsymbol{u}_\tau \in \mathscr{L}^2(\mathbb{R}^d,\varrho_\tau)$ with $\varrho_\tau = \mathbf{t}_\tau\#\boldsymbol{\mu}$.

Proof. Since $\mathbf{u}_\tau := \mathbb{p}_{\mathscr{H}_{\boldsymbol{\mu}(\mathbf{t}_\tau)}}(\mathbf{w}_\tau)$ (orthogonal projection), we have

$$\int_{\mathbb{R}^{2d}} \tfrac{1}{2}|\mathbf{u}_\tau(x,\xi)|^2\,\boldsymbol{\mu}(dx,d\xi) + \int_{\mathbb{R}^{2d}} \tfrac{1}{2}|\mathbf{u}_\tau(x,\xi) - \mathbf{w}_\tau(x,\xi)|^2\,\boldsymbol{\mu}(dx,d\xi)$$
$$= \int_{\mathbb{R}^{2d}} \tfrac{1}{2}|\mathbf{w}_\tau(x,\xi)|^2\,\boldsymbol{\mu}(dx,d\xi).$$

On the other hand, using definition (4.13) of $\mathbf{w}_\tau$ and (4.14) we can write

$$\int_{\mathbb{R}^{2d}} \tfrac{1}{2}|\mathbf{w}_\tau(x,\xi)|^2\,\boldsymbol{\mu}(dx,d\xi) + \frac{1}{6}\int_{\mathbb{R}^{2d}} |\mathbf{w}_\tau(x,\xi) - \xi|^2\,\boldsymbol{\mu}(dx,d\xi)$$
$$= \int_{\mathbb{R}^{2d}} \tfrac{1}{2}|\xi|^2\,\boldsymbol{\mu}(dx,d\xi) - \frac{3}{2\tau^2}\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \mathbf{t}_\tau(x,\xi), \mathbf{t}_\tau(x,\xi)\rangle\,\boldsymbol{\mu}(dx,d\xi)$$
$$+ \frac{3}{2\tau^2}\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \mathbf{t}_\tau(x,\xi), x\rangle\,\boldsymbol{\mu}(dx,d\xi) \tag{4.22}$$

The second integral on the right-hand side of (4.22) vanishes because of (4.4), the last one can be expressed in terms of the stress tensor field $\mathbf{R}_\tau$; see (4.21). □

5. Energy Minimization: Polytropic Gases

We now modify the minimization problem of Section 4.2 for polytropic gases. In this case, the density ϱ must be absolutely continuous with respect to the Lebesgue measure since otherwise the internal energy would be infinite (see Definition 5.10). We need a lower semicontinuity result for the internal energy, suitably redefined as a convex functional on the set of monotone transports.

5.1. **Gradient Young Measures.**

We introduce Young measures to capture oscillations and concentrations of weak* converging sequences of derivatives of functions of bounded variations. They will be used in Section 5.2 to establish a lower semicontinuity result for the internal energy. We follow the presentation of [55, 69].

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and $\boldsymbol{t} \in \mathrm{BV}(\Omega;\mathbb{R}^d)$. Let B_d be the open unit ball in $\mathrm{Mat}_d(\mathbb{R})$ and ∂B_d its boundary. We associate to the derivative $D\boldsymbol{t}$ (which is a measure) a triple $\upsilon = (\nu,\sigma,\mu)$ with

$$\nu \in \mathscr{L}^\infty_{\mathrm{w}}\Big(\Omega;\mathscr{P}\big(\mathrm{Mat}_d(\mathbb{R})\big)\Big), \quad \sigma \in \mathscr{M}_+(\bar\Omega), \quad \mu \in \mathscr{L}^\infty_{\mathrm{w}}\big(\bar\Omega,\sigma;\mathscr{P}(\partial B_d)\big) \tag{5.1}$$

as follows: Consider the Lebesgue-Radon-Nikodým decomposition

$$D\boldsymbol{t} = \nabla\boldsymbol{t}\,\mathcal{L}^d + D^s\boldsymbol{t}, \quad D^s\boldsymbol{t} \perp \mathcal{L}^d, \tag{5.2}$$

and define $\nu_x := \delta_{\nabla\boldsymbol{t}(x)}$ for a.e. $x \in \Omega$ and $\sigma := |D^s\boldsymbol{t}|$. Let further

$$D^s\boldsymbol{t} = \frac{\mathrm{d}D^s\boldsymbol{t}}{\mathrm{d}|D^s\boldsymbol{t}|}\,|D^s\boldsymbol{t}|, \quad p := \frac{\mathrm{d}D^s\boldsymbol{t}}{\mathrm{d}|D^s\boldsymbol{t}|} \in \mathscr{L}^1(\Omega,|D^s\boldsymbol{t}|;\partial B_d).$$

be the polar decomposition of $D^s\boldsymbol{t}$ and define $\mu_x = \delta_{p(x)}$ for $|D^s\boldsymbol{t}|$-a.e. $x \in \Omega$. Here $\mathscr{L}^\infty_{\mathrm{w}}(\Omega;\mathscr{P}(\mathrm{Mat}_d(\mathbb{R})))$ is the space of weakly measurable maps from Ω into the space of probability measures on $\mathrm{Mat}_d(\mathbb{R})$ (similar definition for $\mathscr{L}^\infty_{\mathrm{w}}(\bar\Omega,\sigma;\mathscr{P}(\partial B_d))$). We call $\upsilon = (\nu,\sigma,\mu)$ an elementary Young measure associated to Du.

Consider now a sequence of uniformly bounded maps $\boldsymbol{t}^k \in \mathrm{BV}(\Omega;\mathbb{R}^d)$. Extracting a subsequence, we may assume that $\boldsymbol{t}^k \longrightarrow \boldsymbol{t}$ in $\mathscr{L}^1(\Omega;\mathbb{R}^d)$ and $D\boldsymbol{t}^k \rightharpoonup D\boldsymbol{t}$ weak* in $\mathscr{M}(\Omega;\mathrm{Mat}_d(\mathbb{R}))$, for some $\boldsymbol{t} \in \mathrm{BV}(\Omega;\mathbb{R}^d)$. In this case, we say that $\boldsymbol{t}^k$ converges weak* to $\boldsymbol{t}$ in $\mathrm{BV}(\Omega;\mathbb{R}^d)$. We denote by $\upsilon^k = (\nu^k,\sigma^k,\mu^k)$ the elementary Young measure associated to $D\boldsymbol{t}^k$ as above. Since the spaces in (5.1) are contained in the dual spaces to $\mathscr{L}^1(\Omega;\mathscr{C}_0(\mathrm{Mat}_d(\mathbb{R})))$, $\mathscr{C}(\bar\Omega)$, and $\mathscr{L}^1(\bar\Omega,\sigma;\mathscr{C}(\partial B_d))$ respectively, one can show that there exists a subsequence (which we do not relabel, for simplicity) and a triple $\upsilon=(\nu,\sigma,\mu)$ as in (5.1) with the property that the

$$\begin{aligned}\llbracket f,\upsilon^k\rrbracket &:= \int_\Omega [f(x,\cdot),\nu_x^k]\,dx + \int_{\bar\Omega}[f^\infty(x,\cdot),\mu_x^k]\,\sigma^k(dx) \\ &:= \int_\Omega\int_{\mathrm{Mat}_d(\mathbb{R})} f(x,M)\,\nu_x^k(dM)\,dx + \int_{\bar\Omega}\int_{\partial B_d} f^\infty(x,M)\,\mu_x^k(dM)\,\sigma^k(dx)\end{aligned} \tag{5.3}$$

converge to $\llbracket f,\upsilon\rrbracket$ (defined analogously) as $k\to\infty$, for $f\in\mathscr{R}(\Omega;\mathrm{Mat}_d(\mathbb{R}))$ with

$$\mathscr{R}\big(\Omega;\mathrm{Mat}_d(\mathbb{R})\big) := \left\{\begin{array}{l} f\colon \bar\Omega\times\mathrm{Mat}_d(\mathbb{R})\longrightarrow\mathbb{R}\ : \\ \quad\text{the map } f \text{ is a } \textbf{Carathéodory function} \\ \quad\text{with} \\ \quad\text{linear growth at infinity, and there exists} \\ \quad f^\infty\in\mathscr{C}\big(\bar\Omega\times\mathrm{Mat}_d(\mathbb{R})\big)\end{array}\right\};$$

see Corollary 2 and Proposition 2 in [55]. Recall that the map $f\colon\bar\Omega\longrightarrow\mathbb{R}^d$ is called a Carathéodory function if it is $\mathcal{L}^d\times\mathcal{B}(\mathrm{Mat}_d(\mathbb{R}))$-measurable and if $M\mapsto f(x,M)$ is continuous for a.e. $x\in\bar\Omega$. It is enough to check the measurability of $x\mapsto f(x,M)$ for all $M\in\mathrm{Mat}_d(\mathbb{R})$ fixed; see Proposition 5.6 in [2]. The map f has linear growth at infinity if there exists $c\geqslant 0$ such that $|f(x,M)|\leqslant c(1+\|M\|)$ for a.e. $x\in\bar\Omega$ and all $M\in\mathrm{Mat}_d(\mathbb{R})$. We denote by f^∞ the recession function of f, defined as

$$f^\infty(x,M) := \lim_{\substack{x'\to x\\ M'\to M\\ t\to\infty}} \frac{f(x',tM')}{t} \quad\text{for a.e. } x\in\bar\Omega \text{ and all } M\in\mathrm{Mat}_d(\mathbb{R}). \tag{5.4}$$

Note that the recession function is positively 1-homogeneous in M, if it exists. We call a triple $\upsilon=(\nu,\sigma,\mu)$ obtained as a limit as above a gradient Young measure and denote the space of gradient Young measures by $\mathscr{G}(\Omega;\mathrm{Mat}_d(\mathbb{R}))$. Then

$$D\boldsymbol{t} = [\mathrm{id},\nu]\,\mathcal{L}^d + [\mathrm{id},\mu]\,\sigma,$$

by construction (cf. (5.3)). Moreover, we have

$$\|\nabla\boldsymbol{t}^k\|\,\mathcal{L}^d + \left\|\frac{\mathrm{d}D^s\boldsymbol{t}^k}{\mathrm{d}|D^s\boldsymbol{t}^k|}\right\|\,|D^s\boldsymbol{t}^k| \rightharpoonup [\|\cdot\|,\nu]\,\mathcal{L}^d + [\|\cdot\|,\mu]\,\sigma$$

weak* in $\mathscr{M}(\bar\Omega)$ as $k\to\infty$, which implies that $[\|\cdot\|,\nu]\in\mathscr{L}^1(\Omega)$. We used the fact that the recession function of $f(x,M):=\varphi(x)\|M\|$ with $\varphi\in\mathscr{C}(\bar\Omega)$ coincides with f. We refer the reader to [55] for further information.

We apply this framework to sequences of *monotone* functions $\boldsymbol{t}^k\in\mathrm{BV}(\Omega;\mathbb{R}^d)$ (see Remark 3.1), in which case the derivatives $D\boldsymbol{t}^k$ are positive (that is, matrix-valued and locally finite) measures; see Theorem 5.3 in [1]. Since the map $(M,v)\mapsto v\cdot(Mv)$ is continuous, the set $\mathrm{Mat}_d(\mathbb{R},>)$ is open and convex; the set $\mathrm{Mat}_d(\mathbb{R},\geqslant)$ is a closed convex cone. Recall that a matrix M is an element of $\mathrm{Mat}_d(\mathbb{R},\geqslant)$ (resp. $\mathrm{Mat}_d(\mathbb{R},>)$) if and only if its symmetric part $M^{\mathrm{sym}}\in\mathrm{Sym}_d(\mathbb{R},\geqslant)$ (resp. $\mathrm{Sym}_d(\mathbb{R},>)$).

Proposition 5.1 (Gradient Young Measures)**.** *Let $\Omega\subset\mathbb{R}^d$ be a bounded Lipschitz domain and suppose that $\boldsymbol{t}^k\rightharpoonup\boldsymbol{t}$ weak* in $\mathrm{BV}(\Omega;\mathbb{R}^d)$ with $\boldsymbol{t}^k,\boldsymbol{t}\in\mathrm{BV}(\Omega;\mathbb{R}^d)$ monotone. For all $k\in\mathbb{N}$ we denote by υ^k the elementary gradient Young measure associated to $D\boldsymbol{t}^k$, as introduced above. Then there exists a subsequence (which we* *do not relabel, for simplicity) and a gradient Young measure* $\upsilon \in \mathscr{G}(\Omega; \mathrm{Mat}_d(\mathbb{R}))$ *with the property that* $[\![f, \upsilon^k]\!] \longrightarrow [\![f, \upsilon]\!]$ *for all* $f \in \mathscr{R}_+(\Omega; \mathrm{Mat}_d(\mathbb{R}))$*, where*

$$\mathscr{R}_+\big(\Omega; \mathrm{Mat}_d(\mathbb{R})\big) := \left\{ \begin{array}{l} f\colon \bar{\Omega} \times \mathrm{Mat}_d(\mathbb{R}) \longrightarrow \mathbb{R} : \\ \quad \textit{the map } f \textit{ is a Carathéodory function with} \\ \quad \textit{linear growth at infinity, and there exists} \\ \quad f^\infty \in \mathscr{C}\big(\bar{\Omega} \times \mathrm{Mat}_d(\mathbb{R}, \geqslant)\big) \end{array} \right\}.$$

Proof. It suffices to check continuity of the recession function f^∞ on the smaller set $\mathrm{Mat}_d(\mathbb{R}, \geqslant)$ because all gradient Young measures considered above vanish outside of $\Omega \times \mathrm{Mat}_d(\mathbb{R}, \geqslant)$. Indeed, consider any test function $f \in \mathscr{R}(\Omega; \mathrm{Mat}_d(\mathbb{R}))$ of the form $f(x, M) = \varphi(x) h(M)$, with $\varphi \in \mathscr{C}_\mathrm{c}(\Omega)$ non-negative, $h(M) := \mathrm{dist}(M, \mathrm{Mat}_d(\mathbb{R}, \geqslant))$ for all $M \in \mathrm{Mat}_d(\mathbb{R})$. Then the map h is positively 1-homogeneous. This follows immediately from the fact that $\mathrm{Mat}_d(\mathbb{R}, \geqslant)$ is a cone. It can also be derived from the following observation: Notice first that symmetric and antisymmetric matrices in $\mathrm{Mat}_d(\mathbb{R})$ are orthogonal to each other with respect to the Frobenius inner product. For given $M \in \mathrm{Mat}_d(\mathbb{R})$ let $M^\mathrm{sym} = QP$ be a polar decomposition of its symmetric part (so that $Q^\mathrm{T} Q = \mathbb{1}$ and $P = P^\mathrm{T} \geqslant 0$). Then

$$X_M := M^\mathrm{anti} + (M^\mathrm{sym} + P)/2$$

is the unique element in $\mathrm{Mat}_d(\mathbb{R}, \geqslant)$ closest to M in the Frobenius norm, and

$$\mathrm{dist}(M, \mathrm{Mat}_d(\mathbb{R}, \geqslant))^2 = \sum_{\lambda_i(M^\mathrm{sym}) < 0} \lambda_i(M^\mathrm{sym})^2,$$

with $\lambda_i(M^\mathrm{sym})$ the (real) eigenvalues of M^sym; see [50]. Then the claim follows.

Since h is positively 1-homogeneous it is sufficient to consider the limits $x' \to x$ and $M' \to M$ in (5.4) to define the recession function of f. But φ, h are continuous, and hence f^∞ coincides with f. In particular, this proves that $f \in \mathscr{R}(\Omega; \mathrm{Mat}_d(\mathbb{R}))$. Note that $h(M) = 0$ if and only if $M \in \mathrm{Mat}_d(\mathbb{R}, \geqslant)$. If $\boldsymbol{t}^k$ is monotone, then

$$\nabla \boldsymbol{t}^k(x) \in \mathrm{Mat}_d(\mathbb{R}, \geqslant) \quad \text{and} \quad p^k(x) \in \mathrm{Mat}_d(\mathbb{R}, \geqslant)$$

for a.e. $x \in \Omega$ and $|D^s \boldsymbol{t}^k|$-a.e. $x \in \Omega$, respectively, where

$$D^s \boldsymbol{t}^k = \frac{\mathrm{d} D^s \boldsymbol{t}^k}{\mathrm{d}|D^s \boldsymbol{t}^k|} \, |D^s \boldsymbol{t}^k|, \quad p^k := \frac{\mathrm{d} D^s \boldsymbol{t}^k}{\mathrm{d}|D^s \boldsymbol{t}^k|} \in \mathscr{L}^1(\Omega, |D^s \boldsymbol{t}^k|; \partial B_d)$$

is the polar decomposition of $D^s \boldsymbol{t}^k$. If υ^k is the elementary gradient Young measure associated to $D\boldsymbol{t}^k$, then $[\![f, \upsilon^k]\!] = 0$ for all $k \in \mathbb{N}$ and f as above. Then the gradient Young measure υ generated by $\{\upsilon^k\}_k$ satisfies $[\![f, \upsilon]\!] = 0$ because $[\![f, \upsilon^k]\!] \longrightarrow [\![f, \upsilon]\!]$ for all $f \in \mathscr{R}(\Omega; \mathrm{Mat}_d(\mathbb{R}))$. Since $\varphi \in \mathscr{C}_\mathrm{c}(\Omega)$ non-negative was arbitrary, we get that the gradient Young measure υ vanishes outside of $\Omega \times \mathrm{Mat}_d(\mathbb{R}, \geqslant)$. A careful inspection of the proof of Proposition 2 in [55] now yields the result: Convergence of the gradient Young measures follows from the weak* convergence of

$$\nu^k \, \mathcal{L}^d + \mu^k \, \sigma^k \rightharpoonup \nu \, \mathcal{L}^d + \mu \, \sigma$$

on (a suitable compactification of) $\Omega \times \mathrm{Mat}_d(\mathbb{R})$, which reduces to weak* convergence on $\Omega \times \mathrm{Mat}_d(\mathbb{R}, \geqslant)$ whenever $\boldsymbol{t}^k$ and $\boldsymbol{t}$ are monotone. □

5.2. **Internal Energy.** We introduce a functional on the space of monotone BV-vector fields that represents the internal energy. This functional will be convex and lower semicontinuous with respect to weak* convergence in $\mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$.

Let us start with two auxiliary results.

Lemma 5.2. *For any $\gamma > 1$, the map $h\colon \mathrm{Mat}_d(\mathbb{R}) \longrightarrow [0,\infty]$ defined by*

$$h(M) := \begin{cases} \det(M^{\mathrm{sym}})^{1-\gamma} & \text{if } M \in \mathrm{Mat}_d(\mathbb{R},>), \\ \infty & \text{otherwise,} \end{cases} \tag{5.5}$$

is lower semicontinuous, proper, and convex. For all $M \in \mathrm{Mat}_d(\mathbb{R})$, we have

$$h^\infty(M) := \lim_{t\to\infty} \frac{h(\mathbb{1}+tM)-h(\mathbb{1})}{t} = \begin{cases} 0 & \text{if } M \in \mathrm{Mat}_d(\mathbb{R},\geqslant), \\ \infty & \text{otherwise.} \end{cases} \tag{5.6}$$

Proof. Since $M \mapsto \det(M^{\mathrm{sym}})$ is continuous, the function h is lower semicontinuous. It is proper because $h(\mathbb{1}) = 1$. In order to prove the convexity of h, we observe that $S \mapsto \det(S)^{1/d}$ is concave for all symmetric, positive definite $S \in \mathrm{Mat}_d(\mathbb{R})$. Indeed, pick any two such matrices S^0 and S^1. For all $t \in [0,1]$ we can write

$$\det\big((1-t)S^0 + tS^1\big)^{1/d} = \big(\det(S^0)\det(\mathbb{1}+tB)\big)^{1/d},$$

where $B := C^{-1}(S^1-S^0)C^{-1}$ and $C := \sqrt{S^0}$. The matrix C exists and is invertible since S^0 is symmetric and positive definite, by assumption. Then we compute

$$\begin{aligned} \frac{d}{dt}\det(\mathbb{1}+tB)^{1/d} &= \det(\mathbb{1}+tB)^{1/d}\left\{\frac{1}{d}\mathrm{tr}(D)\right\}, \\ \frac{d^2}{dt^2}\det(\mathbb{1}+tB)^{1/d} &= \det(\mathbb{1}+tB)^{1/d}\left\{\frac{1}{d^2}\mathrm{tr}(D)^2 - \frac{1}{d}\mathrm{tr}(D^2)\right\}, \end{aligned} \tag{5.7}$$

where $D := B(\mathbb{1}+tB)^{-1}$. The matrix D is symmetric. Therefore

$$\mathrm{tr}(D)^2 = (\lambda_1+\cdots+\lambda_d)^2 \leqslant d(\lambda_1^2+\cdots+\lambda_d^2) = d\,\mathrm{tr}(D^2),$$

where $\lambda_1,\dots,\lambda_d$ are the real eigenvalues of D. Hence (5.7) is non-positive for every $s \in [0,1]$. The composition of a concave function with a convex, non-increasing map is convex. Therefore the map $S \mapsto \det(S)^{1-\gamma}$ is convex for all symmetric, positive definite $S \in \mathrm{Mat}_d(\mathbb{R})$. Finally, the composition of any convex function with the linear map $M \mapsto M^{\mathrm{sym}}$ is again convex. Then the result follows.

To prove (5.6), we use that the map $t \mapsto (h(\mathbb{1}+tM)-h(\mathbb{1}))/t$ is non-decreasing (hence $\lim_{t\to\infty} = \sup_{t>0}$), by convexity of h. If $M \notin \mathrm{Mat}_d(\mathbb{R},\geqslant)$, then there exists $v \in \mathbb{R}^d$, $\|v\| = 1$, such that $\langle v, Mv\rangle < 0$. For sufficiently large $t > 0$, we get

$$\langle v, (\mathbb{1}+tM)v\rangle = 1 + t\langle v, Mv\rangle < 0,$$

and thus $h(\mathbb{1}+tM) = \infty$. This proves (5.6) for $M \notin \mathrm{Mat}_d(\mathbb{R},\geqslant)$.

