Nonlinear Fokker–Planck–Kolmogorov equations as gradient flows on the space of probability measures

Marco Rehmeier* † Michael Röckner‡§

June 16, 2023

Abstract

We propose a general method to identify nonlinear Fokker–Planck–Kolmogorov equations (FPK equations) as gradient flows on the space of probability measures on $\mathbb{R}^d$ with a natural differential geometry. Our notion of gradient flow does not depend on any underlying metric structure such as the Wasserstein distance, but is derived from purely differential geometric principles. We explicitly identify the associated energy functions E and show that these are Lyapunov functions for the FPK solutions. Moreover, we show restricted uniqueness results for such gradient flows, and we also prove that the gradient of E is a gradient field on $\mathbb{R}^d$, which can be approximated by smooth gradient fields. These results cover classical and generalized porous media equations, where the latter have a generalized diffusivity function and a nonlinear transport-type first-order perturbation.

Keywords: Gradient flow, nonlinear Fokker–Planck equation, generalized porous media equation, differential geometry, Barenblatt solution.
2020 MSC: 35Q84, 35K55, 76S05, 58B20, 37B35, 35B40.

1 Introduction

In this paper we propose a simple general approach to identify solutions to nonlinear Fokker–Planck–Kolmogorov equations (FPK equations) of type

$$\partial_t \mu_t = \partial_{ij}(a_{ij}(t,\mu_t,x)\mu_t) - \partial_i(b_i(t,\mu_t,x)\mu_t) \tag{FPK}$$

(where $a_{ij}, b_i : \mathbb{R}_+ \times \mathcal{P} \times \mathbb{R}^d \to \mathbb{R}$), which are second-order parabolic equations for measures, as solutions to gradient flows on the space $\mathcal{P}$ of Borel probability measures on $\mathbb{R}^d$, $d \geq 1$, i.e. as solutions to equations of type

$$\frac{d}{dt}\mu_t = -\nabla^{\mathcal{P}} E_{\mu_t}. \tag{GF}$$

Furthermore, we explicitly identify the associated energy functions E as Lyapunov functions for these solutions, which can, e.g., be used to prove existence of stationary solutions to the FPK equation (see [9]). Our approach does not involve any metric on $\mathcal{P}$, but only a general construction of a natural tangent bundle on $\mathcal{P}$ and corresponding gradient $\nabla^{\mathcal{P}}$ (as was done in [2, 4, 21] for configuration spaces replacing $\mathcal{P}$). Therefore, our method is completely different from the well-known approach developed in the theory of optimal transport, e.g. via the theory of gradient flows in metric spaces applied to the Wasserstein space $\mathcal{P}_2$ of Borel probability measures on $\mathbb{R}^d$ with finite second moment, in which

*Faculty of Mathematics, Bielefeld University, 33615 Bielefeld, Germany. E-mail: mrehmeier@math.uni-bielefeld.de
†Scuola Normale Superiore Pisa, Italy. E-mail: marco.rehmeier@sns.it
‡Faculty of Mathematics, Bielefeld University, 33615 Bielefeld, Germany. E-mail: roeckner@math.uni-bielefeld.de
§Academy of Mathematics and System Sciences, CAS, Beijing

case the induced metric, the Wasserstein metric, plays a major role. This approach goes back to the pioneering work by Otto [15] and Jordan, Kinderlehrer, Otto [13]. For very nice presentations of this and related material, see e.g. the books of Ambrosio, Gigli, Savaré [6] and Villani [24], as well as the lecture notes by Figalli and Glaudo [12].

Although our method is expected to be applicable to more general classes of equations, in this work we focus on (FPK) with so-called Nemytskii-type coefficients, which depend on the density u (with respect to Lebesgue measure dx) of $\mu = u(x)dx \in \mathcal{P}$ pointwise via $a_{ij}(\mu, x) = \tilde{a}_{ij}(u(x), x)$ and $b_i(\mu, x) = \tilde{b}_i(u(x), x)$ for measurable $\tilde{a}_{ij}, \tilde{b}_i : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. More precisely, our main results are on generalized porous media equations (PME) of type

$$\partial_t u = \Delta\beta(u) - \operatorname{div}(Db(u)u), \tag{gPME}$$

which are nonlinear FPK equations reformulated as PDEs for the densities $u(t)$ of solutions $t \mapsto \mu_t$. Here, *generalized* PME does not only refer to the generalized diffusivity $\beta : \mathbb{R} \to \mathbb{R}$ ($\beta(r) = |r|^{m-1}r$ gives the classical PME), but also to the additional nonlinear transport-type first order perturbation, composed of a vector field $D : \mathbb{R}^d \to \mathbb{R}^d$ and the nonlinearity $b : \mathbb{R} \to \mathbb{R}$. Though solutions to these equations have densities with respect to dx for $t > 0$, under suitable assumptions on β their initial data may be arbitrary probability measures. To the best of our knowledge, the present paper is the first in which equations of this general type are identified as gradient flows on a space of measures.

The differential geometry on $\mathcal{P}$ used in this paper was first introduced in [19]. But the method how to identity natural tangent bundles over "manifold-like" state spaces $\mathcal{M}$ has been known much longer and goes back to [2, 4, 21]. It goes in two steps. The first is to fix a large enough (i.e., at least point separating) class $\mathcal{F}$ of "test functions" $F : \mathcal{M} \to \mathbb{R}$. The second is to fix for each $x \in \mathcal{M}$ a set $\mathcal{C}$ of "suitable" curves $\gamma^x : (-\varepsilon, \varepsilon) \to \mathcal{M}$ such that $\gamma^x(0) = x$, along which one can differentiate $t \mapsto F(\gamma^x(t))$ at $t = 0$ for each $F \in \mathcal{F}$, which gives derivations at x as linear maps on $\mathcal{F}$, i.e. a "reduced tangent space" at x. Here "reduced" indicates the fact that if $\mathcal{M}$ is a Riemannian manifold, then depending on $\mathcal{F} \subseteq C^1(\mathcal{M})$ and $\mathcal{C}$, one obtains smaller tangent spaces than the usual $T_x\mathcal{M}$. To find $\mathcal{F}$ and, in particular, $\mathcal{C}$, if $\mathcal{M}$ is a space of measures on a Riemannian manifold M is possible due to the general idea of "lifting" the geometry from M to $\mathcal{M}$, which goes as follows: For a smooth, compactly supported section φ in the tangent bundle TM of M (we write $\varphi \in C_c^\infty(M, TM)$), let Φ^φ be the flow of φ on M, i.e. $\Phi^\varphi(0, x) = x$ and $\frac{d}{dt}\Phi^\varphi(t, x) = \varphi(\Phi^\varphi(t, x))$ for all $(t, x) \in \mathbb{R} \times M$. For $\nu \in \mathcal{M}$, let $\tilde{\mu}_t^{\varphi,\nu} := \nu \circ \Phi^\varphi(t)^{-1}$. The curves $t \mapsto \tilde{\mu}_t^{\varphi,\nu}$, $\varphi \in C_c^\infty(M, TM)$ form such a set $\mathcal{C}$ of suitable curves. Then let $\mathcal{F}$ be all functions of the form $F : \mathcal{M} \to \mathbb{R}$, $F(\nu) = f(\nu(h_1), \ldots, \nu(h_k))$, $h_i \in C_c^\infty(M, \mathbb{R})$, $f \in C_b^1(\mathbb{R}^k)$, $k \in \mathbb{N}$. One sets $\nabla_\varphi F(\nu) := \frac{d}{dt}F(\tilde{\mu}_t^{\varphi,\nu})_{|t=0}$ and obtains by the chain rule

$$\nabla_\varphi F(\nu) = \int_M \left\langle \sum_{i=1}^k \partial_i f(\nu(h_1), \ldots, \nu(h_k)) \nabla^M h_i, \varphi \right\rangle_{T_x M} d\nu(x).$$

Hence one defines $T_\nu\mathcal{M} := \overline{C_c^\infty(M, TM)}^{L^2(M,TM;\nu)} = L^2(M, TM); \nu)$ with inner product $\langle \varphi, \bar{\varphi} \rangle_{T_\nu\mathcal{M}} := \int_M \langle \varphi(x), \bar{\varphi}(x) \rangle_{T_x M}\, d\nu(x)$. Then the gradient $\nabla^{\mathcal{M}} F$ is uniquely characterized by

$$\nabla_\varphi F(\nu) = \langle \nabla^{\mathcal{M}} F(\nu), \varphi \rangle_{T_\nu\mathcal{M}} \tag{1.1}$$

for all $\nu \in \mathcal{M}$ and $\varphi \in T_\nu\mathcal{M}$, i.e. for F as above $\nabla^{\mathcal{M}} F(\nu) = \sum_{i=1}^k \partial_i f(\nu(h_1), \ldots, \nu(h_k)) \nabla^M h_i$. The first implementation of this scheme was done in [2, 4, 21] to deduce a natural geometry on the configuration space $\Gamma(M)$ of a Riemannian manifold M, i.e. on the space of $\mathbb{Z}_+ \cup \{\infty\}$-valued Radon measures on M. For the case $M = \mathbb{R}^d$, $\mathcal{M} = \mathcal{P}$, which we consider in the present paper following [19], we may replace $\tilde{\mu}_t^{\varphi,\nu}$ by $\mu_t^{\varphi,\nu} := \nu \circ (\mathrm{Id} + t\varphi)^{-1}$, since clearly $\frac{d}{dt}\tilde{\mu}_t^{\varphi,\nu}{}_{|t=0} = \frac{d}{dt}\mu_t^{\varphi,\nu}{}_{|t=0}$ for all $\varphi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and $\nu \in \mathcal{P}$. The definition of $\mu^{\varphi,\nu}$ can be extended to any $\varphi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$, whereby we arrive at our tangent spaces $T_\nu\mathcal{P} = L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$ for all $\nu \in \mathcal{P}$. Depending on the FPK equation

(FPK) under consideration, we consider $T_\nu\mathcal{P}$ with either the standard L^2-inner product $\langle\cdot,\cdot\rangle_\nu$ or, unlike [19], with a weighted inner product with weight at $\nu = v(x)dx$ given by $\frac{1}{b(v)}$, see (3.8)-(3.9). Please see Section 2.2 for details on our geometry, and the appendix for a further independent deduction of it. That the rather small class $\{t\mapsto\mu_t^{\varphi,\nu},\varphi\in L^2(\mathbb{R}^d,\mathbb{R}^d;\nu),\nu\in\mathcal{P}\}$ of curves is sufficiently large and indeed the "right" class to give a gradient flow representation for all nonlinear FPK equations of type (gPME) is rather surprising, but this is the core result of the present paper.