If $M \in \mathrm{Mat}_d(\mathbb{R},\geqslant)$, then $\mathbb{1}+tM \in \mathrm{Mat}_d(\mathbb{R},>)$ for all $t > 0$, because

$$\langle v, (\mathbb{1}+tM)v\rangle = 1 + t\langle v, Mv\rangle \geqslant 1$$

for all $v \in \mathbb{R}^d$, $\|v\| = 1$. By convavity of $M \mapsto \det(M^{\mathrm{sym}})^{1/d}$, we obtain

$$\begin{aligned} \det(\mathbb{1}+tM^{\mathrm{sym}})^{1/d} &= \det\left((1+t)\left(\frac{1}{1+t}\mathbb{1} + \frac{t}{1+t}M^{\mathrm{sym}}\right)\right)^{1/d} \\ &\geqslant (1+t)\left(\frac{1}{1+t}\det(\mathbb{1})^{1/d} + \frac{t}{1+t}\det(M^{\mathrm{sym}})^{1/d}\right) \geqslant 1 \end{aligned}$$

for $t > 0$. Notice that $\det(M^{\mathrm{sym}}) \geqslant 0$. This implies $\det(\mathbb{1} + tM^{\mathrm{sym}})^{1-\gamma} \leqslant 1$ (recall that $\gamma > 1$, by assumption), and so (5.6) follows for $M \in \mathrm{Mat}_d(\mathbb{R}, \geqslant)$ as well. □

Lemma 5.3. *For any $n \in \mathbb{N}$, we define the* inf*-convolution*

$$h_n(M) := \inf_{B \in \mathrm{Mat}_d(\mathbb{R})} \Big\{ n\|M - B\| + h(B) \Big\} \tag{5.8}$$

for all $M \in \mathrm{Mat}_d(\mathbb{R})$, which has the following properties:

1. *The map h_n is lower semicontinuous, proper, and convex.*
2. *For all $M \in \mathrm{Mat}_d(\mathbb{R})$, we have $h_n(M) \longrightarrow h(M)$ monotonically from below.*
3. *The map h_n is Lipschitz continuous with Lipschitz constant n.*
4. *The map h_n has linear growth at infinity:*
$$h_n(M) \leqslant 1 + n\sqrt{d} + n\|M\| \quad \textit{for all } M \in \mathrm{Mat}_d(\mathbb{R}). \tag{5.9}$$
5. *For all $M \in \mathrm{Mat}_d(\mathbb{R})$, we have that*
$$h_n^\infty(M) := \lim_{t\to\infty} \frac{h_n(\mathbb{1} + tM) - h_n(\mathbb{1})}{t} = n \operatorname{dist}(M, \mathrm{Mat}_d(\mathbb{R}, \geqslant)). \tag{5.10}$$

Proof. Statement (1) follows from Corollary 9.2.2 in [70]: Notice first that the norm and h are lower semicontinuous, convex, and proper; see Lemma 5.2. The recession function of the norm is the norm itself, and it holds

$$n\|M\| + h^\infty(-M) > 0 \quad \text{for all } M \in \mathrm{Mat}_d(\mathbb{R}), M \neq 0.$$

Statements (2) and (3) follow from Lemma 1.61 in [2].

To prove (5.9), we just choose $B = \mathbb{1}$ in (5.8) and use the triangle inequality.

Finally, statement (5) follows from Corollary 9.2.1 in [70]. We must prove that for all pairs of matrices $M_1, M_2 \in \mathrm{Mat}_d(\mathbb{R})$ with the property that

$$n\|M_1\| + h^\infty(M_2) \leqslant 0 \quad \text{and} \quad n\|-M_1\| + h^\infty(-M_2) > 0, \tag{5.11}$$

it holds $M_1 + M_2 \neq 0$. The first condition in (5.11) is only satisfied if

$$M_1 = 0 \quad \text{and} \quad M_2 \in \mathrm{Mat}_d(\mathbb{R}, \geqslant),$$

because of (5.6). Then the second condition requires $-M_2 \notin \mathrm{Mat}_d(\mathbb{R}, \geqslant)$, and thus there exists $v \in \mathbb{R}^d$, $v \neq 0$, with $\langle v, -M_2 v\rangle < 0$ (consistent with $M_2 \in \mathrm{Mat}_d(\mathbb{R}, \geqslant)$). This is only possible if $M_2 \neq 0$, so the claim follows. We then obtain that the recession function (5.10) of the inf-convolution (5.8) is given by

$$h_n^\infty(M) = \inf_{B \in \mathrm{Mat}_d(\mathbb{R})} \Big\{ n\|M - B\| + h^\infty(B) \Big\}$$

for all $M \in \mathrm{Mat}_d(\mathbb{R})$, which implies the result because of (5.6). □

We can now prove the following lower semicontinuity result.

Proposition 5.4 (Internal Energy)**.** *Let $\Omega \subset \mathbb{R}^d$ be open and convex, and h given by* (5.5)*. For $U \in \mathscr{L}^1(\Omega)$ non-negative and $\boldsymbol{t} \in \mathrm{BV}_{\mathrm{loc}}(\Omega; \mathbb{R}^d)$ we define*

$$\mathcal{U}[\boldsymbol{t}] := \begin{cases} \displaystyle\int_\Omega U(x) h\big(\nabla \boldsymbol{t}(x)\big)\, dx & \textit{if } \boldsymbol{t} \textit{ monotone}, \\ +\infty & \textit{otherwise}, \end{cases} \tag{5.12}$$

using again the decomposition (5.2)*. Then the following is true:*

1. *The functional $\mathcal{U}$ is convex.*

(2) *For any* $\boldsymbol{t}^k \rightharpoonup \boldsymbol{t}$ *weak* in* $\mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$ *with* $\boldsymbol{t}^k, \boldsymbol{t} \in \mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$ *monotone, there exists a subsequence (not relabeled) such that*

$$\mathcal{U}[\boldsymbol{t}] \leqslant \liminf_{k\to\infty} \mathcal{U}[\boldsymbol{t}^k].$$

Remark 5.5. Notice that in (5.12) we only consider the part of $D\boldsymbol{t}$ that is absolutely continuous with respect to $\mathcal{L}^d$ and disregard the singular component. The intuition is that (for each direction) only increasing jumps are allowed in the transport map $\boldsymbol{t}$, which correspond to the formation of vacuum, which is admissible.

Proof of Proposition 5.4. We proceed in two steps.

Step 1. Consider $\boldsymbol{t}^k \in \mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$ with $k = 0..1$. For any $s \in (0,1)$ we define $\boldsymbol{t}^s := (1-s)\boldsymbol{t}^0 + s\boldsymbol{t}^1 \in \mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$. If $\mathcal{U}[\boldsymbol{t}^k] = +\infty$ for $k = 0$ or $k = 1$, then there is nothing to prove, so we may assume that both terms are finite. This requires that both $\boldsymbol{t}^k$ are monotone and $\nabla\boldsymbol{t}^k(x) \in \mathrm{Mat}_d(\mathbb{R},>)$ for $U\mathcal{L}^d$-a.e. $x \in \Omega$. It follows that $\boldsymbol{t}^s$ is monotone as well and $\nabla\boldsymbol{t}^s(x) \in \mathrm{Mat}_d(\mathbb{R},>)$ for $U\mathcal{L}^d$-a.e. $x \in \Omega$. Then

$$h\big(\nabla\boldsymbol{t}^s(x)\big) \leqslant (1-s)h\big(\nabla\boldsymbol{t}^0(x)\big) + sh\big(\nabla\boldsymbol{t}^1(x)\big);$$

see Lemma 5.2. Multiplying by $U(x)$ and integrating in Ω, we obtain

$$\mathcal{U}[\boldsymbol{t}^s] \leqslant (1-s)\mathcal{U}[\boldsymbol{t}^0] + s\mathcal{U}[\boldsymbol{t}^1]$$

for all $s \in [0,1]$. This proves the convexity of the functional.

Step 2. We first introduce a sequence of bounded open convex sets

$$\Omega_n := \Big\{x \in B_n(0)\colon \mathrm{dist}(x, \mathbb{R}^d \setminus \Omega) > 1/n\Big\},$$

which are bounded Lipschitz domains. We have $\Omega_{n-1} \subset \Omega_n$ for all $n \in \mathbb{N}$.

We then choose a sequence of cut-off functions $\varphi_n \in \mathscr{C}_{\mathrm{c}}(\Omega;[0,1])$ with $\varphi_n(x) = 1$ for all $x \in \Omega_{n-1}$ and $\varphi_n(x) = 0$ for all $x \notin \Omega_n$. For all $n \in \mathbb{N}$ we define

$$f_n(x,M) := \big(U(x) \wedge n\big)\varphi_n(x)h_n(M) \quad \text{for all } (x,M) \in \Omega \times \mathrm{Mat}_d(\mathbb{R}), \tag{5.13}$$

where h_n is given by (5.8). Because of Lemma 5.3, the map f_n is a Carathéodory function with linear growth at infinity. In fact, we can estimate

$$0 \leqslant f_n(x,M) \leqslant n(1 + n\sqrt{d} + n\|M\|) \quad \text{for all } (x,M) \in \Omega \times \mathrm{Mat}_d(\mathbb{R}).$$

We prove that $f_n^\infty(x,M) = 0$ for all $(x,M) \in \Omega \times \mathrm{Mat}_d(\mathbb{R},\geqslant)$: Note first that

$$\left|\frac{f_n(x', tM')}{t} - 0\right| \leqslant nh_n(tM')/t \leqslant n\Big\{h_n(tM)/t + n\|M' - M\|\Big\},$$

uniformly in $x' \in \Omega$. Recall that h_n is Lipschitz continuous with Lipschitz constant n. Since $h_n(M) < \infty$ for all $M \in \mathrm{Mat}_d(\mathbb{R})$, by Theorem 8.5 in [70] we have

$$h_n(tM)/t \longrightarrow h_n^\infty(M) \quad \text{as } t \to \infty,$$

which vanishes for $M \in \mathrm{Mat}_d(\mathbb{R},\geqslant)$; see Lemma 5.3. Then $f_n^\infty \in \mathscr{C}(\Omega\times\mathrm{Mat}_d(\mathbb{R},\geqslant))$, and so $f_n \in \mathscr{R}_+(\Omega_n;\mathrm{Mat}_d(\mathbb{R}))$ for all $n \in \mathbb{N}$. By construction, it holds

$$f_n(x,M) \leqslant f_{n+1}(x,M) \quad \text{and} \quad U(x)h(M) = \sup_n f_n(x,M) \tag{5.14}$$

for all $(x,M) \in \Omega \times \mathrm{Mat}_d(\mathbb{R})$. We used again Lemma 5.3.

Let us fix $n \in \mathbb{N}$ for a moment. Extracting a subsequence if necessary, we may assume that the sequence of elementary gradient Young measures υ^k generated by $D\boldsymbol{t}^k|\Omega_n$ converges to $\upsilon = (\nu, \sigma, \mu) \in \mathscr{G}(\Omega_n; \mathrm{Mat}_d(\mathbb{R}))$ in the sense that

$$\llbracket f, \upsilon^k \rrbracket \longrightarrow \llbracket f, \upsilon \rrbracket \quad \text{for all } f \in \mathscr{R}_+\big(\Omega_n; \mathrm{Mat}_d(\mathbb{R})\big); \tag{5.15}$$

see Proposition 5.1. It holds $D\boldsymbol{t} = [\mathrm{id}, \nu]\, \mathcal{L}^d + [\mathrm{id}, \mu]\, \sigma$. Comparing this identity with the Lebesgue-Radon-Nikodým decomposition (5.2), we find

$$\nabla \boldsymbol{t} = [\mathrm{id}, \nu] + [\mathrm{id}, \mu] \frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{L}^d} \quad \text{a.e.} \quad \text{and} \quad D^s u = [\mathrm{id}, \mu]\, \sigma^s,$$

where $\sigma^s \perp \mathcal{L}^d$ is the singular part of σ. Note that $[\mathrm{id}, \mu_x] \in \mathrm{Mat}_d(\mathbb{R})$ may not have unit length for σ^s-a.e. $x \in \Omega$. The polar decomposition of $D^s u$ is given by

$$|D^s u| = |[\mathrm{id}, \mu]|\, \sigma^s \quad \text{and} \quad \frac{\mathrm{d}D^s u}{\mathrm{d}|D^s u|} = \frac{[\mathrm{id}, \mu]}{|[\mathrm{id}, \mu]|} \quad |D^s u|\text{-a.e.}$$

We now apply the convergence (5.15) to the function f_n defined in (5.13), whose restriction to Ω_n belongs to $\mathscr{R}_+(\Omega_n; \mathrm{Mat}_d(\mathbb{R}))$. We observe that

$$f_n^\infty(\cdot, [\mathrm{id}, \mu])\, \sigma^s = f_n^\infty\left(\cdot, \frac{\mathrm{d}D^s u}{\mathrm{d}|D^s u|}\right) |D^s u|$$

because the map $M \mapsto f_n^\infty(x, M)$ is positively 1-homogeneous for $x \in \Omega_n$. Then the following Jensen-type inequalities hold (see Theorem 9 in [55]):

$$f_n(\cdot, \nabla u) \leqslant [f_n, \nu] + [f_n^\infty, \mu] \frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{L}^d} \quad \text{a.e.},$$

$$f_n^\infty(\cdot, [\mathrm{id}, \mu]) \leqslant [f_n^\infty, \mu] \quad \sigma^s\text{-a.e.}$$

because the map $M \mapsto f_n(x, M)$ is convex for $x \in \Omega$. We can then estimate

$$\begin{aligned} \lim_{k\to\infty} & \int_{\Omega_n} f_n(\cdot, \nabla u^k) + \int_{\Omega_n} f_n^\infty\left(\cdot, \frac{\mathrm{d}D^s u^k}{\mathrm{d}|D^s u^k|}\right) |D^s u^k| \\ &= \int_{\Omega_n} [f_n, \nu] + \int_{\Omega_n} [f_n^\infty, \mu]\, \sigma \\ &= \int_{\Omega_n} \left([f_n, \nu] + [f_n^\infty, \mu] \frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{L}^d}\right) + \int_{\Omega_n} [f_n^\infty, \mu]\, \sigma^s \\ &\geqslant \int_{\Omega_n} f_n(\cdot, \nabla u) + \int_{\Omega_n} f_n^\infty\left(\cdot, \frac{\mathrm{d}D^s u}{\mathrm{d}|D^s u|}\right) |D^s u|. \end{aligned}$$

Clearly the integrals can be extended to all of Ω because f_n vanishes outside of Ω_n. Moreover, we have shown that the recession function f_n^∞ vanishes. Hence

$$\int_\Omega f_n(x, \nabla u(x))\, dx \leqslant \liminf_{k\to\infty} \int_\Omega U(x) h(\nabla u^k(x))\, dx, \tag{5.16}$$

where we also used (5.14). By a standard diagonal argument (successively extracting subsequences if necessary), we may assume that (5.16) holds for all $n \in \mathbb{N}$. We then use (5.14) and the monotone convergence theorem to obtain the result. □

We finish the section with an estimate on determinants of square matrices.

Lemma 5.6. *Suppose S is a real, positive semidefinite, symmetric $(d \times d)$-matrix. For any real skew-symmetric $(d \times d)$-matrix A we have*

$$\det(S + A) \geqslant \det S \geqslant 0.$$

Proof. We divide the proof into two steps.

Step 1. We will first prove that if $\det S = 0$, then $\det(S+A) \geqslant 0$. Recall that the determinants of square matrices equal the product of their eigenvalues. Non-real eigenvalues of $S+A$ can only occur in complex conjugate pairs because S, A are real matrices. Since the product of two complex conjugate numbers is non-negative, it remains to prove that every *real* eigenvalue of $S+A$ must be non-negative. Let $\lambda \in \mathbb{R}$ be an eigenvalue with corresponding eigenvector v. Note that if v is complex, then its complex conjugate $\bar{v}$ is another eigenvector to the same eigenvalue λ. Taking the sum $v+\bar{v}$ if necessary, we may therefore assume that $v \in \mathbb{R}^d$. We have

$$(S+A)v = \lambda v, \quad \|v\| > 0.$$

We take the inner product with v and obtain (since A is skew-symmetric)

$$\lambda \|v\|^2 = \langle (S+A)v, v\rangle = \langle Sv, v\rangle.$$

The right-hand side is non-negative because S is positive semidefinite. Hence $\lambda \geqslant 0$. From this, we conclude that $\det(S+A) \geqslant \det S$ whenever $\det S = 0$.

Step 2. Consider now $\det S \neq 0$. Since S is positive semidefinite and symmetric, all eigenvalues of S (which are real) are positive. Therefore $\det S > 0$ and $\langle Sv, v\rangle > 0$ for every $v \in \mathbb{R}^d$ with $v \neq 0$. We claim that $\det(S+tA) > 0$ for every $t \in \mathbb{R}$. In fact, assume this is false. Then zero is an eigenvalue of $S+tA$, with corresponding eigenvector $v \in \mathbb{R}^d$ (see above). We have $(S+tA)v = 0$ and $v \neq 0$. We get

$$0 < \langle Sv, v\rangle = \langle (S+tA)v, v\rangle = 0,$$

using again that A is skew-symmetric. This contradiction proves the claim.

For all $t \in \mathbb{R}$, we can now define $f(t) := \log\det(S+tA)$. We compute

$$f'(t) = \operatorname{tr}\big((S+tA)^{-1}A\big).$$

Notice that $t(S+tA)^{-1}A = \mathbb{1} - (S+tA)^{-1}S$. Since S is symmetric, there exists an orthogonal matrix Q such that $Q^{-1}SQ = \Lambda$, where $\Lambda := \operatorname{diag}(\lambda_1, \dots, \lambda_d)$ contains the eigenvalues $\lambda_i > 0$ of S, $i = 1 \dots d$. Let e_i denote the ith standard basis vector of $\mathbb{R}^d$. Since the trace is invariant under changes of basis, we obtain

$$\begin{aligned} t\operatorname{tr}\big((S+tA)^{-1}A\big) &= \operatorname{tr}\big(\mathbb{1} - Q^{-1}(S+tA)^{-1}SQ\big) \\ &= \sum_{i=1}^{d} \big(1 - \langle Q^{-1}(S+tA)^{-1}SQe_i, e_i\rangle\big). \end{aligned}$$

We denote by v_i the ith column vector of Q (hence $v_i = Qe_i$), which is a normalized eigenvector of S corresponding to the eigenvalue λ_i. As $Q^{-1} = Q^{\mathrm{T}}$, we have

$$t\operatorname{tr}\big((S+tA)^{-1}A\big) = \sum_{i=1}^{d} (1 - \lambda_i \langle w_i, v_i\rangle), \quad w_i := (S+tA)^{-1}v_i. \tag{5.17}$$

Since the eigenvectors $v_1, \dots, v_d$ form an orthonormal basis of $\mathbb{R}^d$, there is a unique expansion $w_i = \sum_{k=1}^d \alpha_i^k v_k$ with $\alpha_i^k := \langle w_i, v_k\rangle$. Using this expansion, we get

$$\alpha_i^i = \langle w_i, (S+tA)w_i\rangle = \langle w_i, Sw_i\rangle = \sum_{k=1}^{d} \lambda_k (\alpha_i^k)^2 \tag{5.18}$$

for $i = 1 \dots d$. Recall that the eigenvalues λ_k are all positive and A is skew-symmetric. We conclude that $\alpha_i^i \geqslant 0$. Moreover, rewriting (5.18) in the form

$$\alpha_i^i(1 - \lambda_i \alpha_i^i) = \sum_{k \neq i} \lambda_k (\alpha_i^k)^2 \geqslant 0,$$

we obtain that $1 - \lambda_i \langle w_i, v_i \rangle \geqslant 0$ for each i. Using this estimate in (5.17), we conclude that $f'(t) \geqslant 0$ for all $t > 0$, and so the map $t \mapsto f(t)$ is non-decreasing for such t. In particular, we have that $\det(S + A) = \exp f(1) \geqslant \exp f(0) = \det S > 0$. □

Remark 5.7. Lemma 5.6 can be made more precise if $\det(S) > 0$: We first write

$$\begin{aligned}\det(S + A)^\beta - \det(S)^\beta &= \int_0^1 \frac{d}{dt} \det(S + tA)^\beta \, dt \\ &= \beta \int_0^1 \det(S + tA)^\beta \operatorname{tr}\big((S + tA)^{-1} A\big) \, dt\end{aligned}$$

for $\beta \in \mathbb{R}$. Notice that all terms are well-defined, and the integrand on the right-hand side is non-negative. Since S is positive definite and symmetric, we can compute its root, which is the unique $R \in \mathrm{Sym}_d(\mathbb{R}, >)$ such that $R^2 = S$. Then

$$\begin{aligned}\det(S + tA) &= \det(S) \det(\mathbb{1} + tC), \\ \operatorname{tr}\big((S + tA)^{-1} A\big) &= \operatorname{tr}\big((\mathbb{1} + tC)^{-1} C\big),\end{aligned}$$

with $C := R^{-1} A R^{-1}$ skew-symmetric. We obtain the following identity:

$$\left(\frac{\det(S + A)}{\det(S)}\right)^\beta - 1 = \beta \int_0^1 \det(\mathbb{1} + tC)^\beta \operatorname{tr}\big((\mathbb{1} + tC)^{-1} C\big) \, dt,$$

where the integral on the right-hand side is non-negative.