Our main aim is to reveal the gradient flow structure of nonlinear FPK equations with respect to the aforementioned differential geometry and to identify the corresponding energy functions. The key idea is to evaluate solutions to nonlinear FPK equations, which are weakly continuous paths $t\mapsto\mu_t$ in $\mathcal{P}$, not only through the linear functions $t\mapsto\mu_t(\zeta) := \int_{\mathbb{R}^d}\zeta\,d\mu_t$, $\zeta\in C_c^2(\mathbb{R}^d)$, but through the much larger class of nonlinear finitely based functions on $\mathcal{P}$, i.e. through all $F:\mathcal{P}\to\mathbb{R}$ of type $F(\mu) = f(\mu(h_1),\dots,\mu(h_k)), k\in\mathbb{N}, h_i\in C_c^2(\mathbb{R}^d), f\in C_b^1(\mathbb{R}^k)$ (see the Notation section at the end of this introduction for the notation of the usual function spaces used here and throughout). More precisely, one simply calculates $\frac{d}{dt}F(\mu_t)$ using the fact that $t\mapsto\mu_t$ solves the nonlinear FPK equation. In order for $t\mapsto\mu_t$ to satisfy (GF), this derivative should equal $-\mathrm{diff}E_{\mu_t}(\nabla^{\mathcal{P}}F_{\mu_t})$ for all F as above, which provides an ansatz to find E and hence the right gradient flow equation. Please see Section 2.3 for details. This method, which is rigorously implemented for (gPME) in this paper, seems also applicable for other classes of FPK equations. Since the corresponding energy is a Lyapunov function for the solutions, which can be used to analyze the asymptotic behavior of the gradient flow solutions (see [9]), we hope that one benefit of this paper will be to identify Lyapunov functions for more general classes of nonlinear FPK equations and, thereby, to obtain new results for the asymptotics of their solutions. The following first main result of this paper concerns the classical PME with arbitrary (not necessarily absolutely continuous with respect to Lebesgue measure) probability measures as initial data.

Theorem 1. (see Theorem 3.2 for the precise formulation) Let $\mu_0\in\mathcal{P}$ and $m\geq 2$. The unique probability solution to the classical PME

$$\partial_t u = \Delta(|u|^{m-1}u),\quad t>0 \tag{PME}$$

in $\bigcap_{\delta>0}L^\infty((\delta,\infty)\times\mathbb{R}^d)$ is the restricted unique solution to (GF) on $\mathcal{P}$ with energy $E(vdx) = \frac{1}{m-1}\int_{\mathbb{R}^d}v(x)^m dx$ and gradient $\nabla^{\mathcal{P}}E_\nu = \frac{\nabla(v^m)}{v}$ ($=\frac{m}{m-1}\nabla(v^{m-1})$, if $v^{m-1}\in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$, which is the case for the unique solution to (PME) for $m\geq 3$).

In particular, the famous Barenblatt solutions turn out to satisfy a gradient flow equation on $\mathcal{P}$. Our energy E coincides with the energy function in [15], where a very nice physical interpretation of E is presented, though only absolutely continuous initial measures have been considered there. As the second main result, we extend this theorem to the much larger class of nonlinear FPK equations of type (gPME) as follows.

Theorem 2. (see Theorem 3.7 for the precise formulation) Under suitable assumptions on $\beta, D=-\nabla\Phi$ and b, the unique probability solution to (gPME) is the restricted unique solution to (GF) in $L^\infty([0,\infty)\times\mathbb{R}^d)$ with energy

$$E(v(x)dx) := \int_{\mathbb{R}^d}\eta(v(x))\,dx + \int_{\mathbb{R}^d}\Phi(x)v(x)dx,\quad \eta(r) := \int_0^r g(s)ds := \int_0^r\int_1^s\frac{\beta'(w)}{wb(w)}dw\,ds,$$

and gradient $\nabla_b^{\mathcal{P}}E_\nu = b(v)\nabla\big(g(v)+\Phi\big)$ with respect to a weighted metric tensor $\langle\cdot,\cdot\rangle_b$ depending on b. Moreover, E is a Lyapunov function for u, i.e. $E(u(t))\leq E(u(s))$ for all $s\leq t$.

As mentioned before, due to the nonlinear transport-type perturbation, we are led to consider the weighted L^2-metric tensor $\langle\cdot,\cdot\rangle_b$ with weight at $\nu = v(x)dx$ given by $\frac{1}{b(v)}$. Furthermore, we prove for (sufficiently many) $\nu \in \mathcal{P}$ that $\nabla_b^{\mathcal{P}} E_\nu$ is a gradient vector field on $\mathbb{R}^d$ which, in addition, belongs to the closure of $\{b(v)\nabla\zeta \,|\, \zeta \in C_c^\infty(\mathbb{R}^d)\}$ in $L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$ (see Proposition 3.10 below). We note that in the classical case of the heat equation, i.e. $\beta(r) = r$ and $\Phi = 0$, E is the classical Boltzmann entropy function.

We would like to repeat that our approach is substantially different from [15, 13] and subsequent related works, such as [6, 24, 12, 11]. A more detailed overview of the available literature on these directions is beyond the scope of this introduction. The point we want to stress is that all these works are indeed of a different flavor than the present paper, in which we approach the identification of FPK equations as gradient flows with respect to the differential geometry obtained by the previously mentioned lifting procedure, but without involving any metric.

Organization. The rest of this paper is organized as follows. We repeat the general notion of gradient flows on Riemannian manifolds in Section 2.1 and introduce our geometry on $\mathcal{P}$ in 2.2. In 2.3 we explain our ansatz to reveal the gradient flow structure of nonlinear FPKEs, and we present the notion of gradient flow in our differential geometry on $\mathcal{P}$ in 2.4. Section 3 contains the main results. First, we present our result on the classical PME and equations with general diffusivity functions in 3.1, then those on generalized PMEs in 3.2 including nonlinear transport-type drifts, and, finally, in Section 3.3 we briefly discuss an extension to more general divergence-type equations. In the appendix we present a further natural deduction of our differential geometry.

Notation. We write $\mathcal{P} := \mathcal{P}(\mathbb{R}^d)$ for the set of Borel probability measures on $\mathbb{R}^d$ and $\mathcal{P}_a \subseteq \mathcal{P}$ for its subset of absolutely continuous measures with respect to Lebesgue measure dx. For a measure ν on a measurable space $(X, \mathcal{X})$ and an $\mathcal{X}$-measurable function $f : X \to \mathbb{R}$, we abbreviate $\nu(f) := \int_X f(x)\, d\nu(x)$, provided the integral is defined. A *Borel curve* in $\mathcal{P}$ is a curve $t \mapsto \mu_t$ from an interval $I \subseteq \mathbb{R}$ such that $t \mapsto \mu_t(A)$ is measurable for all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$. Recall that the *topology of weak convergence of measures* is the initial topology of the maps $\mu \mapsto \mu(g)$ for all bounded continuous $g : \mathbb{R}^d \to \mathbb{R}$. Restricted to $\mathcal{P}$, it suffices to consider smooth compactly supported g.

We use the following standard function space notation. $C^m(\mathbb{R}^d, \mathbb{R}^k)$, $C_b^m(\mathbb{R}^d, \mathbb{R}^k)$ and $C_c^m(\mathbb{R}^d, \mathbb{R}^k)$, $m \in \mathbb{N}_0 \cup \{\infty\}$, denote the spaces of m-fold differentiable functions $g : \mathbb{R}^d \to \mathbb{R}^k$ and, respectively, their subsets of bounded and compactly supported functions. For $k = 1$, we write $C^m(\mathbb{R}^d), C_b^m(\mathbb{R}^d)$ and $C_c^m(\mathbb{R}^d)$, and for $m = 0$ $C(\mathbb{R}^d, \mathbb{R}^k)$, $C_b(\mathbb{R}^d, \mathbb{R}^k)$ and $C_c(\mathbb{R}^d, \mathbb{R}^k)$, respectively. For $p \in [1, \infty]$, an open set $U \subseteq \mathbb{R}^d$ and a measure ν on $\mathcal{B}(\mathbb{R}^d)$, the usual L^p-spaces of (equivalence classes of) Borel measurable functions $g : U \to \mathbb{R}^k$ with (locally) ν-integrable p-th power are denoted by $L^p_{(\mathrm{loc})}(U, \mathbb{R}^k; \nu)$. If either $\nu = dx$ or $k = 1$, we simply write $L^p_{(\mathrm{loc})}(U, \mathbb{R}^k)$ or $L^p_{(\mathrm{loc})}(U; \nu)$, respectively. For the usual associated norms, we write $|\cdot|_p$, if no confusion about d, k or ν can occur. Moreover, we denote by $W^{m,p}_{(\mathrm{loc})}(\mathbb{R}^d)$ the usual Sobolev spaces of functions $g : \mathbb{R}^d \to \mathbb{R}$ with weak partial derivatives up to order m in $L^p_{(\mathrm{loc})}(\mathbb{R}^d)$. For $p = 2$, we write $H^m_{(\mathrm{loc})}(\mathbb{R}^d)$. These Sobolev spaces are always considered with respect to dx. $\mathcal{B}_b(\mathbb{R}^d)$ is the space of Borel measurable, bounded functions $g : \mathbb{R}^d \to \mathbb{R}$.

The usual Euclidean norm and inner product on $\mathbb{R}^d$ are denoted by $|\cdot|$ and $x \cdot y$. Depending on context, we write Id both for the identity vector field $\Phi : \mathbb{R}^d \to \mathbb{R}^d$, $\Phi(x) = x$, and for the $d \times d$-identity matrix. For $f : \mathbb{R}^d \to \mathbb{R}$, we write $f^+ := \max(0, f)$ and $f^- := \max(0, -f)$. We set $\mathbb{R}_+ := [0, \infty)$ and, for $x, y \in \mathbb{R}$, $x \wedge y := \min(x, y)$ and $x \vee y := \max(x, y)$.

2 Gradient flows on Riemannian manifolds and on $\mathcal{P}$

Here we briefly recall the notion of gradient flows on Riemannian manifolds from a purely differential geometric perspective. Then we introduce our geometry on $\mathcal{P}$ and the notion of gradient flow on $\mathcal{P}$ based on it.

2.1 Gradient flows on Riemannian manifolds

Let M be a d-dimensional Riemannian manifold, $C^1(M)$ the space of differentiable functions $F: M \to \mathbb{R}$, T_xM the (Hilbert) tangent space at $x \in M$ (i.e. $\ell \in T_xM$ if and only if $\ell : C^1(M) \to \mathbb{R}$ such that $\ell(FG) = \ell(F)G(x) + \ell(G)F(x)$ for all $F, G \in C^1(M)$) with inner product $\langle \cdot, \cdot \rangle_{T_xM}$ and dual space T_xM^*, and let $\mathrm{diff}F_x$ be the differential of F at x. Elements $\ell \in T_xM$ are called *derivations* and act on $F \in C^1(M)$ via

$$\ell(F) = \frac{d}{d\tau}(F \circ \gamma^\ell(\tau))_{\tau=0} = \mathrm{diff}F_x\left(\frac{d}{d\tau}\gamma^\ell(\tau)_{|\tau=0}\right) = \left\langle \nabla F_x, \frac{d}{d\tau}\gamma^\ell(\tau)_{|\tau=0}\right\rangle_{T_xM} = \left\langle \nabla F_x, \ell\right\rangle_{T_xM}.$$

Here γ^ℓ denotes any C^1-curve $\gamma^\ell : (-\varepsilon, \varepsilon) \to M$ such that the first equality holds for all $F \in C^1(M)$. At least one such curve with $\gamma^\ell(0) = x$ exists, since T_xM can equivalently be defined as the set of equivalence classes of C^1-curves $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = x$, with equivalence relation $\gamma_1 \sim \gamma_2 :\iff \frac{d}{d\tau}(F\circ\gamma_1(\tau))_{|\tau=0} = \frac{d}{d\tau}(F\circ\gamma_2(\tau))_{|\tau=0}$ for all $F \in C^1(M)$, i.e. there is an isomorphism between the set of derivations ℓ and such equivalence classes. The second equality is the definition of the differential $\mathrm{diff}F$. Since $\mathrm{diff}F_x \in T_xM^*$, the third equality follows from Riesz' representation theorem and uniquely characterizes the gradient $\nabla F_x \in T_xM$. The final equality follows from the aforementioned equivalence of both definitions of T_xM. In particular, letting $\ell = -\nabla E_x$ for $E \in C^1(M)$, it follows that the derivation $-\nabla E_x$ acts on $F \in C^1(M)$ via

$$-\nabla E_x(F) = -\mathrm{diff}E_x(\nabla F_x) = -\langle \nabla F_x, \nabla E_x \rangle_{T_xM}.$$

Let $I = (a, b)$ with $-\infty \leq a < b \leq +\infty$. A *gradient flow* on M is an equation

$$\frac{d}{d\tau}x(\tau)_{|\tau=t} = -\nabla E_{x(t)}, \quad \forall t \in I \tag{2.1}$$

in the tangent bundle $TM = \bigsqcup_{x\in M} T_xM$, to be solved for differentiable curves $t \mapsto x(t)$ on M. More precisely, the differentiability of $t \mapsto x(t)$ yields $\frac{d}{d\tau}x(\tau)_{|\tau=t} \in T_{x(t)}M$ and, as explained above, $-\nabla E_{x(t)}$ acts via

$$-\nabla E_{x(t)}(F) = -\langle \nabla E_{x(t)}, \nabla F_{x(t)} \rangle_{T_{x(t)}M}, \quad \forall F \in C^1(M).$$

Hence (2.1) implies

$$\frac{d}{d\tau}F\big(x(\tau)\big)_{|\tau=t} = -\mathrm{diff}E_x(\nabla F_x) = -\langle \nabla E_{x(t)}, \nabla F_{x(t)} \rangle_{T_{x(t)}M}, \quad \forall t \in I, F \in C^1(M), \tag{2.2}$$

which is, in fact, equivalent to (2.1). Indeed, for $\xi, \xi' \in T_xM$, $\mathrm{diff}F_x(\xi) = \mathrm{diff}F_x(\xi')$ for all $F \in C^1(M)$ implies $\xi = \xi'$, since every cotangent vector at x is the differential at x of a function in $C^1(M)$, and the set of cotangent elements at x separates tangent vectors at x. E in (2.1) is also called *energy (function)* of the system, and choosing $F = E$ in (2.2) shows

$$E(x(t)) - E(x(s)) = -\int_s^t |\nabla E_{x(r)}|^2_{T_{x(r)}M}\, dr, \quad \forall s \leq t \in I,$$

i.e. for any solution x of (2.1), $t \mapsto E(x(t))$ is non-increasing.

2.2 Differential geometry on $\mathcal{P}$

As mentioned in the introduction, we consider $\mathcal{P}$ as a manifold-like space with the lifted geometry from $\mathbb{R}^d$. Originally, an analogous geometry was introduced in [2, 4, 21] on the space $\Gamma(M)$ of $\mathbb{Z}_+ \cup \{+\infty\}$-valued Radon measures on a Riemannian manifold M (see also the follow-up papers [17, 3, 5, 16, 20, 14, 1]). The same approach for $M = \mathbb{R}^d$ and $\Gamma(M)$ replaced by $\mathcal{P} = \mathcal{P}(\mathbb{R}^d)$, which was first implemented in [19, App.A] and was recalled in the introduction of this paper, leads to the geometry of the present paper. Here we summarize the resulting geometry without repeating in detail the general idea of its construction. The test function class consists of the finitely-based functions

$$\mathcal{F}C_b^2 := \{F : \mathcal{P} \to \mathbb{R} : F(\nu) = f(\nu(h_1), \ldots, \nu(h_k)), k \in \mathbb{N}, h_i \in C_c^2(\mathbb{R}^d), f \in C_b^1(\mathbb{R}^k)\}, \tag{2.3}$$

and the class of differentiable curves $(-\varepsilon, \varepsilon) \ni t \mapsto \mu_t$ on $\mathcal{P}$ passing through $\nu \in \mathcal{P}$ at $t = 0$ is given by all curves of type

$$\mu_t^{\varphi,\nu} := \nu \circ (\mathrm{Id} + t\varphi)^{-1}, \quad \varphi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu).$$

Consequently, the tangent bundle $T\mathcal{P} := \bigsqcup_{\nu \in \mathcal{P}} T_\nu \mathcal{P}$ consists of the Hilbert tangent spaces

$$T_\nu \mathcal{P} := L^2(\mathbb{R}^d, \mathbb{R}^d; \nu), \tag{2.4}$$

with metric tensor

$$\langle \cdot, \cdot \rangle : \nu \mapsto \langle \cdot, \cdot \rangle_\nu \tag{2.5}$$

on $T\mathcal{P}$, where $\langle \cdot, \cdot \rangle_\nu$ denotes the standard inner product on $L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$. Indeed, $\varphi \in T_\nu \mathcal{P}$ acts as a derivation on $F \in \mathcal{F}C_b^2, F(\mu) = f(\mu(h_1), \ldots, \mu(h_k))$, because by the chain rule

$$\frac{d}{dt} F(\mu_t^{\varphi,\nu})_{|t=0} = \sum_{i=1}^k \partial_i f(\nu(h_1), \ldots, \nu(h_k)) \langle \nabla h_i, \varphi \rangle_\nu.$$