5.3. **Minimization Problem.** We now introduce the main minimization problem for (1.1). We represent the state of the fluid by $(\varrho, \boldsymbol{\mu}, \sigma)$, with $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ the density, $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ the velocity distribution, and $\sigma \in \mathscr{M}_+(\mathbb{R}^d)$ the thermodynamic entropy. We assume that $\mathcal{U}[\varrho, \sigma] < \infty$, which implies that $\varrho = r\mathcal{L}^d$ and $\sigma = \varrho S$ for suitable Borel functions r, S; see Definition 1.1. In the isentropic case, S will be constant in time and space. We want to minimize the sum of the internal energy of the transported fluid and the acceleration cost of the transport, over the cone $C_{\boldsymbol{\mu}}$ of monotone maps; see Definition 4.1. The following observation will be useful:

Lemma 5.8. *Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, where $\varrho \ll \mathcal{L}^d$. To every $\mathbf{t} \in C_{\boldsymbol{\mu}}$ we can associate a function $\boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ defined on all of $\mathbb{R}^d$ that satisfies*

$$\mathbf{t}(x, \xi) = \boldsymbol{t}(x) \quad \textit{for } \boldsymbol{\mu}\textit{-a.e. } (x, \xi) \in \mathbb{R}^{2d}. \tag{5.19}$$

The map $\boldsymbol{t}$ is monotone on $\Omega := \operatorname{int} \overline{\operatorname{conv}} \operatorname{spt} \varrho$ (hence $\boldsymbol{t} \in \mathrm{BV}_{\mathrm{loc}}(\Omega; \mathbb{R}^d)$) so that

$$\langle \boldsymbol{t}(x_1) - \boldsymbol{t}(x_2), x_1 - x_2 \rangle \geqslant 0 \quad \textit{for all } x_1, x_2 \in \Omega.$$

Proof. For $\mathbf{t} \in C_{\boldsymbol{\mu}}$ let u be any maximal monotone map associated to $\boldsymbol{\gamma} := (\mathbb{x}, \mathbf{t})\#\boldsymbol{\mu}$, which is in C_ϱ; see Definition 3.3. As shown in Lemma 3.4, the domain of u contains the convex open set Ω. As $\varrho \ll \mathcal{L}^d$, the set Ω must be non-empty and $\varrho(\mathbb{R}^d \setminus \Omega) = 0$ (since the boundary of $\overline{\operatorname{conv}} \operatorname{spt} \varrho$ is a Lipschitz manifold of codimension one, which is a Lebesgue null set and hence ϱ-negligible). Consequently, the maximal monotone map u associated to $\boldsymbol{\gamma}$ is defined ϱ-a.e. The map u is single-valued for all $x \in \Omega \setminus \Sigma^1(u)$

(see Remark 3.1), and $\Sigma^1(u)$ is a Lebesgue null set and hence ϱ-negligible. We now define a (single-valued) function $\boldsymbol{t}$ on all of $\mathbb{R}^d$ as follows:

$$\boldsymbol{t}(x) := \begin{cases} z & \text{if } x \in \Omega \setminus \Sigma^1(u) \text{ and } u(x) =: \{z\}, \\ \bar{z} & \text{if } x \in \Omega \cap \Sigma^1(u) \text{ and } \bar{z} \text{ is the center of mass of } u(x), \\ 0 & \text{if } x \in \mathbb{R}^d \setminus \Omega. \end{cases}$$

Then $\boldsymbol{t}$ is monotone on Ω because $\boldsymbol{t}(x) \in u(x)$ for every $x \in \Omega$. Recall that $u(x)$ is a closed convex set (possibly empty) for all $x \in \mathbb{R}^d$; see Proposition 1.2 in [1].

As shown in Remark 4.4, there exists a Borel set $N_{\mathbf{t}} \subset \mathbb{R}^{2d}$ such that $\boldsymbol{\mu}(N_{\mathbf{t}}) = 0$ and $(x, \mathbf{t}(x,\xi)) \in \operatorname{spt}\boldsymbol{\gamma}$ for all $(x,\xi) \in \mathbb{R}^{2d} \setminus N_{\mathbf{t}}$. This implies that $\mathbf{t}(x,\xi) \in u(x)$ for such (x,ξ), since $\operatorname{graph}(u)$ is an extension of $\operatorname{spt}\boldsymbol{\gamma}$. Therefore

$$\Big\{(x,\xi) \in \mathbb{R}^{2d} \colon \mathbf{t}(x,\xi) \neq \boldsymbol{t}(x)\Big\} \subset N_{\mathbf{t}} \cup (E \times \mathbb{R}^d),$$
$$\text{where} \quad E := (\mathbb{R}^d \setminus \Omega) \cup \big(\Omega \cap \Sigma^1(u)\big).$$

Since $\boldsymbol{\mu}(N_{\mathbf{t}}) = 0$ and $\boldsymbol{\mu}(E \times \mathbb{R}^d) = \varrho(E) = 0$, statement (5.19) follows. Now

$$\int_{\mathbb{R}^d} |\boldsymbol{t}(x)|^2 \,\varrho(dx) = \int_{\mathbb{R}^{2d}} |\boldsymbol{t}(x)|^2 \,\boldsymbol{\mu}(dx,d\xi) = \int_{\mathbb{R}^{2d}} |\mathbf{t}(x,\xi)|^2 \,\boldsymbol{\mu}(dx,d\xi),$$

which is finite. The $\mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$-regularity of $\boldsymbol{t}$ follows from Theorem 5.3 in [1]. □

Lemma 5.8 shows that instead of minimizing over $C_{\boldsymbol{\mu}}$ it is sufficient to consider a minimization over the following convex cone in $\mathscr{L}^2(\mathbb{R}^d,\varrho)$ (we refer the reader to the proof of Proposition 5.15 for topological properties):

Definition 5.9 (Configurations)**.** Let $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$ satisfy $\varrho \ll \mathcal{L}^d$. We denote by C_ϱ the set of all Borel maps $\boldsymbol{t}\colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ with the following properties:

(1) $\boldsymbol{t}$ is monotone on $\Omega := \operatorname{int}\overline{\operatorname{conv}}\operatorname{spt}\varrho$ (hence $\boldsymbol{t} \in \mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$),
(2) $\boldsymbol{t} \in \mathscr{L}^2(\mathbb{R}^d,\varrho)$.

If $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$ and $\mathbf{t} \in C_{\boldsymbol{\mu}}$ are given, and $\tau > 0$, then

$$\begin{aligned} &\int_{\mathbb{R}^{2d}} |(x+\tau\xi) - \mathbf{t}(x,\xi)|^2 \,\boldsymbol{\mu}(dx,d\xi) \\ &\qquad = \tau^2 \int_{\mathbb{R}^{2d}} |\xi - \boldsymbol{u}(x)|^2 \,\boldsymbol{\mu}(dx,d\xi) + \int_{\mathbb{R}^d} \big|\big(x+\tau\boldsymbol{u}(x)\big) - \boldsymbol{t}(x)\big|^2 \,\varrho(dx), \end{aligned} \tag{5.20}$$

for every map $\boldsymbol{t} \in C_\varrho$ satisfying (5.19). Here $\boldsymbol{u}$ is the barycentric projection $\mathrm{b}(\boldsymbol{\mu})$ of $\boldsymbol{\mu}$ (equivalently, the orthogonal projection of $\boldsymbol{\mu}$ onto the space of functions in $\mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$ that depend only on the spatial variable $x \in \mathbb{R}^d$). Notice that the first term on the right-hand side of (5.20) does not depend on $\mathbf{t}$ or $\boldsymbol{t}$.

For any smooth, strictly monotone map $\boldsymbol{t}\colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$, the internal energy of the fluid transported by $\boldsymbol{t}$ is given (after a change of variables) by

$$\begin{aligned} \mathcal{U}[\boldsymbol{t}\#\varrho, \boldsymbol{t}\#\sigma] &= \int_{\mathbb{R}^d} U\left(\left(\frac{r}{\det(\nabla\boldsymbol{t})}\right) \circ \boldsymbol{t}^{-1}(z), S \circ \boldsymbol{t}^{-1}(z)\right) dz \\ &= \int_{\mathbb{R}^d} U\left(\frac{r(x)}{\det\big(\nabla\boldsymbol{t}(x)\big)}, S(x)\right) \det\big(\nabla\boldsymbol{t}(x)\big)\, dx \\ &= \int_{\mathbb{R}^d} U\big(r(x), S(x)\big)\, \det\big(\nabla\boldsymbol{t}(x)\big)^{1-\gamma}\, dx. \end{aligned} \tag{5.21}$$

Since the matrix $\nabla\boldsymbol{t}$ may not be symmetric, the functional $\boldsymbol{t}\mapsto\mathcal{U}[\boldsymbol{t}\#\varrho,\boldsymbol{t}\#\sigma]$ is not convex if $d\geqslant 2$. In order to obtain a *convex* minimization problem, we modify the functional by replacing $\nabla\boldsymbol{t}$ by the deformation, i.e., its symmetric part.

Definition 5.10 (Internal Energy)**.** Suppose that $(\varrho,\sigma)\in\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{M}_+(\mathbb{R}^d)$ with $\varrho=r\mathcal{L}^d$, $\sigma=\varrho S$, and $\mathcal{U}[\varrho,\sigma]<\infty$. For any $\boldsymbol{t}\in C_\varrho$ let

$$D\boldsymbol{t}=\nabla\boldsymbol{t}\,\mathcal{L}^d+D^s\boldsymbol{t},\quad D^s\boldsymbol{t}\perp\mathcal{L}^d \tag{5.22}$$

be the Lebesgue-Radon-Nikodým decomposition of its derivative. Then

$$\mathcal{U}[\boldsymbol{t}|\varrho,\sigma]:=\int_{\mathbb{R}^d}U\big(r(x),S(x)\big)\;h\big(\nabla\boldsymbol{t}(x)\big)\,dx\quad\text{for }\boldsymbol{t}\in C_\varrho. \tag{5.23}$$

Recall that $h(\nabla\boldsymbol{t})$ only depends on the symmetric part of $\nabla\boldsymbol{t}$; see (5.5).

Remark 5.11. We have $U(r,S)\in\mathscr{L}^1(\mathbb{R}^d)$ as $\mathcal{U}[\varrho,\sigma]<\infty$. In (5.23) we may restrict the integration to $\Omega:=\operatorname{int}\overline{\operatorname{conv}}\operatorname{spt}\varrho$ because the measures $\nu:=U(r,S)\mathcal{L}^d$ and ϱ are mutually absolutely continuous, and $\varrho(\mathbb{R}^d\setminus\Omega)=0$ if $\varrho\ll\mathcal{L}^d$.

Remark 5.12. Using only the symmetric part of $\nabla\boldsymbol{t}$ can be justified by the expectation that the map $\boldsymbol{t}$ will be a perturbation of the identity, whose derivative is the identity matrix everywhere, which is symmetric. Using only $\nabla\boldsymbol{t}$ instead of the derivative $D\boldsymbol{t}$ means that the formation of vacuum does not cost any energy.

The following lemma allows us to control (5.21) in terms of (5.23).

Lemma 5.13. *Suppose that density/entropy $(\varrho,\sigma)\in\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{M}_+(\mathbb{R}^d)$ are given with $\varrho=:r\mathcal{L}^d$, $\sigma=:\varrho S$, and $\mathcal{U}[\varrho,\sigma]<\infty$. For any $\boldsymbol{t}\in C_\varrho$ with $\mathcal{U}[\boldsymbol{t}|\varrho,\sigma]<\infty$ there exists a Borel set $\Sigma\subset\mathbb{R}^d$ with $\varrho(\Sigma)=0$ and $\boldsymbol{t}|_{\mathbb{R}^d\setminus\Sigma}$ injective. Then*

$$\mathcal{U}[\boldsymbol{t}\#\varrho,\boldsymbol{t}\#\sigma]\leqslant\mathcal{U}[\boldsymbol{t}|\varrho,\sigma]. \tag{5.24}$$

Proof. We have $\varrho(\mathbb{R}^d\setminus\Omega)=0$ with $\Omega:=\operatorname{int}\overline{\operatorname{conv}}\operatorname{spt}\varrho$. Choose a maximal monotone set-valued map u whose graph is an extension of $\Gamma:=(\mathrm{id},\boldsymbol{t})\#\varrho$. Then $u(x)=\{\boldsymbol{t}(x)\}$ for a.e. $x\in\Omega$ and u is differentiable a.e.: there is a $(d\times d)$-matrix $A(x)$ with

$$\lim_{\substack{x'\to x\\ y\in u(x')}}\frac{y-\boldsymbol{t}(x)-A(x)\cdot(x'-x)}{|x'-x|}=0; \tag{5.25}$$

see Theorem 3.2 in [1]. It follows that the function $\boldsymbol{t}$ is approximately differentiable a.e. in Ω (see Definition 3.70 in [2]) and A coincides with the absolutely continuous part $\nabla\boldsymbol{t}$ of the derivative $D\boldsymbol{t}$; see Theorem 3.83 in [2] and (5.22).

Let D be the set of $x\in\Omega$ for which $u(x)$ is single-valued and u is differentiable at x in the sense of (5.25). Then $\mathcal{L}^d(\Omega\setminus D)=0$. We define

$$N:=\Big\{x\in D\colon\text{there exists }x'\in\Omega,\ x'\neq x,\text{ with }\boldsymbol{t}(x)\in u(x')\Big\}.$$

For given $x\in N$ consider any $x'\in\Omega$, $x'\neq x$, such that $\boldsymbol{t}(x)\in u(x')$. By choice of u, we get $x,x'\in u^{-1}(y)$ with $y:=\boldsymbol{t}(x)$. Since the inverse map u^{-1} is also maximal monotone, the set $u^{-1}(y)$ is closed and convex, containing with x and x' also the segment connecting the two points. Since $\boldsymbol{t}$ is differentiable at x, we obtain

$$0=\lim_{\substack{t\to 0\\ y\in u(x_t)}}\frac{y-\boldsymbol{t}(x)-\nabla\boldsymbol{t}(x)\cdot(x_t-x)}{|x_t-x|}=-\nabla\boldsymbol{t}(x)\cdot\xi,$$

where $x_t := (1-t)x + tx'$ for $t \in [0,1]$ and $\xi := (x'-x)/|x'-x|$. Indeed notice that $\boldsymbol{t}(x) \in u(x_t)$ for all $t \in [0,1]$. Hence $\xi \neq 0$ is an eigenvector of the $(d \times d)$-matrix $\nabla \boldsymbol{t}(x)$, to the eigenvalue zero. Since $x \in N$ was arbitrary, we obtain

$$N \subset \Big\{ x \in D \colon \det\big(\nabla \boldsymbol{t}(x)\big) = 0 \Big\} =: M.$$

Let $\nu := U(r,S)\mathcal{L}^d$. Since $\nu \ll \varrho$, we have that $\nu(\mathbb{R}^d \setminus \Omega) = 0$. Since $\mathcal{U}[\varrho,\sigma] < \infty$ implies that $U(r,S) \in \mathscr{L}^1(\mathbb{R}^d)$, we obtain $\nu(\Omega \setminus D) = 0$. Finally, the assumption $\mathcal{U}[\boldsymbol{t}|\varrho,\sigma] < \infty$ requires that $\nu(M) = 0$. We conclude that the set

$$\Sigma := (\mathbb{R}^d \setminus \Omega) \cup (\Omega \setminus D) \cup M$$

is ν-negligible, hence $\varrho(\Sigma) = 0$; see Remark 5.11. Then $\boldsymbol{t}|_{\mathbb{R}^d \setminus \Sigma}$ is injective, which implies in particular that $\boldsymbol{t}\#\sigma = (S \circ \boldsymbol{t}^{-1})\, \boldsymbol{t}\#\varrho$. Applying Lemma 5.5.3 in [4] we conclude that the equality (5.21) is true for $\boldsymbol{t}$ (with suitable modifications on sets of measure zero). We now use Lemma 5.6 to obtain the estimate

$$0 < \det\big(\nabla \boldsymbol{t}^{\mathrm{sym}}(x)\big) \leqslant \det\big(\nabla \boldsymbol{t}(x)\big) \quad \text{for } \nu\text{-a.e. } x \in \mathbb{R}^d.$$

Then inequality (5.24) follows from the definition (5.23). □

Remark 5.14. Using Remark 5.7, we can give a more precise version of (5.24):

$$\begin{aligned} &\mathcal{U}[\boldsymbol{t}\#\varrho, \boldsymbol{t}\#\sigma] - \mathcal{U}[\boldsymbol{t}|\varrho,\sigma] \qquad (5.26) \\ &\quad = -\int_{\mathbb{R}^d} P(r,S) \left(\det(\nabla \boldsymbol{t}^{\mathrm{sym}})^{1-\gamma} \int_0^1 \det(\mathbb{1} + tC)^{1-\gamma} \mathbf{T}(t,C)\, dt \right) dx, \end{aligned}$$

where $C(x) := R(x)^{-1} \nabla \boldsymbol{t}^{\mathrm{anti}}(x) R(x)^{-1}$ with $R(x) \in \mathrm{Sym}_d(\mathbb{R},>)$ such that

$$R(x)^2 = \nabla \boldsymbol{t}^{\mathrm{sym}}(x) \quad \text{for } \varrho\text{-a.e. } x \in \mathbb{R}^d.$$

For suitable $M \in \mathrm{Mat}_d(\mathbb{R})$ we defined

$$\mathbf{T}(t,M) := \mathrm{tr}\big((\mathbb{1} + tM)^{-1} M\big) \quad \text{for all } t \geqslant 0. \qquad (5.27)$$

Note that the difference (5.26) vanishes if and only if $\nabla \boldsymbol{t}^{\mathrm{anti}}(x) = 0$ for ϱ-a.e. $x \in \mathbb{R}$, i.e., if $\boldsymbol{t}$ is not only monotone, but *optimal* in the sense of Remark 2.2.

Proposition 5.15 (Existence of Minimizers)**.** *Consider some triple $(\varrho, \boldsymbol{\mu}, \sigma)$, with density $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$, velocity distribution $\boldsymbol{\mu} \in \mathscr{P}_\varrho(\mathbb{R}^{2d})$, and entropy $\sigma \in \mathscr{M}_+(\mathbb{R}^d)$. Assume that $\varrho =: r\mathcal{L}^d$, $\sigma =: \varrho S$, and $\mathcal{U}[\varrho,\sigma] < \infty$. Given any timestep $\tau > 0$, there exists a unique $\boldsymbol{t}_\tau \in C_\varrho$ that minimizes the functional*

$$\Psi_\tau[\boldsymbol{t}|\boldsymbol{\mu},\sigma] := \frac{3}{4\tau^2} \int_{\mathbb{R}^{2d}} |(x + \tau\xi) - \boldsymbol{t}(x)|^2\, \boldsymbol{\mu}(dx,d\xi) + \mathcal{U}[\boldsymbol{t}|\varrho,\sigma] \qquad (5.28)$$

with $\boldsymbol{t} \in C_\varrho$. This minimum is finite, which implies in particular that $\mathcal{U}[\boldsymbol{t}_\tau|\varrho,\sigma] < \infty$. For all Borel maps $\boldsymbol{v} \colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ with the property that $\boldsymbol{t}_\tau + \varepsilon\boldsymbol{v} \in C_\varrho$ for some $\varepsilon > 0$, we have the following inequality: let $P(r,S) := U'(r,S)r - U(r,S)$ for $r,S \geqslant 0$ (where $'$ denotes differentiation with respect to r). Then

$$\begin{aligned} &-\frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x + \tau\xi) - \boldsymbol{t}_\tau(x), \boldsymbol{v}(x) \rangle\, \boldsymbol{\mu}(dx,d\xi) \qquad (5.29) \\ &\quad - \int_{\mathbb{R}^d} P\big(r(x),S(x)\big)\ \det\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma} \mathrm{tr}\Big(\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1} \nabla \boldsymbol{v}(x)\Big)\, dx \geqslant 0. \end{aligned}$$

In particular, inequality (5.29) is true for $\boldsymbol{v} \in C_\varrho$ since C_ϱ is a convex cone.

Proof. We proceed in three steps.

Step 1. We observe first that the infimum $\beta := \inf_{\boldsymbol{t}\in C_\varrho} \Psi_\tau[\boldsymbol{t}|\boldsymbol{\mu},\sigma]$ is non-negative. Furthermore β is finite because we may choose $\boldsymbol{t} = \mathrm{id} \in C_\varrho$ to obtain

$$0 \leqslant \beta \leqslant \frac{3}{4}\int_{\mathbb{R}^{2d}} |\xi|^2\, \boldsymbol{\mu}(dx,d\xi) + \mathcal{U}[\varrho,\sigma] < \infty.$$

We consider a sequence of $\boldsymbol{t}^k \in C_\varrho$ such that $\Psi_\tau[\boldsymbol{t}^k|\boldsymbol{\mu},\sigma] \longrightarrow \beta$ as $k\to\infty$. Without loss of generality, we may assume that $\Psi_\tau[\boldsymbol{t}^k|\boldsymbol{\mu},\sigma] \leqslant \beta+1$ for all $k\in\mathbb{N}$. Then

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\boldsymbol{t}^k(x)|^2\,\varrho(dx) \\
&\qquad \leqslant 2\int_{\mathbb{R}^{2d}} |(x+\tau\xi)-\boldsymbol{t}^k(x)|^2\,\boldsymbol{\mu}(dx,d\xi) + 2\int_{\mathbb{R}^{2d}} |x+\tau\xi|^2\,\boldsymbol{\mu}(dx,d\xi) \\
&\qquad \leqslant \frac{8\tau^2}{3}(\beta+1) + 4\bigg\{\int_{\mathbb{R}^d} |x|^2\,\varrho(dx) + \tau^2\int_{\mathbb{R}^{2d}} |\xi|^2\,\boldsymbol{\mu}(dx,d\xi)\bigg\} < \infty.
\end{aligned}$$

Therefore the sequence $\{\boldsymbol{t}^k\}_k$ is precompact with respect to weak convergence in $\mathscr{L}^2(\mathbb{R}^d,\varrho)$: there exists a subsequence (still denoted by $\{\boldsymbol{t}^k\}_k$) and $\boldsymbol{t}\in\mathscr{L}^2(\mathbb{R}^d,\varrho)$ such that $\boldsymbol{t}^k \rightharpoonup \boldsymbol{t}$ weakly. By Mazur's lemma, there exists a map $K\colon \mathbb{N}\longrightarrow\mathbb{N}$ with $K(n)\geqslant n$ for all $n\in\mathbb{N}$, and a sequence of non-negative numbers

$$\{\lambda^n_k \colon k = n\ldots K(n)\}$$

with $\sum_{k=n}^{K(n)} \lambda^n_k = 1$, with the property that

$$\boldsymbol{s}^n := \sum_{k=n}^{K(n)} \lambda^n_k \boldsymbol{t}^k \longrightarrow \boldsymbol{t} \quad \text{strongly in } \mathscr{L}^2(\mathbb{R}^d,\varrho)$$

as $n\to\infty$. Notice that $\boldsymbol{s}^n\in C_\varrho$ since C_ϱ is a convex cone. We apply Proposition 5.4 (the convexity of the quadratic term in (5.28) is easy to check) to estimate

$$\beta \leqslant \Psi_\tau[\boldsymbol{s}^n|\boldsymbol{\mu},\sigma] \leqslant \sum_{k=n}^{K(n)} \lambda^n_k \Psi_\tau[\boldsymbol{t}^k|\boldsymbol{\mu},\sigma] \longrightarrow \beta.$$

Consequently, we obtain a strongly convergent minimizing sequence. Without loss of generality, we may assume that $\Psi_\tau[\boldsymbol{s}^n|\boldsymbol{\mu},\sigma] \leqslant \beta+1$ for all $n\in\mathbb{N}$. Extracting another subsequence if necessary, we may even assume the existence of a Borel set $N\subset\mathbb{R}^d$ with $\varrho(\mathbb{R}^d\setminus N) = 0$ such that $\boldsymbol{s}^n(x)\longrightarrow\boldsymbol{t}(x)$ for all $x\in\mathbb{R}^d\setminus N$.