The *differential* of $F \in \mathcal{F}C_b^2$ at $\nu \in \mathcal{P}$ is the continuous linear functional on $T_\nu \mathcal{P}$

$$\mathrm{diff} F_\nu : \varphi \mapsto \frac{d}{dt} F(\mu_t^{\varphi,\nu})_{|t=0} = \sum_{i=1}^k \partial_i f(\nu(h_1), \ldots, \nu(h_k)) \langle \nabla h_i, \varphi \rangle_\nu, \tag{2.6}$$

and, in analogy to Riemannian geometry, the *gradient* $\nabla^{\mathcal{P}} F$ at ν is defined as the unique element $\nabla^{\mathcal{P}} F_\nu \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$ associated to $\mathrm{diff} F_\nu$ via the Riesz isomorphism, i.e.

$$\nabla^{\mathcal{P}} F_\nu := \sum_{i=1}^k \partial_i f(\nu(h_1), \ldots, \nu(h_k)) \nabla h_i. \tag{2.7}$$

Consequently, by definition we have

$$\mathrm{diff} F_\nu(\varphi) = \langle \nabla^{\mathcal{P}} F_\nu, \varphi \rangle_\nu, \quad \forall \varphi \in T_\nu \mathcal{P} = L^2(\mathbb{R}^d, \mathbb{R}^d; \nu). \tag{2.8}$$

In particular, $\nabla^{\mathcal{P}} F$ is independent of the representation $F(\mu) = f(\mu(h_1), \ldots, \mu(h_k))$ of F. For $G \in \mathcal{F}C_b^2$, analogously to Section 2.1, $\nabla^{\mathcal{P}} G_\nu$ as a derivation acts on $F \in \mathcal{F}C_b^2$ via

$$\nabla^{\mathcal{P}} G_\nu(F) = \mathrm{diff} G_\nu(\nabla^{\mathcal{P}} F_\nu) = \langle \nabla^{\mathcal{P}} G_\nu, \nabla^{\mathcal{P}} F_\nu \rangle_\nu = \sum_{i=1}^k \partial_i f(\nu(h_1), \ldots, \nu(h_k)) \langle \nabla h_i, \nabla^{\mathcal{P}} G_\nu \rangle_\nu. \tag{2.9}$$

The following slight generalizations have not been considered in [19] or, as far as we know, elsewhere, but are necessary for our main results.

Generalized differential and gradient. More generally, for $G : D(G) \subseteq \mathcal{P} \to \mathbb{R}$ not necessarily from $\mathcal{F}C_b^2$, we define $D(\mathrm{diff} G)$ to be the set of all $\nu \in D(G)$ such that

$$\mathrm{diff} G_\nu(\varphi) := \frac{d}{dt} G(\mu_t^{\varphi,\nu})_{|t=0}, \quad \varphi \in C_c^1(\mathbb{R}^d, \mathbb{R}^d), \tag{2.10}$$

is well-defined and linear as well as continuous with respect to the usual $L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$-topology. For such ν, $\mathrm{diff} G_\nu$ has a unique linear and continuous extension to all of $L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$. Then again we define $\nabla^{\mathcal{P}} G_\nu$ as the unique element in $T_\nu \mathcal{P}$ such that

$$\mathrm{diff} G_\nu(\varphi) = \left\langle \nabla^{\mathcal{P}} G_\nu, \varphi \right\rangle_\nu, \quad \forall \varphi \in C_c^1(\mathbb{R}^d, \mathbb{R}^d),$$

i.e. (2.9) remains valid. In particular, we allow $G(\nu)$, $\mathrm{diff} G_\nu$ and $\nabla^{\mathcal{P}} G_\nu$ to be defined for ν from strict subsets of $\mathcal{P}$ only.

Weighted metric tensors and gradient. We also introduce *weighted* metric tensors on the tangent bundle $T\mathcal{P}$, which are needed for our main results, in particular Theorem 3.6, as follows. Let $\alpha : \mathcal{P} \to \mathcal{B}_b(\mathbb{R}^d)$ be such that $C^{-1} \leq \alpha(\nu)(x) \leq C$ for all $x \in \mathbb{R}^d$ and $\nu \in \mathcal{P}$, with $C = C(\alpha) > 1$ independent of x and ν. The inner product

$$\langle \varphi, \tilde{\varphi} \rangle_{\alpha,\nu} := \langle \alpha(\nu)\varphi, \tilde{\varphi} \rangle_\nu, \quad \varphi, \tilde{\varphi} \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$$

is equivalent to the standard inner product $\langle \cdot, \cdot \rangle_\nu$, and we denote the weighted metric tensor $\nu \mapsto \langle \cdot, \cdot \rangle_{\alpha,\nu}$ by $\langle \cdot, \cdot \rangle_\alpha$. For $G : D(G) \subseteq \mathcal{P} \to \mathbb{R}$ and $\nu \in D(\text{diff}G)$, we denote the gradient of G at ν with respect to $\langle \cdot, \cdot \rangle_\alpha$ by $\nabla^{\mathcal{P}}_\alpha G_\nu$, i.e. it is the unique element in $L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$ such that

$$\text{diff}G_\nu(\varphi) = \langle \nabla^{\mathcal{P}}_\alpha G_\nu, \varphi \rangle_{\alpha,\nu}, \quad \forall \varphi \in C^1_c(\mathbb{R}^d, \mathbb{R}^d).$$

For $G \in \mathcal{F}C^2_b$, we have

$$\nabla^{\mathcal{P}}_\alpha G_\nu = \alpha(\nu)^{-1} \nabla^{\mathcal{P}} G_\nu. \tag{2.11}$$

2.3 Gradient flow character of nonlinear Fokker–Planck equations

Here, through a heuristic computation, we explain a general method to reveal the gradient flow character of nonlinear Fokker–Planck–Kolmogorov equations by considering solutions of the latter as curves on $\mathcal{P}$ with the geometry from Section 2.2. First, we repeat the definition of distributional solutions to (FPK).

Definition 2.1. A *distributional probability solution* $(0, \infty) \ni t \mapsto \mu_t$ to (FPK) is a Borel curve $t \mapsto \mu_t$ of probability measures $\mu_t \in \mathcal{P}$ such that $(t,x) \mapsto a_{ij}(t, \mu_t, x)$ and $(t,x) \mapsto b_i(t, \mu_t, x)$, $i, j \in \{1, \dots, d\}$, belong to $L^1_{\text{loc}}([0,\infty) \times \mathbb{R}^d; \mu_t dt)$, and for all $0 \leq s < t$ and $\zeta \in C^2_c(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \zeta(x)\, d\mu_t(x) - \int_{\mathbb{R}^d} \zeta(x)\, d\mu_s(x) = \int_s^t \int_{\mathbb{R}^d} a_{ij}(r, \mu_r, x)\partial_{ij}\zeta(x) + b_i(r, \mu_r, x)\partial_i \zeta(x)\, d\mu_r(x) dr.$$

For brevity, we simply say *solution* instead of *distributional probability solution.*

Such solutions are clearly weakly continuous. Moreover, it follows that $t \mapsto \mu_t(\zeta)$ is differentiable dt-a.s. for every $\zeta \in C^2_c(\mathbb{R}^d)$, with derivative

$$\frac{d}{dt} \int_{\mathbb{R}^d} \zeta(x)\, d\mu_t(x) = \int_{\mathbb{R}^d} a_{ij}(t, \mu_t, x)\partial_{ij}\zeta(x) + b_i(t, \mu_t, x)\partial_i\zeta(x)\, d\mu_t(x) \quad dt\text{-a.s.},$$

where the exceptional set can be chosen independent of ζ. For the following heuristic computation we assume $\mu_t = u(t,x)dx$ and that the dependence of a_{ij} on μ is of Nemytskii-type. For $F \in \mathcal{F}C^2_b, F(\nu) = f(\nu(h_1), \dots, \nu(h_k))$, $t \mapsto F(\mu_t)$ is differentiable dt-a.s. with derivative

$$\begin{aligned}
\frac{d}{d\tau} F(\mu_\tau)_{|\tau=t} &= \sum_{l=1}^k \partial_l f(\mu_t(h_1), \dots, \mu_t(h_k)) \frac{d}{d\tau}\mu_\tau(h_l)_{|\tau=t} \\
&= \sum_{l=1}^k \partial_l f(\mu_t(h_1), \dots, \mu_t(h_k)) \int_{\mathbb{R}^d} a_{ij}(t, \mu_t)\partial_{ij} h_l + b(t, \mu_t)\nabla h_l\, d\mu_t \\
&= \left\langle \sum_{l=1}^k \partial_l f(\mu_t(h_1), \dots, \mu_t(h_k))\nabla h_l, \frac{-\operatorname{div}(a(t, \mu_t)u(t))}{u(t)} + b(t, \mu_t) \right\rangle_{\mu_t} \\
&= \left\langle \nabla^{\mathcal{P}} F_{\mu_t}, \frac{-\operatorname{div}(a(t, \mu_t)u(t))}{u(t)} + b(t, \mu_t) \right\rangle_{\mu_t}
\end{aligned}$$

for dt-a.e. $t > 0$. Here we ignored the question of differentiability of $u(t)$ in x and wrote $a = (a_{ij})_{1\le i,j\le d}$ and $\operatorname{div}(au) \in \mathbb{R}^d$ for the vector with entries $\partial_i(ua_{ij})$, $1 \le j \le d$ (using Einstein summation convention). Hence $t \mapsto \frac{d}{d\tau}\mu_{\tau|\tau=t}$ is a curve of tangent vectors (up to an dt-zero set) with action

$$\frac{d}{d\tau}\mu_{\tau|\tau=t}(F) = \mathrm{diff}F_{\mu_t}(\frac{d}{d\tau}\mu_{\tau|\tau=t}) = \frac{d}{d\tau}F(\mu_\tau)_{|\tau=t} = \left\langle \nabla^{\mathcal{P}}F_{\mu_t}, \frac{-\operatorname{div}(a(t,\mu_t)u(t))}{u(t)} + b(t,\mu_t)\right\rangle_{\mu_t} \quad \forall F \in \mathcal{F}C_b^2.$$

Consequently, if there is $E : \mathcal{P} \to \mathbb{R}$ such that dt-a.s. $\mu_t \in D(\mathrm{diff}E)$ and

$$\mathrm{diff}E_{\mu_t}(\nabla^{\mathcal{P}}F_{\mu_t}) = \left\langle \nabla^{\mathcal{P}}F_{\mu_t}, \frac{\operatorname{div}(a(t,\mu_t)u(t))}{u(t)} - b(t,\mu_t)\right\rangle_{\mu_t}, \tag{2.12}$$

it follows that

$$\nabla^{\mathcal{P}}E_{\mu_t} = \frac{\operatorname{div}(a(t,\mu_t)u(t))}{u(t)} - b(t,\mu_t) \in L^2(\mathbb{R}^d,\mathbb{R}^d;\mu_t) \quad dt\text{-a.s.},$$

and hence that $t \mapsto \mu_t$ satisfies dt-a.s. (compare with (2.2))

$$\frac{d}{d\tau}\mu_{\tau|\tau=t} = -\nabla^{\mathcal{P}}E_{\mu_t}$$

in $T_{\mu_t}\mathcal{P}$. The dt-zero exceptional set cannot be avoided, since in general it is not possible to obtain differentiability of $t \mapsto \mu_t(\zeta)$ for *every* $t > 0$ from Definition 2.1.

Remark 2.2. *In the case* $a_{ij}(t,\nu,x) = \delta_{ij}\frac{\beta(v(x))}{v(x)}$, $\nu = v(x)dx$, *which is treated in Section 3, one has* $\frac{\operatorname{div}(a(t,\mu_t)u(t))}{u(t)} = \frac{\nabla\beta(u(t))}{u(t)}$.

Hence the essence of our ansatz is to first evaluate not only the linear maps $\nu \mapsto \nu(\zeta), \zeta \in C_c^2(\mathbb{R}^d)$, along solutions $t \mapsto \mu_t$ to (FPK), but the much bigger class of nonlinear functions $F \in \mathcal{F}C_b^2$, and then to identify the energy E via (2.12).

2.4 Gradient flows on $\mathcal{P}$

Let $\alpha : \mathcal{P} \to \mathcal{B}_b(\mathbb{R}^d)$, $C^{-1} < \alpha < C$, $C > 1$, be a weight as in Section 2.2 with metric tensor $\langle\cdot,\cdot\rangle_\alpha$ (the non-weighted case corresponds to $\alpha \equiv 1$).

Definition 2.3. Let $I = (a,b)$, $-\infty \le a < b \le +\infty$, and $t \mapsto \mu_t$ be such that there is a dt-zero set $N \subseteq I$ such that for all $t \in N^c$, $t \mapsto \mu_t(\zeta)$ is differentiable for all $\zeta \in C_c^2(\mathbb{R}^d)$. For $E : D(E) \subseteq \mathcal{P} \to \mathbb{R}$ such that $\mu_t \in D(\mathrm{diff}E)$ (and thus $\nabla_\alpha^{\mathcal{P}}E_{\mu_t}$ is defined for all $t \in N^c$), $t \mapsto \mu_t$ is called a solution to the gradient flow with energy E and weight α, if it satisfies

$$\frac{d}{d\tau}\mu_{\tau|\tau=t} = -\nabla_\alpha^{\mathcal{P}}E_{\mu_t} \quad dt\text{-a.s.} \tag{$\mathcal{P}$-GF}$$

($\mathcal{P}$-GF) is equivalent to $\mu_t \in D(\mathrm{diff}E)$ dt-a.s. and

$$\frac{d}{d\tau}F\big(\mu_\tau\big)_{|\tau=t} = -\mathrm{diff}E_{\mu_t}(\nabla_\alpha^{\mathcal{P}}F_{\mu_t}) \ \ \forall F \in \mathcal{F}C_b^2 \quad dt-\text{a.s.}, \tag{2.13}$$

and for $F(\nu) = f(\nu(h_1),\dots,\nu(h_k))$ and $t \in N^c$, by (2.11) the latter is equivalent to

$$\sum_{i=1}^k \partial_i f\big(\mu_t(h_1),\dots,\mu_t(h_k)\big)\frac{d}{d\tau}\mu_\tau(h_i)_{|\tau=t} = -\sum_{i=1}^k \partial_i f\big(\mu_t(h_1),\dots,\mu_t(h_k)\big)\mathrm{diff}E_{\mu_t}(\alpha(\mu_t)^{-1}\nabla h_i).$$

Therefore $t \mapsto \mu_t$ solves ($\mathcal{P}$-GF) if and only if there is a zero set $N \subseteq I$ such that for all $t \in N^c$ and $\zeta \in C_c^2(\mathbb{R}^d)$ (with exceptional set independent of ζ)

$$\frac{d}{d\tau}\mu_\tau(\zeta)_{|\tau=t} = -\mathrm{diff}E_{\mu_t}(\alpha(\mu_t)^{-1}\nabla\zeta). \tag{2.14}$$

Remark 2.4. *For $t \mapsto \mu_t$ and E as above, heuristically choosing $F = E$ in* (2.13) *yields*

$$E(\mu_t) - E(\mu_s) = -\int_s^t |\nabla_\alpha^{\mathcal{P}} E_{\mu_r}|^2_{\alpha,\mu_r}\, dr \le 0, \quad \forall s < t \in I.$$

However, since we do not necessarily have $E \in \mathcal{F}C_b^2$ ($\nabla E_\alpha^{\mathcal{P}}$ might be defined in the generalized sense explained in Section 2.2), the choice $F = E$ in (2.13) *might not be permitted.*

3 Generalized PME as gradient flows on $\mathcal{P}$ and identification of the energy

In this section we present our main results: We show that solutions to a class of generalized PMEs solve gradient flows equations on $\mathcal{P}$ and we identify the corresponding energy function. We also prove a uniqueness result for the gradient flows. First, we consider the classical PME, see Theorem 3.2. In Section 3.2 we consider generalized PMEs with general diffusivity functions and an additional nonlinear transport-type first-order term, see Theorem 3.6. Finally, in Section 3.3, we present a further generalization to a larger class of divergence-type equations. We stress that throughout this section the uniqueness of solutions to the FPK equation does not play any role for our results and proofs, except, of course, for the uniqueness assertions concerning the gradient flow in theorems 3.2 and 3.7.

Consider the equation (gPME). Conditions on the diffusivity function β as well as on the spatial drift-vector field D and the nonlinearity b are given below. In the literature, equations with $D = 0$ are also called generalized PME or PME with generalized diffusivity, see for instance [23]. In the present paper, *generalized* PME refers both to the general diffusivity β and to the additional transport-type drift. (gPME) is a special case of (FPK) with $a_{ij}(t,\mu,x) = \delta_{ij}\frac{\beta(u(x))}{u(x)}$ and $b_i(t,\mu,x) = D^{(i)}(x)b(u(x))$, where $\mu = u(x)dx$. For the case of Nemytskii-type coefficients, the following definition of solution is a more common (but equivalent) formulation than Definition 2.1.

Definition 3.1. $\mu : [0,\infty) \to \mathcal{P}, \mu : t \mapsto \mu_t$, is a *weakly continuous distributional probability solution to* (gPME) (for brevity just *solution*) with initial datum $\mu_0 \in \mathcal{P}$ if $t \mapsto \mu_t$ is weakly continuous, $\mu_t = u(t,x)dx$ dt-a.s.,

$$u, \beta(u) \in L^1_{\text{loc}}([0,\infty) \times \mathbb{R}^d),\ b(u)D \in L^1_{\text{loc}}([0,\infty) \times \mathbb{R}^d, \mathbb{R}^d),$$

and

$$\int_0^\infty \int_{\mathbb{R}^d} u(t,x)(\partial_t \zeta(t,x) + b(u(t,x))D(x) \cdot \nabla\zeta(t,x)) \\ + \beta(u(t,x))\Delta\zeta(t,x) dt\, dx + \mu_0(\zeta(0,\cdot)) = 0,\ \forall \zeta \in C_0^\infty([0,\infty) \times \mathbb{R}^d).$$

It follows that the initial datum μ_0 is attained weakly, i.e. $\mu_t \longrightarrow \mu_0$ as $t \to 0$ in the topology of weak convergence of probability measures.

3.1 The classical porous media equation

For $m > 1$, consider the classical porous media equation (PME) on $\mathbb{R}^d$, i.e. (gPME) with $\beta(r) = |r|^{m-1}r$ and $D = 0$, with an arbitrary initial condition $\mu_0 \in \mathcal{P}$. By [18, Thm.1], for every $\mu_0 \in \mathcal{P}$, there is a unique solution $(0,\infty) \ni t \mapsto u(t,x)dx \in \mathcal{P}$ to (PME) in $\bigcap_{\delta>0} L^\infty((\delta,\infty) \times \mathbb{R}^d)$ such that $u(t,x)dx \longrightarrow \mu_0$ weakly as $t \to 0$. Furthermore, by [7, Sect.5], $u(t)^\alpha \in H^1(\mathbb{R}^d)$ for all $\alpha \ge \frac{m+1}{2}$ for dt-a.a. $t > 0$.