Step 2. It remains to establish the lower semicontinuity of the functional (5.28). The quadratic part is clearly lower semicontinuous with respect to weak convergence in $\mathscr{L}^2(\mathbb{R}^d,\varrho)$. For the internal energy part, we will prove that the sequence $\{\boldsymbol{s}^n\}_n$ is weak* precompact in $\mathrm{BV}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$. Then we apply Proposition 5.4.

For all $m\in\mathbb{N}$, we define the convex compact sets

$$\Omega_m := \Big\{x\in\mathbb{R}^d\colon |x|\leqslant m \text{ and } \mathrm{dist}(x,\mathbb{R}^d\setminus\Omega)\geqslant 1/m\Big\}.$$

Then $\bigcup_{m\in\mathbb{N}}\Omega_m = \Omega$. Let us fix m for the moment. For each $x\in\Omega_{m+1}$ there exist finitely many points in Ω with the property that x is in the interior of the convex hull of these points. Therefore we can even find an open ball centered at x that is contained in the convex hull of these points. The collection of balls obtained in this way form an open covering of Ω_{m+1}. By compactness of Ω_{m+1}, we may choose a finite subcovering. This proves the following statement: there exist finitely many points $x_m^i \in \Omega$, $i = 1 \dots I_m$ for some $I_m \in \mathbb{N}$, with the property that

$$\Omega_{m+1} \subset \operatorname{conv} X_m, \quad \text{where} \quad X_m := \{x_m^i \colon i = 1 \dots I_m\}.$$

By adapting the argument in the proof of Lemma 3.4, we can write each $x_m^i \in X_m$ as a convex combination of points $z_m^{i,j} \in \Omega \setminus N$ with $j = 1 \dots J_m^i$ for some $J_m^i \in \mathbb{N}$. Recall that $\varrho(\mathbb{R}^d \setminus N) = 0$ and $\boldsymbol{s}^n(x) \longrightarrow \boldsymbol{t}(x)$ for all $x \in \mathbb{R}^d \setminus N$. Thus

$$\Omega_{m+1} \subset \operatorname{conv} Z_m, \quad \text{where} \quad Z_m := \{z_m^{i,j} \colon j = 1 \dots J_m^i, i = 1 \dots I_m\}. \tag{5.30}$$

Since the sequence $\{\boldsymbol{s}^n(z_m^{i,j})\}_n$ converges, it must be bounded. Let

$$\beta_m^n := \max_{i=1\dots I_m} \max_{j=1\dots J_m^i} |\boldsymbol{s}^n(z_m^{i,j})|.$$

Then $\{\beta_m^n\}_n$ is uniformly bounded for every $m \in \mathbb{N}$. We now observe that

$$\sup_{x \in \Omega_m} |\boldsymbol{s}^n(x)| \leqslant \frac{\beta_m^n \operatorname{diam}(Z_m)}{\operatorname{dist}(\Omega_m, \mathbb{R}^d \setminus \Omega_{m+1})},$$

which is bounded uniformly in n; see Proposition 1.2 in [1] and (5.30). We conclude that $\{\boldsymbol{s}^n\}_n$ is uniformly bounded in $\mathscr{L}^\infty(\Omega_m; \mathbb{R}^d)$ for all $m \in \mathbb{N}$. Since

$$\int_{\Omega_m} |D\boldsymbol{s}^n| \leqslant c_d \operatorname{diam}(\Omega_m)^{d-1} \operatorname{osc}(\boldsymbol{s}^n, \Omega_m), \tag{5.31}$$

where $c_d > 0$ is a constant depending only on the space dimenension, and where

$$\operatorname{osc}(\boldsymbol{s}^n, A) := \sup_{x_1, x_2 \in A} |\boldsymbol{s}^n(x_1) - \boldsymbol{s}^n(x_2)| \quad \text{for all } A \subset \mathbb{R}^d$$

denotes the oscillation of $\boldsymbol{s}^n$ over A, we obtain that the sequence $\{\boldsymbol{s}^n\}_n$ is uniformly bounded in $\mathrm{BV}(\Omega_m; \mathbb{R}^d)$ for all $m \in \mathbb{N}$, thus precompact in $\mathrm{BV}_{\mathrm{loc}}(\Omega; \mathbb{R}^d)$. We refer the reader to Proposition 5.1 and Remark 5.2 in [1] for a proof of (5.31).

Extracting another subsequence if necessary (not relabeled), we find that $\boldsymbol{s}^n \rightharpoonup \boldsymbol{s}$ weak* in $\mathrm{BV}_{\mathrm{loc}}(\Omega; \mathbb{R}^d)$ for a suitable function $\boldsymbol{s} \in \mathrm{BV}_{\mathrm{loc}}(\Omega; \mathbb{R}^d)$. One can now check that $\boldsymbol{s}$ is again a monotone map on Ω (possibly after redefining $\boldsymbol{s}$ on a set of measure zero). Moreover, we have $\boldsymbol{s}(x) = \boldsymbol{t}(x)$ for ϱ-a.e. $x \in \Omega$, by construction. Defining $\boldsymbol{t}_\tau(x) := \boldsymbol{s}(x)$ for $x \in \Omega$, and $\boldsymbol{t}_\tau(x) := 0$ for $x \in \mathbb{R}^d \setminus \Omega$, we have that

$$\boldsymbol{t}_\tau \in C_\varrho \quad \text{and} \quad \Psi_\tau[\boldsymbol{t}_\tau|\boldsymbol{\mu}, \sigma] \leqslant \liminf_{n \to \infty} \Psi_\tau[\boldsymbol{s}^n|\boldsymbol{\mu}, \sigma].$$

In particular, we get $\Psi_\tau[\boldsymbol{t}_\tau|\boldsymbol{\mu}, \sigma] = \beta$, thus $\boldsymbol{t}_\tau$ is a minimizer. Its uniqueness follows from the strict convexity of the first term in (5.28), which is quadratic in $\boldsymbol{t}$.

Step 3. Consider $\boldsymbol{v} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$ such that $\boldsymbol{t}_\tau + \varepsilon\boldsymbol{v} \in C_\varrho$ for $\varepsilon > 0$ small. Since $\boldsymbol{t}_\tau \in C_\varrho$, we have that $\boldsymbol{v} \in \mathrm{BV}_{\mathrm{loc}}(\Omega; \mathbb{R}^d)$ as well; see Definition 5.9. Then

$$\Psi_\tau[\boldsymbol{t}_\tau + \varepsilon\boldsymbol{v}|\boldsymbol{\mu}, \sigma] - \Psi_\tau[\boldsymbol{t}_\tau|\boldsymbol{\mu}, \sigma] \geqslant 0.$$

We divide by $\varepsilon > 0$ and consider the limit $\varepsilon \to 0$. We obtain that

$$\begin{aligned} \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \bigg\{ \frac{3}{4\tau^2} \int_{\mathbb{R}^{2d}} \big|(x + \tau\xi) - \big(\boldsymbol{t}_\tau(x) + \varepsilon\boldsymbol{v}(x)\big)\big|^2 \, \boldsymbol{\mu}(dx, d\xi) \\ - \frac{3}{4\tau^2} \int_{\mathbb{R}^{2d}} |(x + \tau\xi) - \boldsymbol{t}_\tau(x)|^2 \, \boldsymbol{\mu}(dx, d\xi) \bigg\} \\ = -\frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x + \tau\xi) - \boldsymbol{t}_\tau(x), \boldsymbol{v}(x) \rangle \, \boldsymbol{\mu}(dx, d\xi). \end{aligned}$$

Since $\mathcal{U}[\boldsymbol{t}_\tau|\varrho,\sigma]<\infty$, we can further write

$$\begin{aligned}\lim_{\varepsilon\to 0+} & \frac{\mathcal{U}[\boldsymbol{t}_\tau+\varepsilon\boldsymbol{v}|\varrho,\sigma]-\mathcal{U}[\boldsymbol{t}_\tau|\varrho,\sigma]}{\varepsilon} \\ &= \lim_{\varepsilon\to 0+}\int_{\mathbb{R}^d} U\big(r(x),S(x)\big)\ \frac{1}{\varepsilon}\Big\{\det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)+\varepsilon\nabla\boldsymbol{v}^{\mathrm{sym}}(x)\big)^{1-\gamma} \\ &\qquad\qquad -\det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\Big\}\,dx.\end{aligned}$$

We can restrict the integration to Ω where $\nabla\boldsymbol{t}_\tau$, $\nabla\boldsymbol{v}$ are well-defined; see Remark 5.11. Since $A\mapsto\det(A^{\mathrm{sym}})^{1-\gamma}$ is convex (see Proposition 5.4), the term in curly brackets is non-decreasing for a.e. $x\in\mathbb{R}^d$. By monotone convergence, it follows that

$$\begin{aligned}\lim_{\varepsilon\to 0+} & \frac{\mathcal{U}[\boldsymbol{t}_\tau+\varepsilon\boldsymbol{v}|\varrho,\sigma]-\mathcal{U}[\boldsymbol{t}_\tau|\varrho,\sigma]}{\varepsilon} \\ &= -\int_{\mathbb{R}^d} P\big(r(x),S(x)\big)\ \det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\mathrm{tr}\Big(\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1}\nabla\boldsymbol{v}^{\mathrm{sym}}(x)\Big)\,dx.\end{aligned}$$

We now can replace $\nabla\boldsymbol{v}^{\mathrm{sym}}(x)$ by $\nabla\boldsymbol{v}(x)$ since the antisymmetric part of the derivative cancels in the inner product with a symmetric matrix. □

Remark 5.16. Instead of using Mazur's lemma to get strong $\mathscr{L}^2(\mathbb{R}^d,\varrho)$-convergence (and thus convergence pointwise a.e., up to a subsequence), in Step 2 we can also use narrow convergence of the transport plans $(\mathrm{id},\boldsymbol{t}^k)\#\varrho$ together with Kuratowski convergence of their supports; see Proposition 5.1.8 in [4].

Remark 5.17. Since $\mathcal{U}[\boldsymbol{t}_\tau|\varrho,\sigma]<\infty$, we can apply Lemma 5.13 to conclude that $\boldsymbol{t}_\tau$ is essentially injective and $\mathcal{U}[\varrho_\tau,\sigma_\tau]<\infty$, where $(\varrho_\tau,\sigma_\tau):=\boldsymbol{t}_\tau\#(\varrho,\sigma)$. It follows that ϱ_τ must be absolutely continuous with respect to the Lebesgue measure and $\sigma_\tau=\varrho_\tau S_\tau$ with transported entropy $S_\tau:=S\circ\boldsymbol{t}_\tau^{-1}$; recall Definition 1.1.

Remark 5.18. Using the test functions $\boldsymbol{v}=\pm\boldsymbol{t}_\tau$ in (5.29), we obtain

$$\begin{aligned}\frac{3}{2\tau^2}&\int_{\mathbb{R}^{2d}}\langle(x+\tau\xi)-\boldsymbol{t}_\tau(x),\boldsymbol{t}_\tau(x)\rangle\,\boldsymbol{\mu}(dx,d\xi) \\ &+\int_{\mathbb{R}^d}P\big(r(x),S(x)\big)\ \det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\mathrm{tr}\Big(\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1}\nabla\boldsymbol{t}_\tau(x)\Big)\,dx=0.\end{aligned}\tag{5.32}$$

This is the analogue of equality (4.4) from the pressureless case. As a consequence, we can rewrite (5.29) in the following form (cf. (4.5)): for all $\boldsymbol{s}\in C_\varrho$ we have

$$\begin{aligned}\frac{3}{2\tau^2}&\int_{\mathbb{R}^{2d}}\langle(x+\tau\xi)-\boldsymbol{t}_\tau(x),\boldsymbol{s}(x)\rangle\,\boldsymbol{\mu}(dx,d\xi) \\ &+\int_{\mathbb{R}^d}P\big(r(x),S(x)\big)\ \det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\mathrm{tr}\Big(\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1}\nabla\boldsymbol{s}(x)\Big)\,dx\leqslant 0.\end{aligned}\tag{5.33}$$

Using in (5.33) the constant maps $\boldsymbol{s}(x)=\pm b$ for all $x\in\mathbb{R}^d$, where $b\in\mathbb{R}^d$ is some vector, we conclude that the minimization in Proposition 5.15 again preserves the total momentum; see Remark 4.13 for more details. Similarly, using $\boldsymbol{s}(x):=\pm Ax$ with $A\in\mathrm{Skew}_d(\mathbb{R})$, we obtain global conservation of angular momentum. Notice that in this case, the trace in (5.33) vanishes since $(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x))^{-1}$ is symmetric.

Remark 5.19. In (5.32) we can replace $\nabla \boldsymbol{t}_\tau(x)$ by the deformation $\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)$ since the antisymmetric part cancels in the trace. By Cramer's rule, we obtain

$$\begin{aligned} &-\frac{3}{2\tau^2}\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \boldsymbol{t}_\tau(x), \boldsymbol{t}_\tau(x)\rangle\, \boldsymbol{\mu}(dx,d\xi) \qquad (5.34)\\ &\qquad = d\int_{\mathbb{R}^d} P\big(r(x),S(x)\big)\ \det\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\,dx = d(\gamma-1)\,\mathcal{U}[\boldsymbol{t}_\tau|\varrho,\sigma]. \end{aligned}$$

Definition 5.20. For $(\varrho,\boldsymbol{\mu},\sigma,\tau)$ as in Proposition 5.15, let $\boldsymbol{t}_\tau$ denote the unique minimizer considered there. We define $\mathbf{t}_\tau, \mathbf{w}_\tau, \mathbf{u}_\tau \in \mathscr{L}^2(\mathbb{R}^{2d},\boldsymbol{\mu})$ as follows:

$$\mathbf{t}_\tau(x,\xi) := \boldsymbol{t}_\tau(x), \quad \mathbf{u}_\tau(x,\xi) := \mathbf{w}_\tau(x,\xi) := V_\tau\big(x,\xi,\boldsymbol{t}_\tau(x)\big) \qquad (5.35)$$

for $\boldsymbol{\mu}$-a.e. $(x,\xi)\in\mathbb{R}^{2d}$, with V_τ given by (1.16). Then

$$(\varrho_\tau,\sigma_\tau) := \boldsymbol{t}_\tau\#(\varrho,\sigma), \quad \boldsymbol{\mu}_\tau := (\mathbf{t}_\tau,\mathbf{u}_\tau)\#\boldsymbol{\mu}.$$

Remark 5.21. The definition of $\mathbf{t}_\tau$ in (5.35) is natural in view of Proposition 5.8. If $\boldsymbol{\mu} = (\mathrm{id},\boldsymbol{u})\#\varrho$ for some $\boldsymbol{u}\in\mathscr{L}^2(\mathbb{R}^d,\varrho)$ and $\boldsymbol{\mu}_* := (\mathbf{t}_\tau,\mathbf{w}_\tau)\#\boldsymbol{\mu}$, then

$$\int_{\mathbb{R}^{2d}} \varphi(z,\zeta)\,\boldsymbol{\mu}_*(dz,d\zeta) = \int_{\mathbb{R}^d} \varphi\big(\boldsymbol{t}_\tau(x),W(x)\big)\,\varrho(dx)$$

for all $\varphi\in\mathscr{C}_{\mathrm{b}}(\mathbb{R}^{2d})$, with $W := \frac{3}{2}V - \frac{1}{2}\boldsymbol{u}$ and $V := (\boldsymbol{t}_\tau - \mathrm{id})/\tau$. Let

$$\boldsymbol{u}_\tau(z) := W\big(\boldsymbol{t}_\tau^{-1}(z)\big) \quad \text{for } \varrho_\tau\text{-a.e. } z\in\mathbb{R}^d. \qquad (5.36)$$

The velocity $\boldsymbol{u}_\tau$ is well-defined because $\boldsymbol{t}_\tau$ is essentially injective; see Remark 5.17. It follows that $\boldsymbol{\mu}_* = (\mathrm{id},\boldsymbol{u}_\tau)\#\varrho_\tau$ and $\boldsymbol{u}_\tau\in\mathscr{L}^2(\mathbb{R}^d,\varrho_\tau)$. We would like to emphasize the fact that the minimization preserves the monokinetic structure of the fluid (recall that the velocity update in (5.35) is a consequence of the minimization of the work functional). Since the tangent cone over the cone of monotone maps at ϱ_τ equals $\mathscr{L}^2(\mathbb{R}^d,\varrho_\tau)$, no additional projection is necessary (unlike in the pressureless gas case; see Step (2) in Definition 4.9). We can therefore put $\mathbf{u}_\tau = \mathbf{w}_\tau$.

Proposition 5.22 (Stress Tensor)**.** *Suppose that $\tau>0$ and $(\varrho,\boldsymbol{\mu},\sigma)$ are given, with density $\varrho\in\mathscr{P}_2(\mathbb{R}^d)$, velocity distribution $\boldsymbol{\mu}\in\mathscr{P}_\varrho(\mathbb{R}^{2d})$, and entropy $\sigma\in\mathscr{M}_+(\mathbb{R}^d)$. Assume that $\varrho =: r\mathcal{L}^d$, $\sigma =: \varrho S$, and $\mathcal{U}[\varrho,\sigma]<\infty$. Consider the unique minimizer $\boldsymbol{t}_\tau\in C_\varrho$ from Proposition 5.15. There exists $\mathbf{R}_\tau\in\mathscr{M}(\mathbb{R}^d;\mathrm{Sym}_d(\mathbb{R},\geqslant))$ with*

$$\begin{aligned} \int_{\mathbb{R}^d} \mathrm{tr}\big(\nabla u(x)\mathbf{R}_\tau(dx)\big) &= -\frac{3}{2\tau^2}\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi)-\boldsymbol{t}_\tau(x), u(x)\rangle\,\boldsymbol{\mu}(dx,d\xi)\\ &\quad -\int_{\mathbb{R}^d} P\big(r(x),S(x)\big)\ \det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\mathrm{tr}\Big(\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1}\nabla u(x)\Big)\,dx \qquad (5.37) \end{aligned}$$

for all $u\in\mathscr{C}^1_*(\mathbb{R}^d;\mathbb{R}^d)$. In particular, we have the control

$$\begin{aligned} \int_{\mathbb{R}^d} \mathrm{tr}\big(\mathbf{R}_\tau(dx)\big) &= -\frac{3}{2\tau^2}\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi)-\boldsymbol{t}_\tau(x), x\rangle\,\boldsymbol{\mu}(dx,d\xi)\\ &\quad -\int_{\mathbb{R}^d} P\big(r(x),S(x)\big)\ \det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\mathrm{tr}\Big(\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1}\Big)\,dx. \qquad (5.38) \end{aligned}$$

Proof. Since every $u\in\mathrm{Mon}(\mathbb{R}^d)$ has at most linear growth, we have $u\in\mathscr{L}^2(\mathbb{R}^d,\varrho)$. Thus $\mathrm{Mon}(\mathbb{R}^d)\subset C_\varrho$ and $\boldsymbol{v} := u\in\mathrm{Mon}(\mathbb{R}^d)$ is admissible in (5.29). Let

$$\mathbf{P}(dx) := P\big(r(x),S(x)\big)\ \det\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\big(\nabla\boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1}\,dx.$$

The inverse matrix $(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x))^{-1}$ is symmetric and positive definite for a.e. $x \in \Omega$ because $\boldsymbol{t}_\tau$ is monotone there. Consequently, its norm can be controlled by the trace. Using $\boldsymbol{v} = \mathrm{id}$ (which is an element of C_ϱ) in (5.29), we obtain the estimate

$$0 \leqslant \int_{\mathbb{R}^d} \mathrm{tr}\big(\mathbf{P}(dx)\big) \leqslant -\frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \boldsymbol{t}_\tau(x), x\rangle \, \boldsymbol{\mu}(dx, d\xi),$$

which is finite. Thus $\mathbf{P} \in \mathscr{M}(\mathbb{R}^d; \mathrm{Sym}_d(\mathbb{R}, \geqslant))$. If we define

$$\mathbf{F}(dx) := -\frac{3}{2\tau^2} \Big(\big(x + \tau \boldsymbol{u}(x)\big) - \boldsymbol{t}_\tau(x) \Big) \, \varrho(dx),$$

where $\boldsymbol{u} := \mathrm{b}(\boldsymbol{\mu})$ denotes the barycentric projection of $\boldsymbol{\mu}$ (which is in $\mathscr{L}^2(\mathbb{R}^d, \varrho)$), then $\mathbf{F} \in \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)$ has finite first moment because $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$. We then apply Theorem 4.14 to obtain the representation (5.37)/(5.38); see also Remark 4.15. ☐

Proposition 5.23 (Energy Balance). *Let $\tau > 0$ and $(\varrho, \boldsymbol{u}, \sigma)$ are given, with density $\varrho \in \mathscr{P}_2(\mathbb{R}^d)$, Eulerian velocity $\boldsymbol{u} \in \mathscr{L}^2(\mathbb{R}^d, \varrho)$, and entropy $\sigma \in \mathscr{M}_+(\mathbb{R}^d)$. Suppose that $\varrho =: r\mathcal{L}^d$, $\sigma =: \varrho S$, and $\mathcal{U}[\varrho, \sigma] < \infty$. Let $\boldsymbol{t}_\tau \in C_\varrho$ denote the unique minimizer from Proposition 5.15 (where $\boldsymbol{\mu} := (\mathrm{id}, \boldsymbol{u})\#\varrho$) and $\mathbf{R}_\tau \in \mathscr{M}(\mathbb{R}^d; \mathrm{Sym}_d(\mathbb{R}, \geqslant))$ the stress tensor field in Proposition 5.22. Consider $(\varrho_\tau, \boldsymbol{u}_\tau, \sigma_\tau)$ and $\boldsymbol{w}_\tau$ as defined in the Remarks 5.17/5.21. Then the following energy equality holds:*

$$\begin{aligned} \mathcal{E}[\varrho_\tau, \boldsymbol{u}_\tau, \sigma_\tau] + \int_{\mathbb{R}^d} \tfrac{1}{6} \varrho |\boldsymbol{w}_\tau - \boldsymbol{u}|^2 & \qquad (5.39) \\ + \int_{\mathbb{R}^d} \Big(P(r,S)\, \mathbf{D}^2\big(\nabla \boldsymbol{t}_\tau - \mathbb{1}\big)\Big)\, dx + \mathrm{tr}\big(\mathbf{R}_\tau(dx)\big) \Big) &= \mathcal{E}[\varrho, \boldsymbol{u}, \sigma], \end{aligned}$$

with total energy (recall Definition 1.1)

$$\mathcal{E}[\varrho, \boldsymbol{u}, \sigma] := \int_{\mathbb{R}^d} \tfrac{1}{2} \varrho |\boldsymbol{u}|^2 + \mathcal{U}[\varrho, \sigma].$$

For all matrices $\mathbb{1} + S \in \mathrm{Sym}_d(\mathbb{R}, >)$ and $A \in \mathrm{Skew}_d(\mathbb{R})$ we have

$$\begin{aligned} \mathbf{D}^2(S + A) := \int_0^1 & \det(\mathbb{1} + tS)^{1-\gamma} \Big((\gamma - 1) \mathbf{T}(t, S)^2 + \mathbf{T}_2(t, S) \Big) t \, dt \\ & + \det(\mathbb{1} + S)^{1-\gamma} \int_0^1 \det(\mathbb{1} + tC)^{1-\gamma} \mathbf{T}(t, C)\, dt \geqslant 0. \end{aligned}$$

Here $C := R^{-1} A R^{-1}$ and $R \in \mathrm{Sym}_d(\mathbb{R}, >)$ is uniquely determined by $\mathbb{1} + S =: R^2$. Recall (5.27) *for the definition of* $\mathbf{T}$. *For suitable $M \in \mathrm{Mat}_d(\mathbb{R})$ we define*

$$\mathbf{T}_2(t, M) := \mathrm{tr}\Big(\big((\mathbb{1} + tM)^{-1} M \big)^2 \Big) \quad \textit{for all } t \geqslant 0.$$

Notice that all terms in curly brackets in (5.39) *are non-negative.*

Proof. Let us first consider the kinetic energy. Because of (4.14)/(5.36), we have

$$\begin{aligned} \int_{\mathbb{R}^d} \tfrac{1}{2} |\boldsymbol{u}_\tau(x)|^2 \, \varrho_\tau(dx) &+ \frac{1}{6} \int_{\mathbb{R}^d} |\boldsymbol{w}_\tau(x) - \boldsymbol{u}(x)|^2 \, \varrho(dx) \\ &= \int_{\mathbb{R}^d} \tfrac{1}{2} |\boldsymbol{u}(x)|^2 \, \varrho(dx) - \frac{3}{2\tau^2} \int_{\mathbb{R}^{2d}} \langle (x + \tau\xi) - \boldsymbol{t}_\tau(x), \boldsymbol{t}_\tau(x) - x \rangle \, \boldsymbol{\mu}(dx, d\xi) \end{aligned}$$

Combining (5.32) with the representation (5.38), we find that

$$\begin{aligned}
&- \frac{3}{2\tau^2}\int_{\mathbb{R}^{2d}} \langle (x+\tau\xi) - \boldsymbol{t}_\tau(x), \boldsymbol{t}_\tau(x) - x\rangle\, \boldsymbol{\mu}(dx,d\xi) \qquad (5.40)\\
&\quad = \tau \int_{\mathbb{R}^d} P\big(r(x), S(x)\big)\ \det\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma} \mathrm{tr}\Big(\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1} \nabla \boldsymbol{v}_\tau(x)\Big)\, dx\\
&\qquad - \int_{\mathbb{R}^d} \mathrm{tr}\big(\mathbf{R}_\tau(dx)\big).
\end{aligned}$$

Let $\boldsymbol{t}(s,x) := x + s\tau \boldsymbol{v}_\tau(x)$ for $s\in[0,1]$. Taylor expanding around $s=1$, we get

$$\begin{aligned}
&\det\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma} = 1\\
&\quad - \tau(\gamma-1)\det\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{1-\gamma}\ \mathrm{tr}\Big(\big(\nabla \boldsymbol{t}_\tau^{\mathrm{sym}}(x)\big)^{-1} \nabla \boldsymbol{v}_\tau(x)\Big) \qquad (5.41)\\
&\quad - \int_0^1 \det\big(\nabla \boldsymbol{t}^{\mathrm{sym}}(s,x)\big)^{1-\gamma} \bigg\{ (\gamma-1)^2 \Big(\mathrm{tr}\Big(\big(\nabla \boldsymbol{t}^{\mathrm{sym}}(s,x)\big)^{-1} \tau \nabla \boldsymbol{v}_\tau^{\mathrm{sym}}(x)\Big)\Big)^2\\
&\qquad\qquad + (\gamma-1)\,\mathrm{tr}\Big(\Big(\big(\nabla \boldsymbol{t}^{\mathrm{sym}}(s,x)\big)^{-1} \tau \nabla \boldsymbol{v}_\tau^{\mathrm{sym}}(x)\Big)^2\Big)\bigg\} s\, ds
\end{aligned}$$

for a.e. $x\in\Omega$. We now multiply by $U(r(x),S(x))$ and integrate in $x\in\mathbb{R}^d$. Then the integral of (5.41) equals the negative of the first term on the right-hand side of (5.40). Combining all terms and using Remark 5.14, we conclude the proof. □

Remark 5.24 (Bregman Divergence). We observe that the function

$$\begin{aligned}
D_{\mathcal{U}}(S) &:= \Big(1 - \det(\mathbb{1}+S)^{1-\gamma}\Big) - (\gamma-1)\det(\mathbb{1}+S)^{1-\gamma}\mathrm{tr}\Big((\mathbb{1}+S)^{-1}S\Big) \qquad (5.42)\\
&= \int_0^1 \det(\mathbb{1}+tS)^{1-\gamma}\Big((\gamma-1)\mathbf{T}(t,S)^2 + \mathbf{T}_2(t,S)\Big) t\, dt \geqslant 0,
\end{aligned}$$

defined for every $S\in\mathrm{Sym}_d(\mathbb{R})$ with $\mathbb{1}+S$ positive definite, is the Bregman divergence for 0 and S associated to the convex function $S\mapsto \det(\mathbb{1}+S)^{1-\gamma}$.

The following result will be useful to control the momentum equation of (1.1).

Lemma 5.25. *For every $\varepsilon>0$ there exists a constant $C_\varepsilon>0$ with the following property: For all $S\in\mathrm{Sym}_d(\mathbb{R})$ with $\mathbb{1}+S$ positive definite, we have*

$$\sup_{z\in\mathbb{R}^d, |z|=1} \big|\big\langle z, \big(\mathbb{1} - \det(\mathbb{1}+S)^{1-\gamma}(\mathbb{1}+S)^{-1}\big)z\big\rangle\big| \leqslant \varepsilon + C_\varepsilon D_{\mathcal{U}}(S), \qquad (5.43)$$

where $D_{\mathcal{U}}$ is defined in (5.42).

Similar, for any $\varepsilon>0$ there exists $C_\varepsilon>0$ such that

$$\sup_{z\in\mathbb{R}^d, |z|=1} \big|\big\langle z, \big(\boldsymbol{w}\otimes(\boldsymbol{v}-\boldsymbol{w})\big)z\big\rangle\big| \leqslant \varepsilon|\boldsymbol{w}|^2 + C_\varepsilon D_{\mathcal{K}}|\boldsymbol{w}-\boldsymbol{u}|^2 \qquad (5.44)$$

for all $\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^d$ and $\boldsymbol{u} := 3\boldsymbol{v} - 2\boldsymbol{w}$.

Proof. Notice first that the map $S\mapsto \mathbb{1} - \det(\mathbb{1}+S)^{1-\gamma}(\mathbb{1}+S)^{-1}$ vanishes for $S=0$ and is continuous there. Consequently, for any $\varepsilon>0$ there exists a $\delta>0$ such that the left-hand side of (5.43) is less than ε for all $S\in\mathrm{Sym}_d(\mathbb{R})$ with $\|S\|<\delta$.

For any $S \in \mathrm{Sym}_d(\mathbb{R})$ with $\mathbb{1} + S$ positive definite, we rewrite

$$\begin{aligned}&\mathbb{1} - \det(\mathbb{1}+S)^{1-\gamma}(\mathbb{1}+S)^{-1}\\ &\quad = \Big(1 - \det(\mathbb{1}+S)^{1-\gamma}\Big)\mathbb{1} + \det(\mathbb{1}+S)^{1-\gamma}\Big(\mathbb{1} - (\mathbb{1}+S)^{-1}\Big)\end{aligned}$$

Because of the spectral theorem, there exist real eigenvalues λ_i and a corresponding system of orthonormal eigenvalues $e_i \in \mathbb{R}^d$, $i = 1 \dots d$, such that $S = \sum_{i=1}^d \lambda_i e_i \otimes e_i$. We also have the identity $\sum_{i=1}^d e_i \otimes e_i = \mathbb{1}$. We can then write

$$\begin{aligned}&\mathbb{1} - \det(\mathbb{1}+S)^{1-\gamma}(\mathbb{1}+S)^{-1}\\ &\quad = \bigg(1 - \prod_{i=1}^d (1+\lambda_i)^{1-\gamma}\bigg)\mathbb{1} + \prod_{i=1}^d (1+\lambda_i)^{1-\gamma} \sum_{i=1}^d \frac{\lambda_i}{1+\lambda_i} e_i \otimes e_i.\end{aligned}$$

Multiplying from left and right by a vector $z \in \mathbb{R}^d$ with $|z| = 1$, we obtain

$$\begin{aligned}&\big\langle z, \big(\mathbb{1} - \det(\mathbb{1}+S)^{1-\gamma}(\mathbb{1}+S)^{-1}\big) z\big\rangle\\ &\quad = \bigg(1 - \prod_{i=1}^d (1+\lambda_i)^{1-\gamma}\bigg) + \prod_{i=1}^d (1+\lambda_i)^{1-\gamma} \sum_{i=1}^d c_i^2 \frac{\lambda_i}{1+\lambda_i},\end{aligned}$$

where $c_i := z \cdot e_i$ and $\sum_{i=1}^d c_i^2 = 1$. Similarly, we can rewrite (5.42) as

$$\begin{aligned}D_{\mathcal{U}}(S) &= \bigg(1 - \prod_{i=1}^d (1+\lambda_i)^{1-\gamma}\bigg) - (\gamma-1)\prod_{i=1}^d (1+\lambda_i)^{1-\gamma} \sum_{i=1}^d \frac{\lambda_i}{1+\lambda_i}.\\ &= \int_0^1 \prod_{i=1}^d (1+t\lambda_i)^{1-\gamma} \Bigg\{(\gamma-1)\bigg(\sum_{i=1}^d \frac{\lambda_i}{1+t\lambda_i}\bigg)^2 + \sum_{i=1}^d \bigg(\frac{\lambda_i}{1+t\lambda_i}\bigg)^2\Bigg\} t\,dt,\end{aligned}$$

from which we conclude that $D_{\mathcal{U}}(S) = 0$ if and only if all eigenvalues λ_i vanish, thus $S = 0$. Recall that $\gamma - 1 > 0$, by assumption. In particular, we have $D_{\mathcal{U}}(S) > 0$ for all $S \in \mathrm{Sym}_d(\mathbb{R})$ such that $\|S\| \geqslant \delta$. By continuity and compactness, for any $\gamma < 1$ there exists a constant $c_\gamma > 0$ with $D_{\mathcal{U}}(S) \geqslant c_\gamma$ for all $\gamma \geqslant \|S\| \geqslant \delta$.

To simplify the notation, we will write

$$d(\lambda) := \prod_{i=1}^d (1+\lambda_i)^{1-\gamma}, \quad S_c(\lambda) := \sum_{i=1}^d c_i^2 \frac{\lambda_i}{1+\lambda_i}$$

for all $\lambda := (\lambda_1, \dots, \lambda_d)$ with $\lambda_i > -1$ and $c := (c_1, \dots, c_d)$. We claim that

$$F(\lambda) := \frac{1 - d(\lambda)\Big(1 - S_c(\lambda)\Big)}{1 - d(\lambda)\Big(1 + (\gamma-1)S_{(1,\dots,1)}(\lambda)\Big)} \tag{5.45}$$

is uniformly bounded away from $\lambda = 0$. Then the estimate (5.43) follows.

In order to prove the claim, we first observe that the level sets of $d(\lambda)$ generate a partition of the orthant $(-1,\infty)^d$ into hyperboloids. For simplicity, we only consider the case $d = 2$. The general case can be handled similarly. We introduce a coordinate system adapted to $(-1,\infty)^2$ as follows: For all $(\alpha,\beta) \in (0,\pi/2)^2$ let

$$\lambda_1(\alpha,\beta) := \sqrt{\tan(\alpha)\cot(\beta)} - 1, \quad \lambda_2(\alpha,\beta) := \sqrt{\tan(\alpha)\tan(\beta)} - 1.$$

Notice that with this choice β parameterizes the level curves of $d(\lambda) = \tan(\alpha)^{1-\gamma}$. Expressed in these coordinates, the function (5.45) takes the form

$$F(\alpha,\beta) = \frac{1 - \tan(\alpha)^{1-\gamma}\left(c_1^2\sqrt{\frac{\tan(\beta)}{\tan(\alpha)}} + c_2^2\sqrt{\frac{\cot(\beta)}{\tan(\alpha)}}\right)}{1 - \tan(\alpha)^{1-\gamma}\left((2\gamma-1) - (\gamma-1)\left(\sqrt{\frac{\tan(\beta)}{\tan(\alpha)}} + \sqrt{\frac{\cot(\beta)}{\tan(\alpha)}}\right)\right)}, \tag{5.46}$$

where we have used that $c_1^2 + c_2^2 = 1$. For any $\alpha \in (0, \pi/2)$ fixed, we find that

$$\lim_{\beta\to 0} F(\alpha,\beta) = \frac{c_2^2}{1-\gamma}, \quad \lim_{\beta\to\pi/2} F(\alpha,\beta) = \frac{c_1^2}{1-\gamma}. \tag{5.47}$$

Similarly, for any $\beta \in (0, \pi/2)$ fixed, we have

$$\lim_{\alpha\to 0} F(\alpha,\beta) = \frac{c_1^2\sqrt{\tan(\beta)} + c_2^2\sqrt{\cot(\beta)}}{(1-\gamma)\left(\sqrt{\tan(\beta)} + \sqrt{\cot(\beta)}\right)}, \quad \lim_{\alpha\to\pi/2} F(\alpha,\beta) = 1.$$

Notice that $\lim_{\alpha\to 0} F(\alpha,\beta)$ converges to the limits in (5.47) as $\beta \to 0$ or $\pi/2$.

We now consider the limit $\alpha \to 0$ with $\beta(\alpha) := k\alpha$ for $k > 0$. We find that

$$\lim_{\alpha\to 0} F(\alpha, k\alpha) = \frac{c_2^2}{1-\gamma} \quad \text{for any } k > 0,$$

hence $F(\alpha,\beta)$ can be continuously extended to $(\alpha,\beta) = (0,0)$ by $c_2^2/(1-\gamma)$. Recall that $\tan(\theta) \approx \theta$ for small θ. Similarly, we compute the limit

$$\lim_{\beta\to 0} F(\pi/2 - k\beta, \beta) = 1 \quad \text{for any } k > 0.$$

We used that $\tan(\pi/2 - k\beta) = \cot(k\beta)$. The behavior of the map $(\alpha,\beta) \mapsto F(\alpha,\beta)$ at the other corners of the domain $(0,\pi/2)^2$ can be studied analogously. We conclude that $F(\alpha,\beta)$ remains bounded for (α,β) near the boundary of $(0,\pi/2)^2$, uniformly in $c = (c_1, c_2)$. It is continuous up to the boundary expect for the points $(0,\pi/2)$ and $(\pi/2,\pi/2)$. As long as we stay away from the unique root of the denominator in (5.46), the function F is uniformly bounded as claimed, by continuity.

The estimate (5.44) follows from Young inequality. □

6. Measure-Valued Solutions

In this section, we use the minimizations in Sections 4.2/5.3 to define approximate solutions to the compressible gas dynamics equations (1.1), for suitable initial data and timestep $\tau > 0$. We establish uniform bounds and prove that a subsequence converges to a measure-valued solution of (1.1) in the limit $\tau \to 0$. We will cover the pressureless case and the Euler case simultaneously, with the understanding that for the pressureless case the internal energy is set to zero. Similarly, the specific entropy is considered constant in all cases other than the full Euler case.

6.1. Approximate Solutions.

We will construct approximate solutions to (1.1) on time intervals $[0,\infty)$ by successively applying the variational minimization step introduced in the previous sections and then utilizing a suitable interpolation between discrete times. Consider initial density, velocity distribution, and entropy

$$\bar{\varrho} \in \mathscr{P}_2(\mathbb{R}^d), \quad \bar{\boldsymbol{\mu}} \in \mathscr{P}_{\bar{\varrho}}(\mathbb{R}^{2d}), \quad \bar{\sigma} \in \mathscr{M}_+(\mathbb{R}^d).$$

Suppose $\mathcal{U}[\bar\varrho,\bar\sigma]<\infty$ so that $\bar\varrho =: \bar r\mathcal{L}^d$ and $\bar\sigma =: \bar\varrho\bar S$ for suitable Borel functions $\bar r,\bar S$; see Definition 1.1. Assume further that $\bar{\boldsymbol\mu} =: (\mathrm{id},\bar{\boldsymbol v})\#\bar\varrho$ with

$$\bar{\boldsymbol v}\in\mathscr{L}^2(\mathbb{R}^d,\bar\varrho)\quad\text{satisfying}\quad\int_{\mathbb{R}^d}\bar{\boldsymbol v}(x)\,\bar\varrho(dx)=0. \tag{6.1}$$

Notice that since the hyperbolic conservation law (1.3) is invariant under transformations to a moving reference frame, the assumption (6.1) is not restrictive.

For later use, let us introduce the initial total energy

$$\bar{\mathcal{E}} := \int_{\mathbb{R}^d}\tfrac12\bar\varrho|\bar{\boldsymbol v}|^2+\mathcal{U}[\bar\varrho,\bar\sigma]<\infty. \tag{6.2}$$

In order to simplify the notation, in this section we will not indicate the dependence of various quantities on the timestep $\tau>0$, which will be arbitrary, but fixed for the following construction. Let $s^k := k\tau$ for all $k\in\mathbb{N}_0$. We define

$$\varrho^0:=\bar\varrho,\quad \boldsymbol\mu^0:=\bar{\boldsymbol\mu},\quad \sigma^0:=\bar\sigma.$$

Then we proceed recursively: For any $k\in\mathbb{N}_0$ we define

$$\begin{aligned}\mathbf{t}^{k+1}&:=\mathbf{t}_\tau,\quad \mathbf{w}^{k+1}:=\mathbf{w}_\tau,\quad \mathbf{u}^{k+1}:=\mathbf{u}_\tau,\\ \varrho^{k+1}&:=\varrho_\tau,\quad \boldsymbol\mu^{k+1}:=\boldsymbol\mu_\tau,\quad \sigma^{k+1}:=\sigma_\tau,\end{aligned}$$

with $(\mathbf{t}_\tau,\mathbf{w}_\tau,\mathbf{u}_\tau)$ and $(\varrho_\tau,\boldsymbol\mu_\tau,\sigma_\tau)$ taken from Definitions 4.9/5.20, for the choice

$$\varrho:=\varrho^k,\quad \boldsymbol\mu:=\boldsymbol\mu^k,\quad \sigma:=\sigma^k.$$

By induction in k, we observe first that $\boldsymbol\mu^k$ is monokinetic for every $k\in\mathbb{N}_0$. For $k=0$ this follows from our assumption on the initial data. For $k\geqslant1$ we refer the reader to Definition 4.9 and Remark 5.21, respectively. Thus

$$\boldsymbol u^k\in\mathscr{L}^2(\mathbb{R}^d,\varrho^k)\quad\text{such that}\quad \boldsymbol\mu^k=:(\mathrm{id},\boldsymbol u^k)\#\varrho^k.$$

is well-defined. Similarly, from Propositions 4.17/5.23 and (6.2), we obtain that

$$\mathcal{E}[\varrho^k,\boldsymbol u^k,\sigma^k]=\int_{\mathbb{R}^d}\tfrac12\varrho^k|\boldsymbol u^k|^2+\mathcal{U}[\varrho^k,\sigma^k]\leqslant\bar{\mathcal{E}}$$

for every $k\in\mathbb{N}_0$. Therefore the following maps are all well-defined as well:

$$\varrho^k=:r^k\mathcal{L}^d,\quad \sigma^k=:\varrho^kS^k.$$

For ϱ^k-a.e. $x\in\mathbb{R}^d$ and $k\in\mathbb{N}_0$, we now define

$$(\boldsymbol t^{k+1},W^{k+1},U^{k+1})(x):=(\mathbf{t}^{k+1},\mathbf{w}^{k+1},\mathbf{u}^{k+1})\big(x,\boldsymbol u^k(x)\big),$$

which are in $\mathscr{L}^2(\mathbb{R}^d,\varrho^k)$. Rewriting Propositions 4.17/5.23, we obtain

$$\begin{aligned}\mathcal{E}[\varrho^{k+1},\boldsymbol u^{k+1},\sigma^{k+1}]&+\int_{\mathbb{R}^d}\Big(\tfrac16|W^{k+1}-\boldsymbol u^k|^2+\tfrac12|U^{k+1}-W^{k+1}|^2\Big)\,\varrho^k(dx)\\ &+\int_{\mathbb{R}^d}\Big(P(r^k,S^k)\,\mathbf{D}^2\big(\nabla\boldsymbol t^{k+1}-\mathbb{1}\big)\Big)\,dx+\mathrm{tr}\big(\mathbf{R}^{k+1}(dx)\big)\Big)\\ &=\mathcal{E}[\rho^k,\boldsymbol u^k,\sigma^k]\quad\text{for all } k\in\mathbb{N}_0.\end{aligned} \tag{6.3}$$

Here $\mathbf{R}^{k+1}$ is the residual tensor corresponding to the minimizer $\boldsymbol t^{k+1}$.