Identification of the energy. According to Section 2.3 (in particular, Remark 2.2), the ansatz to find the energy E for the gradient flow equation is to find E such that $\mu_t \in D(\mathrm{diff}E)$ dt$-a.s.$ and

$$\mathrm{diff}E_\nu(\varphi) = \left\langle \varphi, \frac{\nabla(v^m)}{v} - b(t,\nu) \right\rangle_\nu \quad \forall \varphi \in C_c^1(\mathbb{R}^d, \mathbb{R}^d)$$

for all $\nu \in D(\mathrm{diff}E)$. We shall obtain

$$D_{\mathrm{nice}} := \left\{ \nu = v(x)dx \in \mathcal{P} \,\middle|\, v \in L^\infty(\mathbb{R}^d), v^m \in W^{1,1}_{\mathrm{loc}}, \frac{\nabla(v^m)}{v} \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu) \right\} \subseteq D(\mathrm{diff}E).$$

As shown in the proof of Theorem 3.2 below, the right choice of E is

$$E : D(E) \subseteq \mathcal{P}_a \to \mathbb{R}, \quad E(udx) := \int_{\mathbb{R}^d} \eta(u(x))dx, \quad \eta(r) := m \int_0^r \int_1^s w^{m-2}\, dw\, ds, r \geq 0, \tag{3.1}$$

i.e. we have, since $D(E) \subseteq \mathcal{P}$,

$$E(udx) = \frac{1}{m-1}\bigg(\int_{\mathbb{R}^d} u(x)^m dx - m \bigg), \tag{3.2}$$

and $D(E) = \mathcal{P}_a \cap L^m(\mathbb{R}^d)$. Since E appears only through its differential (a first-order functional), the (compared with Otto [15]), additional zero order summand $-\frac{m}{m-1}$ can be dropped.

Main result. Note that for $\varphi \in C_c^1(\mathbb{R}^d, \mathbb{R}^d)$, $\nu = v(x)dx \in \mathcal{P}_a$ and $|\tau| < \varepsilon = \varepsilon(\varphi) > 0$, by the transformation rule the measure $\mu_\tau^{\varphi,\nu} = \nu \circ (\mathrm{Id} + \tau\varphi)^{-1}$ is absolutely continuous with respect to dx with density

$$\frac{d\mu_\tau^{\varphi,\nu}}{dx}(x) = v\big((\mathrm{Id} + \tau\varphi)^{-1}(x)\big) |\det D(\mathrm{Id} + \tau\varphi)(x)|^{-1}, \tag{3.3}$$

where $D\psi$ denotes the Jacobian of $\psi \in C^1(\mathbb{R}^d, \mathbb{R}^d)$.

Theorem 3.2. *(i) Let $m \geq 2$ and $\mu_0 \in \mathcal{P}$. The unique solution $t \mapsto \mu_t = u(t,x)dx$ to* (PME) *in $\bigcap_{\delta > 0} L^\infty((\delta,\infty) \times \mathbb{R}^d)$ with initial datum μ_0 is a solution to* (GF) *in $\mathcal{P}$ with E as in* (3.1) *and the standard L^2-metric tensor $\langle \cdot, \cdot \rangle$ on $T\mathcal{P}$ (i.e. with weight $\alpha \equiv 1$). The gradient of E for $\nu = v(x)dx \in D_{nice}$ is*

$$\nabla^{\mathcal{P}} E_\nu = \frac{\nabla(v^m)}{v} \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu).$$

(ii) If, in addition, $m \geq 3$, for $t \mapsto \mu_t$ from (i) we have

$$\nabla^{\mathcal{P}} E_{\mu_t} = \frac{\nabla(u(t)^m)}{u(t)} = \frac{m}{m-1} \nabla(u(t)^{m-1}) \quad dt\text{-}a.s.,$$

i.e. in this case $\nabla^{\mathcal{P}} E_{\mu_t}$ is a gradient vector field in $T_{\mu_t}\mathcal{P} = L^2(\mathbb{R}^d, \mathbb{R}^d; \mu_t)$.

(iii) The curve $t \mapsto \mu_t$ from (i) is the unique solution to the gradient flow with E as in (i) in $\big(\bigcap_{\delta>0} L^\infty((\delta,\infty) \times \mathbb{R}^d)\big) \cap L^1_{loc}([0,\infty) \times \mathbb{R}^d) \cap D_{nice}$ such that $\mu_t \longrightarrow \mu_0$ weakly as $t \to 0$.

Proof. (i) For $t \mapsto \mu_t = u(t,x)dx$ from the assertion we have for all $\zeta \in C_c^2(\mathbb{R}^d)$ and dt-a.e. $t > 0$ (with exceptional set independent of ζ)

$$\frac{d}{d\tau}\mu_\tau(\zeta)_{|\tau = t} = \int_{\mathbb{R}^d} \frac{-\nabla(u(t,x)^m)}{u(t,x)} \cdot \nabla\zeta(x)\, d\mu_t(x).$$

Hence, considering (2.14), to prove the assertion, it is sufficient to show

$$D_{\mathrm{nice}} \subseteq D(\mathrm{diff}E), \tag{3.4}$$

$$\int_{\mathbb{R}^d} \frac{-\nabla(v(x)^m)}{v(x)} \cdot \varphi(x)\, d\nu(x) = -\mathrm{diff}E_\nu(\varphi) \tag{3.5}$$

for every $\varphi \in C_c^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\nu = v(x)dx \in D_{\text{nice}}$, and $\mu_t \in D_{\text{nice}}$ dt-a.s. Concerning (3.4), by (3.3) it is easy to see that $\mu_\tau^{\varphi,\nu} \in D(E)$ for all $\varphi \in C_c^1(\mathbb{R}^d, \mathbb{R}^d)$, $\nu = v(x)dx \in \mathcal{P}_a$ such that $v \in L^\infty(\mathbb{R}^d)$ and $|\tau| < \varepsilon = \varepsilon(\varphi, \nu) > 0$. Moreover, $\tau \mapsto E(\mu_\tau^{\varphi,\nu})$ is differentiable at $\tau = 0$, since by the transformation rule

$$\begin{aligned} E(\mu_\tau^{\varphi,\nu}) &= \int_{\mathbb{R}^d} \eta\Big(\frac{d\mu_\tau^{\varphi,\nu}}{dx}(x)\Big)\, dx \\ &= \frac{1}{m-1}\int_{\mathbb{R}^d} v(x)^m |\det D(\mathrm{Id} + \tau\varphi)(x + \tau\varphi(x))|^{-m}| \cdot |\det D(\mathrm{Id} + \tau\varphi)(x)|\, dx - \frac{m}{m-1} \end{aligned}$$

and since by Lemma 3.3 below and because $v \in L^\infty(\mathbb{R}^d)$ the integrand on the right is differentiable in $\tau \in (-\varepsilon, \varepsilon)$ for all $x \in \mathbb{R}^d$ with uniformly in τ $L^1(\mathbb{R}^d)$-bounded derivative. Hence, for $\nu \in D_{\text{nice}}$, Lemma 3.3 yields $\nu \in D(\mathrm{diff}E)$ and

$$\mathrm{diff}E_\nu(\varphi) = \frac{d}{d\tau} E(\mu_\tau^{\varphi,\nu})_{|\tau=0} = -\int_{\mathbb{R}^d} v(x)^m \,\mathrm{div}\,\varphi(x)\, dx = \Big\langle \frac{\nabla(v^m)}{v}, \varphi \Big\rangle_\nu.$$

Consequently, also (3.5) holds. It remains to show $\mu_t \in D_{\text{nice}}$ dt-a.s. To this end, recall $u(t)^{m-\frac{1}{2}} \in H^1(\mathbb{R}^d)$ dt-a.s. This follows from [7, Rem.5.4] since $m - \frac{1}{2} \geq \frac{m+1}{2}$ for $m \geq 2$. Note that [7, Rem.5.4] is restricted to $d \geq 3$. However, this restriction is only needed for the more general diffusivity functions β instead of $r \mapsto |r|^{m-1}r$ considered in [7]. More precisely, tracing through the proof of [7, Thm.5.2], one sees that $d \geq 3$ is only needed for the $L^1 - L^\infty$-regularization result of Theorem 4.1 of the same reference. But this result is true for (PME) for any $d \geq 1$, see for instance [23, Thm.9.12]. Since $u(t)^{m-\frac{1}{2}} \in H^1(\mathbb{R}^d)$ implies $u(t)^m \in W^{1,1}_{\text{loc}}(\mathbb{R}^d)$ as well as $\frac{\nabla(u(t)^m)}{u(t)} \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu_t)$, the proof is complete.

(ii) For $m \geq 3$ we have $m - 1 \geq \frac{m+1}{2}$, hence by [9, Rem.5.4], $u(t)^{m-1} \in H^1(\mathbb{R}^d)$ dt-a.s. Now the additional equality of the assertion follows from a straightforward calculation.

(iii) Any such solution $t \mapsto \nu_t$ to (GF) with E as in (i) is a solution to (PME). Hence the claim follows from the restricted uniqueness result [18, Thm.1] for (PME).