6.2. **Interpolation in Time.** In the previous section, we introduced approximate solutions of (1.13) at discrete times $s^k := k\tau$ for any timestep $\tau > 0$. Here we want to interpolate in time to define functions/measures of time and space.

For this, we could use the path of minimal acceleration

$$Y_s(x,\xi) := x + s\xi - \left(\frac{s^2}{\tau} - \frac{s^3}{3\tau^2}\right)\frac{3}{2\tau}\Big((x+\tau\xi) - \mathbf{t}_\tau(x,\xi)\Big)$$

for location $x \in \mathbb{R}^d$, velocity $\xi \in \mathbb{R}^d$, and $s \in [0,\tau]$, suitably shifted in s. This would be the natural choice in view of the derivation of the work functional, which featured in our minimization problem. Instead we prefer to apply the convex interpolation that we have already utilized to derive the displacement convexity of the internal energy and hence the energy inequality in Proposition 5.23. One can show that the differences between both the positions and velocities of the minimal acceleration paths and the convex interpolations remain bounded and vanish as $\tau \to 0$. Let

$$\boldsymbol{t}_s(x) := x + (s - s^k)\, V^{k+1}(x), \quad V^{k+1}(x) := \frac{\boldsymbol{t}^{k+1}(x) - x}{\tau}$$

for ϱ^k-a.e. $x \in \mathbb{R}^d$ and $s \in [s^k, s^{k+1})$. Notice that $\boldsymbol{t}_s$ is strictly monotone, therefore invertible for any $s \in [s^k, s^{k+1})$ since $\boldsymbol{t}^{k+1}$ is monotone. Moreover, in the cases with pressure the map $\boldsymbol{t}^{k+1}$ is essentially injective; see Lemma 5.13. We can therefore track the path of each fluid element starting from a generic position $\bar{x} \in \mathbb{R}^d$. By composing the transport maps of successive timesteps, we define the transport/velocity

$$X_s := \boldsymbol{t}_s \circ \boldsymbol{t}^k \circ \cdots \circ \boldsymbol{t}^1, \quad \Xi_s := \begin{cases} W^{k+1} \circ \boldsymbol{t}^k \circ \cdots \circ \boldsymbol{t}^1 & \text{if } s \in (s^k, s^{k+1}), \\ \boldsymbol{u}^k \circ \boldsymbol{t}^k \circ \cdots \circ \boldsymbol{t}^1 & \text{if } s = s^k, \end{cases} \tag{6.4}$$

where $k := \lfloor s/\tau \rfloor$ (the largest integer not bigger than s/τ). Since $\boldsymbol{t}^{l+1}$ is defined only ϱ^l-a.e., we may have to discard a ϱ^l-null set, for every $l \in \mathbb{N}_0$. The preimages of these null sets under the preceding transport maps, however, can be traced back to a $\bar{\varrho}$-negligible set, hence X_s and Ξ_s are well-defined $\bar{\varrho}$-a.e. The map $s \mapsto X_s(\bar{x})$ is Lipschitz continuous for $\bar{\varrho}$-a.e. $\bar{x} \in \mathbb{R}^d$ because V^{k+1} is finite ϱ^k-a.e. To simplify the notation, let $\boldsymbol{X}_s := (X_s, \Xi_s)$ for all $s \geqslant 0$. We now define

$$(\varrho_s, \sigma_s) := \boldsymbol{t}_s\#(\varrho^k, \sigma^k), \quad \boldsymbol{\mu}_s := \begin{cases} (\boldsymbol{t}_s, W^{k+1})\#\varrho^k & \text{if } s \in (s^k, s^{k+1}), \\ \boldsymbol{\mu}^k & \text{if } s = s^k. \end{cases} \tag{6.5}$$

Using the transport/velocity (6.4), we can express $\boldsymbol{\mu}_s$ by following the characteristic lines back to the initial data. More precisely, we have $\boldsymbol{\mu}_s = \boldsymbol{X}_s\#\bar{\varrho}$ for $s \geqslant 0$.

It follows from the proof of Proposition 5.23 that in the cases with pressure

$$\mathcal{U}[\varrho_s, \sigma_s] \leqslant \mathcal{U}[\boldsymbol{t}_s|\varrho^k, \sigma^k] \leqslant \big(1 - \ell^k(s)\big)\, \mathcal{U}[\varrho^k, \sigma^k] + \ell^k(s)\, \mathcal{U}[\boldsymbol{t}^{k+1}|\varrho^k, \sigma^k]$$

for every $s \in [s^k, s^{k+1}]$ and $k \in \mathbb{N}_0$. Here $\ell^k(s) := (s - s^k)/\tau$. Applying this estimate recursively, we conclude that $\mathcal{U}[\varrho_s, \sigma_s]$ remains finite for all $s \geqslant 0$, thus

$$\varrho_s =: r_s \mathcal{L}^d, \quad \sigma_s = \varrho_s S_s \quad \text{with} \quad S_s := S^k \circ \boldsymbol{t}_s^{-1}$$

for $s \in [s^k, s^{k+1})$. The specific entropy S_s is simply transported along with the flow. The velocity distribution $\boldsymbol{\mu}_s$ is monokinetic for all $s \geqslant 0$, which defines

$$\boldsymbol{w}_s \in \mathscr{L}^2(\mathbb{R}^d, \varrho_s) \quad \text{such that} \quad \boldsymbol{\mu}_s =: (\mathrm{id}, \boldsymbol{w}_s)\#\varrho_s.$$

The case $s = s^k$ has been discussed above. For $s \in (s^k, s^{k+1})$ the claim follows from $\boldsymbol{\mu}^k$ monokinetic and the invertibility of $\boldsymbol{t}_s$. In the transition from $\boldsymbol{u}^k$ to W^{k+1} the kinetic energy may increase; it remains bounded by $\mathcal{E}[\rho^k, \boldsymbol{u}^k, \sigma^k]$. Over the course of the time interval, the total energy then decreases so that (6.3) holds. This explains the additional factor 2 on the right-hand side of the following energy bound:

$$\mathcal{E}[\varrho_s, \boldsymbol{w}_s, \sigma_s] = \int_{\mathbb{R}^d} \tfrac{1}{2}\varrho_s |\boldsymbol{w}_s|^2 + \mathcal{U}[\varrho_s, \sigma_s] \leqslant 2\bar{\mathcal{E}} \quad \text{for all } s \geqslant 0. \tag{6.6}$$

For any $s \in [s^k, s^{k+1})$, we define the transport velocity

$$\boldsymbol{v}_s := V^{k+1} \circ \boldsymbol{t}_s^{-1} \quad \text{so that} \quad \dot{X}_s = \boldsymbol{v}_s \circ X_s. \tag{6.7}$$

Because of (4.15), we have

$$\int_{\mathbb{R}^d} |\boldsymbol{v}_s(x)|^2 \, \varrho_s(dx) \leqslant \frac{2}{3} \int_{\mathbb{R}^d} |W^{k+1}(x)|^2 \, \varrho^k(dx) + \frac{1}{3} \int_{\mathbb{R}^d} |\boldsymbol{u}^k(x)|^2 \, \varrho^k(dx), \tag{6.8}$$

which can be bounded in terms of $\bar{\mathcal{E}}$ for all $s \geqslant 0$; see (6.6).

The momentum $\boldsymbol{m}_s := \varrho_s \boldsymbol{w}_s$ has zero mean: We can write

$$\int_{\mathbb{R}^d} \boldsymbol{v}_s(z) \, \varrho_s(dz) = \int_{\mathbb{R}^d} W^{k+1}(x) \, \varrho^k(dx) = \int_{\mathbb{R}^d} \boldsymbol{u}^k(x) \, \varrho^k(dx)$$

for all $s \in (s^k, s^{k+1})$; see Remarks 4.13/5.18 and (4.15). Similarly, we have

$$\begin{aligned} \int_{\mathbb{R}^d} \boldsymbol{v}_s(z) \, \varrho_s(dz) &= \int_{\mathbb{R}^d} U^{k+1}(x) \, \varrho^k(dx) = \int_{\mathbb{R}^d} W^{k+1}(x) \, \varrho^k(dx) \\ &= \int_{\mathbb{R}^d} \boldsymbol{u}^k(x) \, \varrho^k(dx) \end{aligned}$$

for $s = s^{k+1}$. We have used that the barycentric projection preserves the momentum and that $U^{k+1} = W^{k+1}$ in the cases with pressure. Applying this identity recursively and using assumption (6.1), we obtain the result.

6.3. **Regularity in Time.** In the following, we will use the subscript τ to indicate explicitly the dependence of various quantities on the timestep $\tau > 0$.

Lemma 6.1. *For suitable initial data $(\bar{\varrho}, \bar{\boldsymbol{v}}, \bar{\sigma})$ and $\tau > 0$, consider approximate solutions $(\varrho_\tau, \boldsymbol{v}_\tau, \boldsymbol{w}_\tau, \sigma_\tau)$ as defined in Section 6.2. For any $T > 0$ it holds*

$$\sup_{\tau > 0} \|\varrho_\tau\|_{\mathrm{Lip}([0,T];\mathscr{P}_2(\mathbb{R}^d))} \leqslant (2\bar{\mathcal{E}})^{1/2}.$$

The second moments remain finite: for all $s \in [0, T]$ we have

$$\sup_\tau \bigg(\int_{\mathbb{R}^d} |x|^2 \varrho_{\tau,s}(dx) \bigg)^{1/2} \leqslant \bigg(\int_{\mathbb{R}^d} |x|^2 \bar{\varrho}(dx) \bigg)^{1/2} + s(2\bar{\mathcal{E}})^{1/2}. \tag{6.9}$$

For any sequence $\tau_n \longrightarrow 0$, there exist a subsequence (not relabeled, for simplicity of notation) and a map $\varrho \in \mathrm{Lip}([0,T]; \mathscr{P}_2(\mathbb{R}^d))$ such that

$$\varrho_{\tau_n, s} \longrightarrow \varrho_s \quad \textit{narrowly as } n \to \infty, \textit{ for all } s \in [0, T]. \tag{6.10}$$

An analogous statement holds for the entropy density σ_τ. Moreover, for the limit map $\sigma \in \mathrm{Lip}([0,T]; \mathscr{M}_{\mathrm{Ent}}(\mathbb{R}^d))$ we can write $\sigma_s =: \varrho_s S_s$ with

$$S_s \in \mathscr{L}_+^\infty(\mathbb{R}^d, \varrho_s) \quad \textit{for all } s \in [0, T].$$

Proof. We divide the proof into three steps.

Step 1. Consider first $s_1 < s_2$ with $s_1, s_2 \in [s_\tau^k, s_\tau^{k+1})$ for some $k \in \mathbb{N}_0$. Since $\boldsymbol{t}_{\tau,s}(x) = x + (s - s_\tau^k)V_\tau^{k+1}(x)$ for ϱ_τ^k-a.e. $x \in \mathbb{R}^d$ and $s \in [s_\tau^k, s_\tau^{k+1})$, we get

$$\begin{aligned} \mathrm{W}_2(\varrho_{\tau,s_2}, \varrho_{\tau,s_1})^2 &\leqslant \int_{\mathbb{R}^d} |\boldsymbol{t}_{\tau,s_2}(x) - \boldsymbol{t}_{\tau,s_1}(x)|^2 \, \varrho_\tau^k(dx) \\ &= (s_2 - s_1)^2 \int_{\mathbb{R}^d} |V_\tau^{k+1}(x)|^2 \, \varrho_\tau^k(dx). \end{aligned} \tag{6.11}$$

For every $s \in (t_\tau^k, t_\tau^{k+1})$, we have

$$\int_{\mathbb{R}^d} |V_\tau^{k+1}(x)|^2 \, \varrho_\tau^k(dx) = \int_{\mathbb{R}^d} |\boldsymbol{v}_{\tau,s}(x)|^2 \, \varrho_{\tau,s}(dx)$$

(see (6.5)/(6.7)), which is bounded uniformly in τ, k because of (6.8) and (6.6). The estimate (6.11) remains true also for $s_2 = s_\tau^{k+1}$, by continuity.

Step 2. Consider now $0 \leqslant s_1 < s_2$ with the property that there exists at least one $k \in \mathbb{N}$ with $s_1 \leqslant s_\tau^k < s_2$. We use the triangle inequality to estimate

$$\mathrm{W}_2(\varrho_{\tau,s_2}, \varrho_{\tau,s_1}) \leqslant \mathrm{W}_2(\varrho_{\tau,s_2}, \varrho_\tau^{k_2}) + \sum_{k=k_1+1}^{k_2-1} \mathrm{W}_2(\varrho_\tau^{k+1}, \varrho_\tau^k) + \mathrm{W}_2(\varrho_\tau^{k_1+1}, \varrho_{\tau,s_1}),$$

where $k_i := \lfloor s_i/\tau \rfloor$ for $i = 1..2$. For each term, we can now apply the estimate from Step 1. Summing up all contributions, we obtain the inequality

$$\mathrm{W}_2(\varrho_{\tau,s_2}, \varrho_{\tau,s_1}) \leqslant |s_2 - s_1|(2\bar{\mathcal{E}})^{1/2} \quad \text{for all } 0 \leqslant s_1 < s_2.$$

To control the second moments, we write

$$\begin{aligned} \left(\int_{\mathbb{R}^d} |z|^2 \, \varrho_{\tau,s_2}(dz) \right)^{1/2} &= \left(\int_{\mathbb{R}^{2d}} |z|^2 \, \boldsymbol{\gamma}(dx, dz) \right)^{1/2} \\ &\leqslant \left(\int_{\mathbb{R}^{2d}} |x|^2 \, \boldsymbol{\gamma}(dx, dz) \right)^{1/2} + \left(\int_{\mathbb{R}^{2d}} |z - x|^2 \, \boldsymbol{\gamma}(dx, dz) \right)^{1/2} \\ &= \left(\int_{\mathbb{R}^d} |x|^2 \, \varrho_{\tau,s_1}(dx) \right)^{1/2} + \mathrm{W}_2(\varrho_{\tau,s_2}, \varrho_{\tau,s_1}), \end{aligned}$$

with $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})$ an optimal transport plan connecting ϱ_{τ,s_1} and ϱ_{τ,s_2}.

The uniform bound (6.9) implies that the family $\{\varrho_{\tau,s}\}_\tau$ is tight, thus precompact with respect to narrow convergence, for any $s \in [0, T]$. We can then apply Arzelà-Ascoli theorem to conclude; see Proposition 3.3.1 in [4], for example.

Step 3. The statement for σ_τ follows analogously. Note that the specific entropy $S_{\tau,s}$ is simply transported along the flow and hence bounded in $\mathscr{L}_+^\infty(\mathbb{R}^d, \varrho_{\tau,s})$. This implies, in particular, that σ_s must be absolutely continuous with respect to ϱ_s. □

Lemma 6.2. *For suitable initial data $(\bar{\varrho}, \bar{\boldsymbol{v}}, \bar{\sigma})$ and $\tau > 0$, consider approximate solutions $(\varrho_\tau, \boldsymbol{v}_\tau, \boldsymbol{w}_\tau, \sigma_\tau)$ as defined in Section 6.2. Let $T > 0$ be given. For any sequence $\tau_n \longrightarrow 0$, there exist a subsequence (not relabeled, for simplicity of notation) and a map $\boldsymbol{m} \in \mathrm{Lip}([0, T]; \mathscr{M}_{\mathrm{K}}(\mathbb{R}^d; \mathbb{R}^d))$ with the property that*

$$\|\boldsymbol{m}_{\tau_n,s} - \boldsymbol{m}_s\|_{\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d)} \longrightarrow 0 \quad \textit{as } n \to \infty, \textit{ for a.e. } s \in [0, T]. \tag{6.12}$$

Here $\boldsymbol{m}_{\tau,s} := \varrho_{\tau,s}\boldsymbol{w}_{\tau,s}$ for all τ, s. If the subsequence $\tau_n \longrightarrow 0$ is such that statement (6.10) *of Lemma 6.1 holds as well, then we have, for a.e. $s \in [0, T]$, that*

$$\boldsymbol{m}_s =: \varrho_s \boldsymbol{v}_s \quad \textit{with} \quad \boldsymbol{v}_s \in \mathscr{L}^2(\mathbb{R}^d, \varrho_s). \tag{6.13}$$

Proof. We divide the proof into three steps.

Step 1. Consider first $s_1 < s_2$ with $s_1, s_2 \in [s_\tau^k, s_\tau^{k+1})$ for some $k \in \mathbb{N}_0$. Since $\boldsymbol{t}_{\tau,s}(x) = x + (s - s_\tau^k) V_\tau^{k+1}(x)$ for ϱ_τ^k-a.e. $x \in \mathbb{R}^d$ and $s \in [s_\tau^k, s_\tau^{k+1})$, we get

$$\begin{aligned}
&\int_{\mathbb{R}^d} \zeta(z) \cdot \Big(\boldsymbol{m}_{\tau,s_2}(dz) - \boldsymbol{m}_{\tau,s_1}(dz) \Big) \\
&\qquad = \int_{\mathbb{R}^d} \Big(\zeta\big(\boldsymbol{t}_{\tau,s_2}(x)\big) - \zeta\big(\boldsymbol{t}_{\tau,s_1}\big) \Big) \cdot W_\tau^{k+1}(x)\, \varrho_\tau^k(dx) \\
&\qquad \leqslant |s_2 - s_1| \bigg(\int_{\mathbb{R}^d} |V_\tau^{k+1}(x)|^2\, \varrho_\tau^k(dx) \bigg)^{1/2} \bigg(\int_{\mathbb{R}^d} |W_\tau^{k+1}(x)|^2\, \varrho_\tau^k(dx) \bigg)^{1/2},
\end{aligned}$$

for any $\zeta \in \mathrm{BL}_1(\mathbb{R}^d; \mathbb{R}^d)$. We have used that the Lipschitz constant of ζ is bounded by 1; see Definition 1.4. Consider now $\varphi \in \mathscr{C}_c^1(\mathbb{R}^d)$ with $\eta(\mathbb{R}^d) \subset [0,1]$ and

$$\varphi(x) = 1 \text{ if } |x| \leqslant 1, \quad \varphi(x) = 0 \text{ if } |x| \geqslant 2.$$

For any $R, \varepsilon > 0$ we define the rescaled cut-off function/mollifier

$$\eta_R(x) := \varphi(x/R), \quad \varphi_\varepsilon(x) := \varepsilon^{-d} \varphi(x/\varepsilon)$$

for all $x \in \mathbb{R}^d$. Then we can decompose

$$\begin{aligned}
&\int_{\mathbb{R}^d} \zeta(x) \cdot \Big(W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x) \Big)\, \varrho_\tau^k(dx) \\
&\qquad = \int_{\mathbb{R}^d} \big(1 - \eta_R(x)\big) \zeta(x) \cdot \Big(W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x) \Big)\, \varrho_\tau^k(dx) \qquad (6.14)\\
&\qquad\quad + \int_{\mathbb{R}^d} \Big(\eta_R(x) \zeta(x) - \zeta_{R,\varepsilon}(x) \Big) \cdot \Big(W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x) \Big)\, \varrho_\tau^k(dx) \\
&\qquad\quad + \int_{\mathbb{R}^d} \zeta_{R,\varepsilon}(x) \cdot \Big(W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x) \Big)\, \varrho_\tau^k(dx)
\end{aligned}$$

with $\zeta_{R,\varepsilon} := (\eta_R \zeta) \star \varphi_\varepsilon$. The first term on the right-hand side of (6.14) satisfies

$$\begin{aligned}
&\bigg| \int_{\mathbb{R}^d} \big(1 - \eta_R(x)\big) \zeta(x) \cdot \Big(W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x) \Big)\, \varrho_\tau^k(dx) \bigg| \qquad (6.15)\\
&\qquad \leqslant \frac{C}{R} \bigg(\int_{\mathbb{R}^d} |x|^2\, \varrho_\tau^k(dx) \bigg)^{1/2} \\
&\qquad\qquad \times \left\{ \bigg(\int_{\mathbb{R}^d} |W_\tau^{k+1}|^2\, \varrho_\tau^k(dx) \bigg)^{1/2} + \bigg(\int_{\mathbb{R}^d} |\boldsymbol{u}_\tau^k|^2\, \varrho_\tau^k(dx) \bigg)^{1/2} \right\}
\end{aligned}$$

with constant C depending on the sup-norm of ζ. Recall that the second moment of ϱ_τ^k is bounded uniformly in τ, k, as shown in Lemma 6.1. Moreover, the terms in curly brackets are uniformly bounded because of (6.8) and (6.6). We conclude that (6.15) vanishes as $R \to \infty$, uniformly in τ, k. We observe that

$$\|\eta_R \zeta - \zeta_{R,\varepsilon}\|_{\mathscr{L}^\infty(\mathbb{R}^d)} \longrightarrow 0 \quad \text{as } \varepsilon \to 0,$$

by standard properties of mollication. Notice that $\eta_R \zeta$ has compact support in $\mathbb{R}^d$. Arguing as above, we find that the second term in (6.14) also converges to zero, uniformly in τ, k and R, as $\varepsilon \to 0$. Finally, we have the identity

$$\int_{\mathbb{R}^d} \zeta_{R,\varepsilon}(x) \cdot \Big(W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x) \Big) \varrho_\tau^k(dx) \tag{6.16}$$
$$= \tau \int_{\mathbb{R}^d} \nabla \zeta_{R,\varepsilon}(x) : \Big(\mathbf{P}_\tau^k(x)\, dx + \mathbf{R}_\tau^{k+1}(dx) \Big),$$

with $\mathbf{R}_\tau^{k+1}$ is the residual tensor corresponding to $\boldsymbol{t}_\tau^{k+1}$ and pressure term

$$\mathbf{P}_\tau^k(dx) := P\big(r_\tau^k(x), S_\tau^k(x)\big)\ \det\big(\nabla \boldsymbol{t}_\tau^{k+1,\mathrm{sym}}(x)\big)^{1-\gamma} \big(\nabla \boldsymbol{t}_\tau^{k+1,\mathrm{sym}}(x)\big)^{-1} dx.$$

We use (5.38) (with $x = \boldsymbol{t}_\tau^{k+1}(x) - \tau V_\tau^{k+1}(x)$) and (5.34) to get

$$\int_{\mathbb{R}^d} \Big(\mathrm{tr}\big(\mathbf{P}_\tau^k(x)\big)\, dx + \mathrm{tr}\big(\mathbf{R}_\tau^{k+1}(dx)\big) \Big) \leqslant d(\gamma-1)\, \mathcal{U}[\boldsymbol{t}_\tau^{k+1} | \varrho_\tau^k, \sigma_\tau^k]$$
$$+ \bigg(\int_{\mathbb{R}^d} |W_\tau^{k+1}(x) - \boldsymbol{u}_\tau^k(x)|^2\, \varrho_n^k(dx) \bigg)^{1/2} \bigg(\int_{\mathbb{R}^d} |V_\tau^{k+1}(x)|^2\, \varrho_\tau^k(dx) \bigg)^{1/2},$$

which can be bounded in terms of $\bar{\mathcal{E}}$, thus uniformly in τ, k, because of the energy balance (6.3). The sup-norm of $\nabla \zeta_{R,\varepsilon}$ in (6.16) can be bounded uniformly in $R \geqslant 1$ and $\varepsilon > 0$, by choice of η_R and because $\zeta \in \mathrm{BL}_1(\mathbb{R}^d; \mathbb{R}^d)$.

Collecting all terms and letting first $R \to \infty$, then $\varepsilon \to 0$, we conclude that

$$\|\boldsymbol{m}_{\tau,s_2} - \boldsymbol{m}_{\tau,s_1}\|_{\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d)} \leqslant C\big(|s_2 - s_1| + \tau\big) \quad \text{for all } s_1, s_2 \in [0,T], \tag{6.17}$$

for some constant C that can bounded in terms of $\bar{\mathcal{E}}$, hence uniformly in τ. The additional τ on the right-hand side of (6.17) occurs since the jumps in $\boldsymbol{m}_\tau$ at discrete times t_τ^k are always of order τ, not fractions of timesteps.

Step 2. Using (6.17), we conclude that for any choice of times $t_0 \leqslant t_1 \leqslant \ldots \leqslant t_m$ contained in $[0,T]$, we can bound the variation uniformly in τ as

$$\sum_{i=1}^m \|\boldsymbol{m}_{\tau,t_{i-1}} - \boldsymbol{m}_{\tau,t_i}\|_{\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d)} \leqslant C\big((t_m - t_0) + \tau\big).$$

Therefore the map $t \mapsto \boldsymbol{m}_{\tau,t} \in \mathscr{M}_{\mathrm{K}}(\mathbb{R}^d; \mathbb{R}^d)$ is of uniform bounded variation.

By Cauchy-Schwarz inequality, for each $s \geqslant 0$ we can estimate

$$\int_{\mathbb{R}^d} |\boldsymbol{m}_{\tau,s}(dx)| \leqslant \bigg(\int_{\mathbb{R}^d} |\boldsymbol{w}_{\tau,s}(x)|^2\, \varrho_{\tau,s}(dx) \bigg)^{1/2},$$

which is uniformly bounded because of (6.6). The Monge-Kantorovich norm can be controled by the total variation, which implies that $\{\boldsymbol{m}_{\tau,s}\}_\tau$ is precompact in the dual space $\mathrm{BL}(\mathbb{R}^d; \mathbb{R}^d)^*$, for all $s \geqslant 0$. On the other hand, a sequence of $\mathbb{R}^d$-valued measures with uniformly bounded total variation, converging in Monge-Kantorovich norm, has as limit again a measure; see Theorem 3.2 in [51]. We now apply Helly's theorem in the form of Theorem 2.3 in [43] to obtain (6.12).

Step 3. The limit map $s \mapsto \boldsymbol{m}_s \in \mathscr{M}_{\mathrm{K}}(\mathbb{R}^d; \mathbb{R}^d)$ satisfies the inequality

$$\|\boldsymbol{m}_{s_2} - \boldsymbol{m}_{s_1}\|_{\mathscr{M}_{\mathrm{K}}(\mathbb{R}^d)} \leqslant C|s_2 - s_1| \quad \text{for all } s_1, s_2 \in [0,T]$$

(recall (6.12) with $\tau \to 0$) and is therefore Lipschitz continuous, as claimed. Moreover, if ϱ and $\boldsymbol{m}$ are obtained from the same sequence $\tau_n \to 0$, then by lower semicontinuity of the kinetic energy functional with respect to narrow convergence of density and momentum, we conclude that $\boldsymbol{m}_s$ must be absolutely continuous with respect to ϱ_t, which proves the decomposition (6.13) for all $s \in [0,T]$. □

6.4. **Compactification.** We will need compactifications of the state space.

Lemma 6.3. *Let X be a completely regular space and $\mathscr{F} \subset \mathscr{C}(X, I)$, with $I := [0,1]$, a set of continuous functions that separates points and closed sets: for every closed set $E \subset X$ and every $x \in X \setminus E$, there exists $\Phi \in \mathscr{F}$ with $\Phi(u) \notin \overline{\Phi(E)}$. Then there exist a compact Hausdorff space $\mathfrak{X}$ and an embedding $e\colon X \longrightarrow \mathfrak{X}$ such that $e(X)$ is dense in $\mathfrak{X}$. Moreover, for any $\Phi \in \mathscr{F}$, the composition $\Phi \circ e^{-1}\colon e(X) \longrightarrow \mathbb{R}$ has a continuous extension to all of $\mathfrak{X}$. If $\mathscr{F}$ is countable, then $\mathfrak{X}$ is metrizable.*

Proof. Consider the product space $I^{\mathscr{F}}$, which is compact in the product topology, by Tykhonov's theorem. Let the map $e\colon X \longrightarrow I^{\mathscr{F}}$ be defined by

$$\pi_\Phi\big(e(u)\big) := \Phi(u) \quad \text{for all } u \in X \text{ and } \Phi \in \mathscr{F},$$

where $\pi_\Phi\colon I^{\mathscr{F}} \longrightarrow I$ denotes the projection onto the Φ-component. Since $\mathscr{F}$ separates points and closed sets, the map e is in fact an embedding (a homeomorphism between X and its image, with $e(X)$ given the relative topology of $I^{\mathscr{F}}$). We refer the reader to Proposition 4.53 of [44] for a proof. We now define $\mathfrak{X}$ to be the closure of $e(X)$ in $I^{\mathscr{F}}$. Being a closed subset of a compact Hausdorff space, the set $\mathfrak{X}$ is itself compact and Hausdorff. The set $e(X)$ is dense in $\mathfrak{X}$, by construction. We denote by $\mathscr{A}$ the smallest closed subalgebra in $\mathscr{C}_\mathrm{b}(X)$ containing $\mathscr{F}$. For any $\Phi \in \mathscr{A}$, there exists a continuous extension of $\Phi \circ e^{-1}$ to all of $\mathfrak{X}$; see Proposition 4.56 in [44]. If $\mathscr{F}$ is countable, then the set $I^{\mathscr{F}}$ is metrizable. Therefore, since every subset of a metrizable space is metrizable, we obtain that $\mathfrak{X}$ is metrizable. We refer the reader to Section 4.8 of [44] for additional information on compactifications. □

For simplicity of notation, we will identify X with its image $e(X)$. Then every function $\Phi \in \mathscr{A}$ can be extended as a continuous function on $\mathfrak{X}$. Notice that such an extension is uniquely determined because $e(X)$ is dense in $\mathfrak{X}$. We denote by $\mathscr{C}(\mathfrak{X})$ the space of all extensions obtained this way, and we will use the same symbols to indicate functions in $\mathscr{A}$ and their extensions in $\mathscr{C}(\mathfrak{X})$.

6.5. **Young Measures.** We will use Young measures to capture the behavior of weakly convergent sequences of approximate solutions of the compressible Euler equations (1.1). Recall that we assumed the specific entropy S to be non-negative and bounded at initial time. Since S is simply transported along with the flow, the same is true for all times, thus $S \in [0, S_\mathrm{max}]$ for some $S_\mathrm{max} \geqslant 0$. The state space for density, velocity, and specific entropy $(\varrho, \boldsymbol{v}, S)$ is therefore given by

$$X := [0, \infty) \times \mathbb{R}^d \times [0, S_\mathrm{max}]. \tag{6.18}$$

Equipped with the usual topology, it is a completely regular space.

Our goal is to define a suitable compactification of the state space. Equivalently, we must specify the set of continuous and bounded functions on X, for which we need to be able to describe weak limits of compositions with approximate solutions. Let us first consider a function that represents the total energy and mass. In slight abuse of notation, we use the same symbols $(\varrho, \boldsymbol{v}, S)$ for elements in X. Let

$$h(\varrho, \boldsymbol{v}, S) := \varrho + \Big(\tfrac{1}{2}\varrho|\boldsymbol{v}|^2 + U(\varrho, S)\Big); \tag{6.19}$$

see Definition 1.1. We now introduce the set

$$\mathscr{W}(X) := \left\{ \begin{array}{r} \varphi + \left(c_\varrho \cdot \begin{pmatrix} \varrho \\ \varrho \boldsymbol{v} \end{pmatrix} + c_\sigma \cdot \begin{pmatrix} \varrho S \\ \varrho \boldsymbol{v} S \end{pmatrix} + c_K : \varrho \boldsymbol{v} \otimes \boldsymbol{v} + c_U U(\varrho, S) \right) / h \colon \\ \varphi \in \mathscr{C}_0(X),\ c_\varrho, c_\sigma \in \mathbb{R}^{d+1},\ c_K \in \mathrm{Sym}_d(\mathbb{R}),\ c_U \in \mathbb{R} \end{array} \right\}.$$

One can check that the functions in $\mathscr{W}(X)$ are continuous and bounded. For this, it is convenient to introduce a parameterization of the state space X, similarly to the construction in the the proof of Lemma 5.25. One possible choice is

$$\varrho(\alpha) := \tan(\alpha), \quad \boldsymbol{v}(u) := \frac{u}{\sqrt{1 - |u|^2}}$$

for $\alpha \in [0, \pi/2)$ and $u \in B := \{u \in \mathbb{R}^d \colon |u| < 1\}$. For any $\Phi \in \mathscr{W}(X)$, the map

$$(\alpha, u, S) \mapsto \Phi\big(\varrho(\alpha), \boldsymbol{v}(u), S\big)$$

can be extended to a bounded function on the compact set $[0, \pi/2] \times \bar{B} \times [0, S_{\max}]$. This extension may be discontinuous at parts of the boundary. One can also check that the set $\mathscr{W}(X)$ is a closed *separable* vector space with respect to the sup-norm. To this end, notice that $\mathscr{W}(X)$ is a finite-dimensional augmentation of the vector space $\mathscr{C}_0(X)$, which is known to be separable; see also Lemma 2 in [7].

There exists a countable set $\mathscr{F}$ that is dense in $\mathscr{W}(X) \cap \mathscr{C}(X, I)$, $I = [0, 1]$, and separates points and closed sets. Indeed consider any closed set $E \subset X$ and any point $u \in X \setminus E$. One can find a $\Psi \in \mathscr{C}_0(X, I)$ with $\Psi(u) = 1$ and $\Psi | E \equiv 0$, and since $\mathscr{F}$ is dense there exists $\Phi \in \mathscr{F}$ with $\|\Phi - \Psi\|_{\mathscr{C}(X)} < \varepsilon$ for some $0 < \varepsilon < 1/2$. Applying Lemma 6.3, we obtain a compactification $\mathfrak{X}$ (a compact, metrizable Hausdorff space) of (6.18). The closed subalgebra $\mathscr{A}$ in Lemma 6.3 contains the set $\mathscr{W}(X)$.

Recall that $\dot{\mathbb{R}}^d$ is the one-point compactification of $\mathbb{R}^d$; see Section 4.3. Then

$$\mathbb{E} := \mathscr{L}^1\big([0, \infty), \mathscr{C}(\dot{\mathbb{R}}^d \times \mathfrak{X})\big)$$

is the space of (equivalence classes of) measurable maps $\boldsymbol{\phi} \colon [0, \infty) \longrightarrow \mathscr{C}(\dot{\mathbb{R}}^d \times \mathfrak{X})$ (i.e., pointwise limits of sequences of simple functions) with finite norm:

$$\|\boldsymbol{\phi}\|_{\mathbb{E}} := \int_0^\infty \|\boldsymbol{\phi}(s, \cdot)\|_{\mathscr{C}(\dot{\mathbb{R}}^d \times \mathfrak{X})} \, ds < \infty.$$

Notice that $\mathfrak{X}$ is compact and metrizable, hence separable. One can then show that $\mathbb{E}$ is a separable Banach space. Its topological dual is given by

$$\mathbb{E}^* := \mathscr{L}^\infty_w\big([0, \infty), \mathscr{M}_+(\dot{\mathbb{R}}^d \times \mathfrak{X})\big),$$

the space of (equivalence classes of) $\nu \colon [0, \infty) \longrightarrow \mathscr{M}_+(\dot{\mathbb{R}}^d \times \mathfrak{X})$ with

$$s \mapsto \int_{\dot{\mathbb{R}}^d \times \mathfrak{X}} \phi(x, \mathfrak{x}) \, \nu_s(dx, d\mathfrak{x}) \text{ measurable for all } \phi \in \mathscr{C}(\dot{\mathbb{R}}^d \times \mathfrak{X}), \text{ and}$$

$$\|\nu\|_{\mathbb{E}^*} := \operatorname*{ess\,sup}_{s \in [0, \infty)} \|\nu_s\|_{\mathscr{M}(\dot{\mathbb{R}}^d \times \mathfrak{X})} < \infty$$

(we write $s \mapsto \nu_s$ and $\mathfrak{x} := (\varrho, \boldsymbol{v}, S) \in \mathfrak{X}$). The duality is induced by the pairing

$$\langle \nu, \boldsymbol{\phi} \rangle := \int_0^\infty \int_{\dot{\mathbb{R}}^d \times \mathfrak{X}} \boldsymbol{\phi}(s, x, \mathfrak{x}) \, \nu_s(dx, d\mathfrak{x}) \, ds \tag{6.20}$$

for $\boldsymbol{\phi} \in \mathbb{E}$ and $\nu \in \mathbb{E}^*$. Bounded closed balls in $\mathbb{E}^*$ endowed with the weak* topology are metrizable and (sequentially) compact, by Banach-Alaoglu theorem.

For any timestep $\tau > 0$, we now define $\nu^1_\tau \in \mathbb{E}^*$ by

$$\int_{\dot{\mathbb{R}}^d\times\mathfrak{X}} \phi(x,\mathfrak{x})\,\nu^1_{\tau,s}(dx,d\mathfrak{x}) \tag{6.21}$$
$$:= \int_{\mathbb{R}^d} \phi\big(x, r_{\tau,s}(x), \boldsymbol{w}_{\tau,s}(x), S_{\tau,s}(x)\big)\, h\big(r_{\tau,s}(x), \boldsymbol{w}_{\tau,s}(x), S_{\tau,s}(x)\big)\,dx$$

for all $\phi \in \mathscr{C}(\dot{\mathbb{R}}^d \times \mathfrak{X})$ and $s \geqslant 0$; see (6.19). As usual, we have

$$\varrho_{\tau,s} =: r_{\tau,s}\mathcal{L}^d, \quad \sigma_{\tau,s} =: \varrho_{\tau,s}S_{\tau,s},$$

with approximate solutions $(\varrho_\tau, \boldsymbol{v}_\tau, \sigma_\tau)$ constructed in Section 6.2. Because of (6.6), the family $\{\nu^1_\tau\}_{\tau>0}$ is uniformly bounded in $\mathbb{E}^*$: We have

$$\|\nu^1_{\tau,s}\|_{\mathscr{M}(\dot{\mathbb{R}}^d\times\mathfrak{X})} = \int_{\dot{\mathbb{R}}^d\times\mathfrak{X}} \nu^1_{\tau,s}(dx,d\mathfrak{x}) = 1 + \mathcal{E}[\varrho_{\tau,s}, \boldsymbol{w}_{\tau,s}, \sigma_{\tau,s}]$$

for $\tau > 0$ and $s \geqslant 0$. Notice that $\nu^1_{\tau,s}$ is non-negative. From this, we get the relative (sequential) compactness of $\{\nu^1_\tau\}_{\tau>0}$ with respect to the weak* topology.

In summary, we have the following result:

Proposition 6.4 (Young Measure). *Consider a sequence $\tau_n \longrightarrow 0$ for $n \to \infty$ and let $\nu^1_n := \nu^1_{\tau_n} \in \mathbb{E}^*$ be defined by (6.21). Then there exist $\nu^1 \in \mathbb{E}^*$ and a subsequence (still denoted by $\{\nu^1_n\}_n$ for simplicity) with the property that*

$$\lim_{n\to\infty} \int_0^\infty \int_{\mathbb{R}^d} \boldsymbol{\phi}\big(s, x, r_{n,s}(x), \boldsymbol{w}_{n,s}(x), S_{n,s}(x)\big)\, h\big(r_{n,s}(x), \boldsymbol{w}_{n,s}(x), S_{n,s}(x)\big)\,dx\,ds$$
$$= \int_0^\infty \int_{\dot{\mathbb{R}}^d\times\mathfrak{X}} \boldsymbol{\phi}(s,x,\mathfrak{x})\,\nu^1_s(dx,d\mathfrak{x})\,ds \quad \textit{for all } \boldsymbol{\phi} \in \mathbb{E}.$$

To simplify the notation, we have used the subscript n instead of τ_n. We write

$$[f(\varrho,\boldsymbol{v},S)]_s(dx) := \int_{\mathfrak{X}} f(\mathfrak{x})/h(\mathfrak{x})\,\nu^1_s(dx,d\mathfrak{x}) \quad \textit{for a.e. } s \geqslant 0 \tag{6.22}$$

and for any $f\colon X \longrightarrow \mathbb{R}$ such that $f/h \in \mathscr{A}$; see Lemma 6.3.

Remark 6.5. In the same way, we define a second Young measure ν^2, using piecewise constant instead of piecewise linear interpolation in time. More precisely, for any $\tau > 0$, using the same notation as above, we define $\nu^2_\tau \in \mathbb{E}^*$ by

$$\int_{\dot{\mathbb{R}}^d\times\mathfrak{X}} \phi(x,\mathfrak{x})\,\nu^2_{\tau,s}(dx,d\mathfrak{x})$$
$$:= \int_{\mathbb{R}^d} \phi\big(x, r_{\tau,s_\tau}(x), \boldsymbol{w}_{\tau,s_\tau}(x), S_{\tau,s_\tau}(x)\big)\, h\big(r_{\tau,s_\tau}(x), \boldsymbol{w}_{\tau,s_\tau}(x), S_{\tau,s_\tau}(x)\big)\,dx$$

for all $\phi \in \mathscr{C}(\dot{\mathbb{R}}^d \times \mathfrak{X})$ and $s \geqslant 0$, where $s_\tau := \lfloor s/\tau\rfloor\tau$ denotes the largest integer multiple of τ less than or equal to s. Passing to the limit along a suitable sequence $\tau_n \longrightarrow 0$, we obtain the Young measure $\nu^2 \in \mathbb{E}^*$, which again can be used to capture concentrations/oscillations in weakly convergent sequences of approximate solutions of (1.1); see Proposition 6.4. Similar to (6.21), we use double brackets $[\![\cdot]\!]$ to indicate the pairing of ν^2 with suitable functions of $(\varrho, \boldsymbol{v}, S)$.

6.6. **Global Existence.** In this section, we establish the global existence of measure-valued solutions to (1.1), using the results of Sections 6.2 and 6.5.

Proof of Theorem 1.8. We consider a sequence of timesteps $\tau_n \longrightarrow 0$ as $n \to \infty$ with the property that the pointwise in time convergence in Lemmas 6.1/6.2 holds and that the approximate Young measures in Proposition 6.4 and Remark 6.5 converge to ν^1 and ν^2 along $\{\tau_n\}_n$. We will use the notation introduced in Sections 6.2 and 6.5, but with subscript n in place of τ_n, for simplicity.

By construction, it holds that

$$\left.\begin{aligned} \partial_s \varrho_n + \nabla\cdot(\varrho_n \boldsymbol{v}_n) = 0 \\ \partial_s \sigma_n + \nabla\cdot(\sigma_n \boldsymbol{v}_n) = 0 \end{aligned}\right\} \quad \text{in } \Big(\mathscr{C}_c^1([0,T)) \otimes \mathfrak{A}\Big)^*.$$

Recall that density $\varrho_{n,s}$ and entropy $\sigma_{n,s}$ have finite second moments for all $s \geqslant 0$. Passing to the limit $n \to \infty$, we get the first and third equations in (1.13).

It remains to prove the momentum equation. We observe that

$$\begin{aligned} -\int_{\mathbb{R}^d} \eta(0)\zeta(x)\cdot\bar{\boldsymbol{v}}(x)\,\bar{\varrho}(dx) &= \int_0^T \frac{d}{ds}\bigg(\int_{\mathbb{R}^d} \eta(s)\zeta(x)\cdot\boldsymbol{w}_{n,s}(x)\,\varrho_{n,s}(dx)\bigg)\,ds \\ &= \sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta'(s)\zeta(z)\cdot\boldsymbol{w}_{n,s}(z)\,\varrho_{n,s}(dz)\,ds \qquad (6.23) \\ &\quad + \sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta(s)\nabla\zeta(z) : \Big(\boldsymbol{w}_{n,s}(z)\otimes\boldsymbol{v}_{n,s}(z)\Big)\,\varrho_{n,s}(dz)\,ds \\ &\quad + \sum_{k\in\mathbb{N}_0} \tau_n \int_{\mathbb{R}^d} \eta(s_n^k)\nabla\zeta(x) : \Big(\mathbf{P}_n^k(x)\,dx + \mathbf{R}_n^{k+1}(dx)\Big) \end{aligned}$$

for any $\eta \in \mathscr{C}_c^1([0,T))$ and $\zeta \in \mathfrak{A}$. We proceed in four steps.