□

We used the following lemma for the previous proof. Recall that $D\psi$ denotes the Jacobian for $\psi \in C^1(\mathbb{R}^d, \mathbb{R}^d)$.

Lemma 3.3. *Let $\varphi \in C_b^1(\mathbb{R}^d, \mathbb{R}^d)$. Then, there is $\varepsilon = \varepsilon(\varphi) > 0$, which does not depend on $x \in \mathbb{R}^d$, such that for all $x \in \mathbb{R}^d$:*

(i) $\tau \mapsto \det D(\mathrm{Id} + \tau\varphi)(x)$ is differentiable on $(-\varepsilon, \varepsilon)$ with derivative $\mathrm{div}\,\varphi(x)$ at $\tau = 0$.

(ii) $\tau \mapsto \det D(\mathrm{Id} + \tau\varphi)(x + \tau\varphi(x))^{-1}$ is differentiable on $(-\varepsilon, \varepsilon)$ with derivative $-\mathrm{div}\,\varphi(x)$ at $\tau = 0$.

Moreover, both maps and their first derivatives in τ are uniformly bounded in $(\tau, x) \in (-\varepsilon, \varepsilon) \times \mathbb{R}^d$.

Proof. Choose $\varepsilon > 0$ such that $\det D(\mathrm{Id} + \tau\varphi)(x) > 0$ for all $(t, x) \in (-\varepsilon, \varepsilon) \times \mathbb{R}^d$. This is possible, since $\varphi \in C_b^1(\mathbb{R}^d, \mathbb{R}^d)$, $\det D(\mathrm{Id})(x) = 1$ and since $(t, x) \mapsto \det D(\mathrm{Id} + \tau\varphi)(x)$ is continuous.

(i) The differentiability follows directly from the Leibniz formula. Concerning the derivative at $\tau = 0$, note that in dimension $d = 1$ one has

$$\frac{d}{d\tau}\big(\det D(\mathrm{Id} + \tau\varphi)(x)\big)_{|\tau=0} = \frac{d}{d\tau}\big(1 + \tau\varphi'(x)\big)_{|\tau=0} = \varphi'(x) = \mathrm{div}\,\varphi(x).$$

Now the claim follows by induction over d and by using the Laplace expansion for the determinant of a quadratic matrix. Indeed, regrouping all summands in the Laplace expansion in terms of their order in t, the only summand of order one is $\operatorname{div}\varphi(x)$. Hence, by differentiating in t, the claim follows.

(ii) Consider the map $(t,y) \mapsto \det D(\mathrm{Id}+\tau\varphi)(y)$. Again appealing to the Leibniz formula, clearly $(\tau,y) \mapsto \det D(\mathrm{Id}+\tau\varphi)(y)$ is differentiable on $(-\varepsilon,\varepsilon)\times\mathbb{R}^d$. Hence, for $x\in\mathbb{R}^d$ fixed (but arbitrary), (i) implies

$$\begin{aligned}\frac{d}{d\tau}\big(\det D(\mathrm{Id}+\tau\varphi)(x+\tau\varphi(x))\big)^{-1}_{|\tau=0} &= -(1,\operatorname{div}\varphi(x))\cdot(\operatorname{div}\varphi(x),(\partial_y \det D(\mathrm{Id}+\tau\varphi)(y))_{|(\tau,y)=(0,x)})\\ &= -\operatorname{div}\varphi(x),\end{aligned}$$

since $(\partial_y \det D(\mathrm{Id}+\tau\varphi)(y))_{|(\tau,y)} = 0$. Indeed, by the Leibniz formula, $\partial_y \det D(\mathrm{Id}+\tau\varphi)(y)$ consists of summands of order of at least 1 in t, and hence its evaluation at any point $(0,y)$ equals 0.

The final claim follows directly from the Leibniz formula. □

Remark 3.4. *The restriction $m\geq 2$ in Theorem 3.6 is only needed to apply [7, Rem.5.4] in order to obtain $\frac{\nabla(u(t)^m)}{u(t)}\in L^2(\mathbb{R}^d,\mathbb{R}^d;\mu_t)$. Without this, we cannot prove $\nabla^{\mathcal{P}}E_{\mu_t}\in L^2(\mathbb{R}^d,\mathbb{R}^d;\mu_t)=T_{\mu_t}\mathcal{P}$, i.e. we cannot prove that $\nabla^{\mathcal{P}}E$ is a section along $t\mapsto\mu_t$ in the tangent bundle $T\mathcal{P}$. But even for $m\in(1,2)$ we have $\nabla(u(t)^m)\in L^2(\mathbb{R}^d,\mathbb{R}^d)$ and thus $\frac{\nabla(u(t)^m)}{u(t)}\in L^1_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d;\mu_t)$. Then $\nabla^{\mathcal{P}}E_{\mu_t}$ can be considered as the unique representing element in $L^1_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d;\mu_t)$ of $\mathrm{diff}E_{\mu_t}$ on $(C^1_c(\mathbb{R}^d,\mathbb{R}^d),|\cdot|_\infty)$, and solutions to* (PME) *can be understood as solutions to the gradient flow in such a generalized sense.*

Remark 3.5. *For $d\geq 3$, the assertion and the proof of Proposition 3.2 can be extended to the case of a more general diffusivity function $\beta: r\mapsto\beta(r)$ in place of $r\mapsto|r|^{m-1}r$ in* (PME)*, with the following assumptions on β.*

$$\beta\in C^2(\mathbb{R}),\ \beta(0)=0,\ |\beta(r)|\leq Cr^\alpha,\ \beta'(r)\geq C|r|^{\alpha-1}$$

for $C>0$, $\alpha\geq 1$, and β' is such that η given below belongs to $C^1(\mathbb{R})$ (which is, e.g., the case, if we have $\beta'(r)\leq C|r|^{\alpha-1}$ for some $\alpha\geq 2$). Then the energy E is given by

$$E: D(E)\subseteq\mathcal{P}_a\to\mathbb{R},\quad E(udx):=\int_{\mathbb{R}^d}\eta(u(x))dx,\quad \eta(r):=\int_0^r\int_1^s\frac{\beta'(w)}{w}dwds, r\geq 0,$$

where $D(E):=\{\nu=v(x)dx:\eta(v)\in L^1(\mathbb{R}^d)\}$. Indeed, in this case, by [7, Sect.5], for any $\mu_0\in\mathcal{P}$ there is a solution $t\mapsto\mu_t=u(t,x)dx$ to (PME) *(with $r\mapsto\beta(r)$ replacing $r\mapsto|r|^{m-1}r$) in $\bigcap_{\delta>0}L^\infty((\delta,\infty)\times\mathbb{R}^d)$ with initial datum μ_0. However, here we need to assume additionally $\mu_t\in D(E)$ dt-a.s (which again follows if, e.g., $\beta'(r)\leq C|r|^{\alpha-1}$ for some $\alpha\geq 2$). Then a similar calculation as in the proof of Theorem 3.2 shows*

$$\mathit{diff}E_{\mu_t}(\varphi)=\frac{d}{dt}E(\mu^{\varphi,\nu}_\tau)_{|\tau=0}=\int_{\mathbb{R}^d}\big(\eta(u(t,x))-\eta'(u(t,x))u(t,x)\big)\operatorname{div}\varphi(x)\,dx,\quad\forall\varphi\in C^1_c(\mathbb{R}^d,\mathbb{R}^d).$$

Since $\big(\eta(r)-\eta'(r)r\big)'=-\beta'(r)$ and since [7, Rem.5.4] yields $\beta(u(t))\in H^1(\mathbb{R}^d)$ dt-a.s., we have dt-a.s.

$$\mathit{diff}E_{\mu_t}(\varphi)=\int_{\mathbb{R}^d}\nabla(\beta(u(t,x)))\cdot\varphi(x)\,dx.$$

Hence, as in Remark 3.4, $\nabla^{\mathcal{P}}E_{\mu_t}=-\frac{\nabla(\beta(u(t)))}{u(t)}$ is the unique $L^1_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d;\mu_t)$-element representing $\mathrm{diff}E_{\mu_t}$ on $(C^1_c(\mathbb{R}^d,\mathbb{R}^d),|\cdot|_\infty)$, and $t\mapsto\mu_t$ can be understood as a solution to the gradient flow in this generalized sense.

3.2 Generalized porous media equation

Now we consider equation (gPME). We will prove that solutions to this equation solve ($\mathcal{P}$-GF) with a weighted metric tensor $\langle \cdot, \cdot \rangle_\alpha$ (see Section 2.2). Consider the following set of assumptions on the coefficients β, b and D.

Hypothesis 1

(i) $\beta \in C^1(\mathbb{R}),\ \beta(0) = 0,\ \gamma \le \beta'(r) \le \gamma_1,\ r \in \mathbb{R}$, for $0 < \gamma < \gamma_1 < \infty$.

(ii) $b \in C_b(\mathbb{R}) \cap C^1(\mathbb{R}), b \ge b_0 > 0$.

(iii) $\Phi \in C^1(\mathbb{R}^d)$, $\nabla\Phi \in C_b(\mathbb{R}^d, \mathbb{R}^d)$, $D = -\nabla\Phi$.

(iv) $(\operatorname{div} D)^- \in L^\infty(\mathbb{R}^d)$ and $(\operatorname{div} D)^+ \in (L^2(\mathbb{R}^d) + L^\infty(\mathbb{R}^d))$.