Step 1. In the first term of the right-hand side of (6.23), we directly apply (6.21) and pass to the limit. Using the definition of $\boldsymbol{v}$ in Lemma 6.2, we get that

$$\begin{aligned} &\lim_{n\to\infty} \sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta'(s)\zeta(z)\cdot\boldsymbol{w}_{n,s}(z)\,\varrho_{n,s}(dz)\,ds \\ &\qquad = \int_0^\infty \int_{\mathbb{R}^d} \eta'(s)\zeta(z)\cdot\boldsymbol{v}_s(z)\,\varrho_s(dz)\,ds. \end{aligned}$$

Recall that the momentum $\boldsymbol{m}_{n,s} := \varrho_{n,s}\boldsymbol{v}_{n,s}$ has finite first moment for all $s \geqslant 0$.

Step 2. In the second term on the right-hand side of (6.23), we need to replace the transport velocity $\boldsymbol{v}_{n,s}$ by the transported velocity $\boldsymbol{w}_{n,s}$ because the approximative Young measure (6.21) only captures the latter. We first rewrite

$$\begin{aligned} &\int_{\mathbb{R}^d} \nabla_z\zeta(z) : \Big(\boldsymbol{w}_{n,s}(z)\otimes\big(\boldsymbol{v}_{n,s}(z)-\boldsymbol{w}_{n,s}(z)\big)\Big)\,\varrho_{n,s}(dz) \\ &\qquad = \int_{\mathbb{R}^d} \nabla_z\zeta\big(\boldsymbol{t}_{n,s}(x)\big) : \Big(W^{k+1}(x)\otimes\big(V^{k+1}(x)-W^{k+1}(x)\big)\Big)\,\varrho_n^k(dx) \end{aligned}$$

for all $s \in (s_n^k, s_n^{k+1})$ and $k \in \mathbb{N}_0$. Then we apply (5.44) to estimate

$$\begin{aligned} &\bigg| \sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta(s)\nabla_z\zeta(z) : \Big(\boldsymbol{w}_{n,s}(z) \otimes \big(\boldsymbol{v}_{n,s}(z) - \boldsymbol{w}_{n,s}(z)\big)\Big)\, \varrho_{n,s}(dz)\, ds \bigg| \\ &\qquad \leqslant C \bigg\{ \varepsilon T \max_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} |W_n^{k+1}(x)|^2\, \varrho_n^k(dx) \\ &\qquad\qquad + C_\varepsilon \tau_n \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} |W_n^{k+1}(x) - \boldsymbol{u}^k(x)|^2\, \varrho_n^k(dx) \bigg\} \end{aligned} \tag{6.24}$$

for any $\varepsilon > 0$ and suitable constant C_ε. Here C depends on the sup-norm of $\eta\nabla_z\zeta$, which is finite; recall (1.12). Both the max and sum on the right-hand side of (6.24) are bounded by $\bar{\mathcal{E}}$ uniformly in n, because of energy equality (6.3) and (4.15). Since ε was arbitrary, we find that the left-hand side of (6.24) vanishes as $n \to \infty$.

We can now write

$$\sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta(s)\nabla\zeta(z) : \Big(\boldsymbol{w}_{n,s}(z) \otimes \boldsymbol{w}_{n,s}(z)\Big)\, \varrho_{n,s}(dz)\, ds = \langle \nu_n^1, \boldsymbol{\phi} \rangle$$

(recall definition (6.20) of the dual pairing), with test function

$$\boldsymbol{\phi}(s, x, \mathfrak{x}) := \eta(s)\nabla\zeta(x) : \varrho(\boldsymbol{w} \otimes \boldsymbol{w})/h(\varrho, \boldsymbol{w}, S)$$

for all $s \geqslant 0$, $x \in \dot{\mathbb{R}}^d$, and $\mathfrak{x} = (\varrho, \boldsymbol{w}, S) \in \mathfrak{X}$. From Proposition 6.4, we get

$$\begin{aligned} &\lim_{n\to\infty} \sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta(s)\nabla\zeta(z) : \Big(\boldsymbol{w}_{n,s}(z) \otimes \boldsymbol{w}_{n,s}(z)\Big)\, \varrho_{n,s}(dz)\, ds \\ &\qquad = \int_0^\infty \int_{\dot{\mathbb{R}}^d} \eta(s)\nabla\zeta(x) : [\varrho\boldsymbol{v} \otimes \boldsymbol{v}]_s(dx)\, ds. \end{aligned}$$

We refer the reader to (6.22) for notation.

Step 3. In the pressure term in (6.23), we want to replace

$$\det\big(\nabla \boldsymbol{t}_n^{k+1,\mathrm{sym}}(x)\big)^{1-\gamma} \big(\nabla \boldsymbol{t}_n^{k+1,\mathrm{sym}}(x)\big)^{-1}$$

by $\mathbb{1}$. Using (5.43), we obtain the estimate

$$\begin{aligned} &\bigg| \tau_n \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} \eta(s_n^k)\nabla\zeta(x) : \Big(\mathbf{P}_n^k(x) - P\big(r_n^k(x), S_n^k(x)\big)\mathbb{1}\Big)\, dx \bigg| \\ &\qquad \leqslant C \bigg\{ \varepsilon T \max_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} U\big(r_n^k(x), S_n^k(x)\big)\, dx \\ &\qquad\qquad + C_\varepsilon \tau_n \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} P\big(r_n^k(x), S_n^k(x)\big) D_{\mathcal{U}}\big(\nabla \boldsymbol{t}_n^{k+1,\mathrm{sym}}(x) - \mathbb{1}\big)\, dx \bigg\} \end{aligned} \tag{6.25}$$

for any $\varepsilon > 0$ and suitable constant C_ε. Here C depends on the sup-norm of $\eta\nabla\zeta$, which is bounded. Both max and sum on the right-hand side of (6.25) are bounded by $\bar{\mathcal{E}}$ uniformly in n, because of the energy equality (6.3). Since ε was arbitrary, we conclude that the left-hand side of (6.25) vanishes as $n \to \infty$.

Similarly, we can estimate

$$\begin{aligned}\Bigg| \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} \left(\tau_n \eta(s_n^k) - \int_{s_n^k}^{s_n^{k+1}} \eta(s)\,ds \right) \nabla\zeta(x) : \Big(P\big(r_n^k(x), S_n^k(x)\big)\mathbb{1}\Big)\,dx \Bigg| \\ \leqslant CT\omega(\tau_n,\eta)\ \max_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} U\big(r_n^k(x), S_n^k(x)\big)\,dx, \qquad (6.26)\end{aligned}$$

with C depending on the sup-norm of $\nabla\zeta$ and modulus of continuity

$$\omega(\tau_n,\eta) := \sup_{\substack{s_1,s_2\in[0,T]\\ |s_2-s_1|\leqslant\tau_n}} |\eta(s_2)-\eta(s_1)|.$$

The max on the right-hand side of (6.26) is bounded by $\bar{\mathcal{E}}$ uniformly in n, because of the energy balance (6.3). The left-hand side therefore vanishes as $n\to\infty$.

We can now write

$$\sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta(s)\nabla\zeta(z) : \Big(P\big(r_n^k(z), S_n^k(z)\big)\mathbb{1}\Big)\,dz\,ds = \langle \nu_n^2, \boldsymbol{\phi}\rangle$$

(recall definition (6.20) of the dual pairing), with test function

$$\boldsymbol{\phi}(s,x,\mathfrak{x}) := \eta(s)\nabla\zeta(x) : \Big(P(\varrho,S)\mathbb{1}\Big)/h(\varrho,\boldsymbol{w},S)$$

for all $s\geqslant 0$, $x\in\dot{\mathbb{R}}^d$, and $\mathfrak{x}=(\varrho,\boldsymbol{w},S)\in\mathfrak{X}$. From Remark 6.5, we get

$$\begin{aligned}\lim_{n\to\infty} \sum_{k\in\mathbb{N}_0} \int_{s_n^k}^{s_n^{k+1}} \int_{\mathbb{R}^d} \eta(s)\nabla\zeta(z) : \Big(P\big(r_n^k(z), S_n^k(z)\big)\mathbb{1}\Big)\,dz\,ds \\ = \int_0^\infty \int_{\dot{\mathbb{R}}^d} \eta(s)\nabla\zeta(z) : [\![P(\varrho,S)\mathbb{1}]\!]_s(dz)\,ds.\end{aligned}$$

We refer the reader to (6.22) for notation.

Step 4. The residual term in (6.23) can be estimated as

$$\left| \tau_n \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} \eta(s_n^k)\nabla\zeta(x) : \mathbf{R}_n^{k+1}(dx) \right| \leqslant C\tau_n \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}^d} \operatorname{tr}\big(\mathbf{R}_n^{k+1}(dx)\big),$$

with the sum on the right-hand side bounded by $\bar{\mathcal{E}}$ uniformly in n, because of the energy balance (6.3). Here C is some constant depending on the sup-norm of $\eta\nabla\zeta$. The left-hand side therefore vanishes as $n\to\infty$; see also Remark 5.24.

Combining Steps 1–4, we have proved the momentum equation. □

References

[1] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in $\boldsymbol{R}^n$*, Math. Z. **230** (1999), no. 2, 259–316.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.

[3] L. Ambrosio and N. Gigli, *Construction of the parallel transport in the Wasserstein space*, Methods Appl. Anal. **15** (2008), no. 1, 1–29.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, Birkhäuser Verlag, Basel, 2008.

[5] L. Ambrosio and D. Trevisan, *Well-posedness of Lagrangian flows and continuity equations in metric measure spaces*, Anal. PDE **7** (2014), no. 5, 1179–1234.

[6] H. H. Bauschke and X. Wang, *The kernel average for two convex functions and its application to the extension and representation of monotone operators*, Trans. Amer. Math. Soc. **361** (2009), no. 11.

[7] D. Blount and M. A. Kouritzin, *On convergence determining and separating classes of functions*, Stochastic Process. Appl. **120** (2010), no. 10, 1898–1907.

[8] G. Bouchitté, G. Buttazzo, and L. De Pascale, *The Monge-Kantorovich problem for distributions and applications*, J. Convex Anal. **17** (2010), no. 3-4, 925–943.

[9] F. Bouchut and F. James, *Équations de transport unidimensionnelles à coefficients discontinus*, C. R. Acad. Sci. Paris Sér. I Math. **320** (1995), no. 9, 1097–1102.

[10] ———, *Duality solutions for pressureless gases, monotone scalar conservation laws, and uniqueness*, Comm. Partial Differential Equations **24** (1999), no. 11-12, 2173–2189.

[11] Y. Brenier, *L^2 formulation of multidimensional scalar conservation laws*, Arch. Ration. Mech. Anal. **193** (2009), no. 1, 1–19.

[12] Y. Brenier, W. Gangbo, G. Savaré, and M. Westdickenberg, *Sticky particle dynamics with interactions*, J. Math. Pures Appl. (9) **99** (2013), no. 5, 577–617.

[13] Y. Brenier and E. Grenier, *Sticky particles and scalar conservation laws*, SIAM J. Numer. Anal. **35** (1998), no. 6, 2317–2328 (electronic).

[14] A. Bressan and T. Nguyen, *Non-existence and non-uniqueness for multidimensional sticky particle systems*, Kinet. Relat. Models **7** (2014), no. 2, 205–218.

[15] J. Brezina and E. Eduard, *Maximal dissipation principle for the complete Euler system* (2017), available at `arXiv:1712.04761[math.AP]`.

[16] T. Buckmaster, *Onsager's conjecture almost everywhere in time*, Comm. Math. Phys. **3** (2015), 1175–1198.

[17] T. Buckmaster, C. De Lellis, and Jr. Székelyhidi L., *Dissipative Euler flows with Onsager-critical spatial regularity*, Comm. Pure Appl. Math. **69** (2016), no. 9, 1613–1670.

[18] C. Castaing, P. Raynaud de Fitte, and M. Valadier, *Young measures on topological spaces*, Mathematics and its Applications, vol. 571, Kluwer Academic Publishers, Dordrecht, 2004. With applications in control theory and probability theory.

[19] F. Cavalletti, M. Sedjro, and M. Westdickenberg, *A Simple Proof of Global Existence for the 1D Pressureless Gas Dynamics Equations*, SIAM Math. Anal. **44** (2015), no. 1, 66–79.

[20] F. Cavalletti and M. Westdickenberg, *The polar cone of the set of monotone maps*, Proc. Amer. Math. Soc. **143** (2015), 781–787.

[21] G.-Q. Chen, *Compactness methods and nonlinear hyperbolic conservation laws*, Some current topics on nonlinear conservation laws, 2000, pp. 33–75.

[22] G.-Q. Chen and P. G. LeFloch, *Compressible Euler equations with general pressure law*, Arch. Ration. Mech. Anal. **153** (2000), no. 3, 221–259.

[23] G.-Q. Chen and M. Perepelitsa, *Vanishing viscosity limit of the Navier-Stokes equations to the Euler equations for compressible fluid flow*, Comm. Pure Appl. Math. **63** (2010), no. 11, 1469–1504.

[24] E. Chiodaroli, *A Counterexample to Well-Posedness of Entropy Solutions to the Compressible Euler System*, Accepted for publication in J. Hyperbolic Differ. Equ. (2014).

[25] E. Chiodaroli and O. Kreml, *On the Energy Dissipation Rate of Solutions to the Compressible Isentropic Euler System*, Accepted for publication in Arch. Ration. Mech. Anal. (2014).

[26] E. Chiodaroli, C. De Lellis, and O. Kreml, *Global Ill-Posedness of the Isentropic System of Gas Dynamics*, Accepted for publication in Comm. Pure App. Math. (2014).

[27] I. Chitescu, R. Miculescu, L. Nita, and L. Ioana, *Monge-Kantorovich norms on spaces of vector measures*, Results Math. **70** (2016), no. 3-4, 349–371.

[28] C. De Lellis and Jr. Székelyhidi L., *The Euler equations as a differential inclusion*, Ann. of Math. (2) **170** (2009), no. 3, 1417–1436.

[29] ———, *On admissibility criteria for weak solutions of the Euler equations*, Arch. Ration. Mech. Anal. **195** (2010), no. 1, 225–260.

[30] C. M. Dafermos, *The entropy rate admissibility criterion for solutions of hyperbolic conservation laws*, J. Differential Equations **14** (1973), 202–212.

[31] S. Demoulini, D. M. A. Stuart, and A. E. Tzavaras, *A variational approximation scheme for three-dimensional elastodynamics with polyconvex energy*, Arch. Ration. Mech. Anal. **157** (2001), no. 4, 325–344.

[32] ———, *Weak-strong uniqueness of dissipative measure-valued solutions for polyconvex elastodynamics*, Arch. Ration. Mech. Anal. **205** (2012), no. 3, 927–961.

[33] E. De Giorgi, *New problems on minimizing movements*, Boundary value problems for partial differential equations and applications, 1993, pp. 81–98.

[34] A. Dermoune, *d-dimensional pressureless gas equations*, Teor. Veroyatn. Primen. **49** (2004), no. 3, 610–614.

[35] X. X. Ding, G.-Q. Chen, and P. Z. Luo, *Convergence of the Lax-Friedrichs scheme for the system of equations of isentropic gas dynamics. I*, Acta Math. Sci. (Chinese) **7** (1987), no. 4, 467–480.

[36] ______, *Convergence of the Lax-Friedrichs scheme for the system of equations of isentropic gas dynamics. II*, Acta Math. Sci. (Chinese) **8** (1988), no. 1, 61–94.

[37] R. J. DiPerna, *Convergence of the viscosity method for isentropic gas dynamics*, Comm. Math. Phys. **91** (1983), no. 1, 1–30.

[38] R. M. Dudley, *Convergence of Baire measures*, Studia Math. **27** (1966), 251–268.

[39] W. E, Yu. G. Rykov, and Ya. G. Sinai, *Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics*, Comm. Math. Phys. **177** (1996), no. 2, 349–380.

[40] E. Feireisl, *Maximal dissipation and well-posedness for the compressible Euler system*, J. Math. Fluid Mech. **16** (2014), no. 3, 447–461.

[41] E. Feireisl, P. Gwiazda, A. Świerczewska-Gwiazda, and E. Wiedemann, *Regularity and energy conservation for the compressible Euler equations*, Arch. Ration. Mech. Anal. **223** (2017), no. 3, 1375–1395.

[42] U. S. Fjordholm, R. Kaeppeli, S. Mishra, and E. Tadmor, *Construction of approximate entropy measure valued solutions for hyperbolic systems of conservation laws*, Preprint (2014).

[43] I. Fleischer and J. E. Porter, *Convergence of metric space-valued BV functions*, Real Anal. Exchange **27** (2001/02), no. 1, 315–319.

[44] G. B. Folland, *Real analysis*, Second, Pure and Applied Mathematics (New York), John Wiley & Sons Inc., New York, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[45] W. Gangbo, T. Nguyen, and A. Tudorascu, *Euler-Poisson systems as action-minimizing paths in the Wasserstein space*, Arch. Ration. Mech. Anal. **192** (2009), no. 3, 419–452.

[46] W. Gangbo and M. Westdickenberg, *Optimal transport for the system of isentropic Euler equations*, Comm. PDE **34** (2009), no. 9, 1041–1073.

[47] N. Ghoussoub, *A variational theory for monotone vector fields*, J. Fixed Point Theory Appl. **4** (2008), no. 1, 107–135.

[48] N. Gigli, *On the geometry of the space of probability measures endowed with the quadratic optimal transport distance*, Ph.D. Thesis, 2004.

[49] E. Grenier, *Existence globale pour le système des gaz sans pression*, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), no. 2, 171–174.

[50] N. J. Higham, *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra Appl. **103** (1988), 103–118.

[51] S. C. Hille, T. Szarek, D. T. H. Worm, and M. A. Ziemlańska, *On a Schur-like property for spaces of measures* (2017), available at `arXiv:1703.00677[math.FA]`.

[52] F. Huang and Z. Wang, *Well posedness for pressureless flow*, Comm. Math. Phys. **222** (2001), no. 1, 117–146.

[53] P. Isett, *Hölder continuous Euler flows in three dimensions with compact support in time*, Annals of Mathematics Studies, vol. 196, Princeton University Press, Princeton, NJ, 2017.

[54] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17.

[55] J. Kristensen and F. Rindler, *Characterization of generalized gradient Young measures generated by sequences in $W^{1,1}$ and BV*, Arch. Ration. Mech. Anal. **197** (2010), no. 2, 539–598.

[56] P. G. LeFloch and M. Westdickenberg, *Finite energy solutions to the isentropic Euler equations with geometric effects*, J. Math. Pures Appl. (9) **88** (2007), no. 5, 389–429.

[57] H. Lim, Y. Yu, J. Glimm, X. L. Li, and D. H. Sharp, *Chaos, transport and mesh convergence for fluid mixing*, Acta Math. Appl. Sin. Engl. Ser. **24** (2008), no. 3, 355–368.

[58] H. Lim, J. Iwerks, J. Glimm, and D. H. Sharp, *Nonideal Rayleigh-Taylor mixing*, Proc. Natl. Acad. Sci. USA **107** (2010), no. 29, 12786–12792.

[59] P.-L. Lions, B. Perthame, and P. E. Souganidis, *Existence and stability of entropy solutions for the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian coordinates*, Comm. Pure Appl. Math. **49** (1996), no. 6, 599–638.

[60] P.-L. Lions, B. Perthame, and E. Tadmor, *Kinetic formulation of the isentropic gas dynamics and p-systems*, Comm. Math. Phys. **163** (1994), no. 2, 415–431.

[61] M. Mandelkern, *Metrization of the one-point compactification*, Proc. Amer. Math. Soc. **107** (1989), no. 4, 1111–1115.

[62] O. Moutsinga, *Convex hulls, sticky particle dynamics and pressure-less gas system*, Ann. Math. Blaise Pascal **15** (2008), no. 1, 57–80.

[63] J. Nash, *C^1 isometric imbeddings*, Ann. of Math. **60** (1954), no. 2, 383–396.

[64] ______, *The imbedding problem for Riemannian manifolds*, Ann. of Math. **63** (1956), no. 2, 20–63.

[65] L. Natile and G. Savaré, *A Wasserstein approach to the one-dimensional sticky particle system*, SIAM J. Math. Anal. **41** (2009), no. 4, 1340–1365.

[66] T. Nguyen and A. Tudorascu, *Pressureless Euler/Euler-Poisson systems via adhesion dynamics and scalar conservation laws*, SIAM J. Math. Anal. **40** (2008), no. 2, 754–775.

[67] L. Onsager, *Statistical hydrodynamics*, Nuovo Cimento (9) **6** (1949), no. Supplemento, 2 (Convegno Internazionale di Meccanica Statistica), 279–287.

[68] F. Poupaud and M. Rascle, *Measure solutions to the linear multi-dimensional transport equation with non-smooth coefficients*, Comm. Partial Differential Equations **22** (1997), no. 1-2, 337–358.

[69] F. Rindler, *Lower semicontinuity for integral functionals in the space of functions of bounded deformation via rigidity and Young measures*, Arch. Ration. Mech. Anal. **202** (2011), no. 1, 63–113.

[70] R. T. Rockafellar, *Convex analysis*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[71] M. Sever, *An existence theorem in the large for zero-pressure gas dynamics*, Differential Integral Equations **14** (2001), no. 9, 1077–1092.

[72] E. Tadmor, *A minimum entropy principle in the gas dynamics equations*, Appl. Numer. Math. **2** (1986), no. 3-5, 211–219.

[73] M. Westdickenberg, *Projections onto the cone of optimal transport maps and compressible fluid flows*, J. Hyperbolic Differ. Equ. **7** (2010), 605–649.

[74] G. Wolansky, *Dynamics of a system of sticking particles of finite size on the line*, Nonlinearity **20** (2007), no. 9, 2175–2189.

[75] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory. I. Projections on convex sets*, Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Madison, WI, 1971), Academic Press, New York, 1971, pp. 237–341.

[76] Ya. B. Zel'dovich, *Gravitational instability: An approximate theory for large density perturbations*, Astro. Astrophys. **5** (1970), 84–89.

Fabio Cavalletti, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy
Email address: `cavallet@sissa.it`

Marc Sedjro, AIMS Tanzania, Plot No: 288, Makwahiya Street, Regent Estate, Dar es Salaam, Tanzania
Email address: `sedjro@aims.ac.tz`

Michael Westdickenberg, Lehrstuhl für Mathematik (Analysis), RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany
Email address: `mwest@instmath.rwth-aachen.de`