(v) $\Phi \in W^{2,1}_{\text{loc}}(\mathbb{R}^d),\ \Phi \ge 1,\ \lim\limits_{|x|\to\infty} \Phi(x) = +\infty$ and there exists $m \in [2, \infty)$ such that $\Phi^{-m} \in L^1(\mathbb{R}^d)$.

First, we only assume (i)-(iii) . The ansatz for the following energy functional comes again from Section 2.3. Set

$$\eta : \mathbb{R}_+ \to \mathbb{R}, \quad \eta(r) := \int_0^r g(s) ds := \int_0^r \int_1^s \frac{\beta'(w)}{w b(w)} dw\, ds, \tag{3.6}$$

and

$$E : D_0(E) \subseteq \mathcal{P}_a \to \mathbb{R} \cup \{+\infty\}, \quad E(v dx) := \int_{\mathbb{R}^d} \eta(v(x))\, dx + \int_{\mathbb{R}^d} \Phi(x)\, v(x) dx, \tag{3.7}$$

where $D_0(E) := \{\nu = v(x)dx \in \mathcal{P}_a \,|\, \Phi \in L^1(\mathbb{R}^d; \nu)\}$. Note that by (i) and (ii) we have for $r \ge 0$

$$\frac{\gamma_1}{b_0} \mathbb{1}_{[0,1]}(r) r(\log r - 1) + \frac{\gamma}{|b|_\infty} \mathbb{1}_{(1,\infty)}(r) r(\log r - 1) \le \eta(r) \le \frac{\gamma}{|b|_\infty} \mathbb{1}_{[0,1]}(r) r(\log r - 1) + \frac{\gamma_1}{b_0} \mathbb{1}_{(1,\infty)}(r) r(\log r - 1),$$

which yields the second equality in the next line

$$D(E) = D_0(E) \cap \{\nu = v(x)dx \in \mathcal{P}_a \,|\, E(v(x)dx) < +\infty\} = D_0(E) \cap \{\nu = v(x)dx \in \mathcal{P}_a \,|\, v \log v \in L^1(\mathbb{R}^d)\}.$$

We point out that in general E is not convex. For main results of this section, Theorems 3.6 and 3.7, we consider the weight α,

$$\alpha : \mathcal{P} \to \mathcal{B}_b(\mathbb{R}^d), \quad \alpha(\nu) := \begin{cases} x \mapsto \frac{1}{b(v(x))} & , \text{ if } \nu \in \mathcal{P}_a, \nu = v(x)dx, \\ 1 & , \text{ else}, \end{cases} \tag{3.8}$$

and we denote the corresponding metric tensor and gradient by $\langle \cdot, \cdot \rangle_b$ and $\nabla^{\mathcal{P}}_b$, i.e.

$$\langle \cdot, \cdot \rangle_b : \nu \mapsto \langle \cdot, \cdot \rangle_{b,\nu} = \left\langle \frac{1}{b(v(x))} \cdot, \cdot \right\rangle_\nu \tag{3.9}$$

for $\nu = v(x)dx \in \mathcal{P}_a$, and $\langle \cdot, \cdot \rangle_{b,\nu} = \langle \cdot, \cdot \rangle_\nu$ if $\nu \in \mathcal{P} \backslash \mathcal{P}_a$. For $G : D(G) \subseteq \mathcal{P} \to \mathbb{R}$ and $\nu \in D(\text{diff}G)$, $\nabla^{\mathcal{P}}_b F_\nu$ denotes the unique element in $L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)$ such that $\text{diff}F_\nu(\varphi) = \langle \nabla^{\mathcal{P}}_b F_\nu, \varphi \rangle_{b,\nu}$ for all $\varphi \in C^1_c(\mathbb{R}^d, \mathbb{R}^d)$. We introduce the set

$$D'_{\text{nice}} := \left\{ \nu = v(x)dx \in \mathcal{P} \,|\, v \in L^\infty(\mathbb{R}^d) \cap W^{1,1}_{\text{loc}}(\mathbb{R}^d), \frac{\beta'(v)\nabla v}{v} \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu) \right\},$$

and we will prove $D'_{\text{nice}} \cap D(E) \subseteq D(\text{diff}E)$ under the assumptions of Hypothesis 1 (see Theorem 3.7).

Theorem 3.6. *Suppose (i)-(iii) of Hypothesis 1 hold. Let $t \mapsto \mu_t = u(t,x)dx$ be a solution to* (gPME) *with initial datum $\mu_0 \in \mathcal{P}$ such that $\mu_t \in D'_{nice} \cap D(E)$ for dt-a.e. $t > 0$. Then $t \mapsto \mu_t$ solves* ($\mathcal{P}$-GF) *with E as in* (3.7) *and with the weighted metric tensor $\langle \cdot,\cdot \rangle_b$ from* (3.9). *We have $D'_{nice} \cap D(E) \subseteq D(diffE)$ and the weighted gradient of E for $\nu = v(x)dx \in D'_{nice} \cap D(E)$ is*

$$\nabla_b^{\mathcal{P}} E_\nu = \frac{\nabla(\beta(v))}{v} - b(v)D = b(v)\nabla\big(g(v) + \Phi\big) \tag{3.10}$$

in $(L^2(\mathbb{R}^d,\mathbb{R}^d;\nu), \langle \cdot,\cdot \rangle_{b,\nu})$.

Proof. For $t \mapsto \mu_t = u(t,x)dx$ as in the assertion we have for all $\zeta \in C_c^2(\mathbb{R}^d)$ and dt-a.e. $t > 0$ (with exceptional set independent of ζ)

$$\frac{d}{d\tau}\mu_\tau(\zeta)_{|\tau=t} = \int_{\mathbb{R}^d} \frac{\beta(u(t,x))}{u(t,x)}\Delta\zeta + D(x)b(u(t,x)) \cdot \nabla\zeta \, d\mu_t(x).$$

Hence, by Section 2.4, it remains to prove

$$D'_{\text{nice}} \cap D(E) \subseteq D(\text{diff}E), \tag{3.11}$$

$$-\text{diff}E_\nu(b(v)\varphi) = -\int_{\mathbb{R}^d} \left(\frac{\nabla\beta(v(x))}{v(x)} - D(x)b(v(x))\right) \cdot \varphi \, d\nu(x) \tag{3.12}$$

for every $\varphi \in C_c^1(\mathbb{R}^d,\mathbb{R}^d)$ and $\nu \in v(x)dx \in D'_{\text{nice}} \cap D(E)$. Concerning (3.11), let $\nu = v(x)dx \in D'_{\text{nice}} \cap D(E)$. Recall that for $\varphi \in C_c^1(\mathbb{R}^d,\mathbb{R}^d)$ and $|\tau| < \varepsilon = \varepsilon(\varphi) > 0$, $\mu_\tau^{\varphi,\nu} = \nu \circ (\text{Id} + \tau\varphi)^{-1}$ is absolutely continuous with density as in (3.3). Moreover, $\mu_\tau^{\varphi,\nu} \in D(E)$ for $|\tau| < \varepsilon$ (with $\varepsilon > 0$ as above). Indeed,

$$\begin{aligned}\int_{\mathbb{R}^d} |\Phi(x + \tau\varphi(x))| v(x)dx = &\int_{(\text{supp}\,\varphi)^c} |\Phi(x)|v(x)dx \\ &+ \int_{(\text{Id}+\tau\varphi)(\text{supp}\,\varphi)} |\Phi(x)|v(\text{Id}+\tau\varphi)^{-1}(x))| \det D(\text{Id}+\tau\varphi)(x)|^{-1}dx,\end{aligned}$$

and both summands on the right-hand side are finite, since $v \in D(E) \cap L^\infty(\mathbb{R}^d)$, $\Phi \in L^1_{\text{loc}}(\mathbb{R}^d)$, and due to Lemma 3.3. Hence $\mu_\tau^{\varphi,\nu} \in D_0(E)$, so by the transformation rule it remains to show, abbreviating $j(\tau,x) := |\det D(\text{Id}+\tau\varphi)(x + \tau\varphi(x))|^{-1}$, that $v(x)j(\tau,x)\log\big(v(x)j(\tau,x)\big) \in L^1(\mathbb{R}^d)$. But this follows, since $j(\tau,x)$ is uniformly in $(\tau,x) \in (-\varepsilon,\varepsilon) \times \mathbb{R}^d$ contained in an interval $(1-\delta, 1+\delta)$, where (after decreasing ε if necessary) $\delta = \delta(\varepsilon) \in (0,1/2)$, and hence

$$\int_{\mathbb{R}^d} v(x)j(\tau,x)|\log\big(v(x)j(\tau,x)\big)| \, dx \leq 2\int_{\mathbb{R}^d} v(x)(|\log v(x)| + \log 2) \, dx < \infty.$$

Moreover, $\tau \mapsto E(\mu_\tau^{\varphi,\nu})$ is differentiable in $\tau = 0$, since by the transformation rule

$$\begin{aligned}E(\mu_\tau^{\varphi,\nu}) &= \left[\int_{\mathbb{R}^d} \eta\left(\frac{d\mu_\tau^{\varphi,\nu}}{dx}(x)\right) dx + \int_{\mathbb{R}^d} \Phi(x)\left(\frac{d\mu_\tau^{\varphi,\nu}}{dx}(x)\right) dx\right] \\ &= \left[\int_{\mathbb{R}^d} \eta\Big(v(x)|\det D(\text{Id}+\tau\varphi)(x + \tau\varphi(x))|^{-1}\Big)|\det D(\text{Id}+\tau\varphi)(x)| \, dx \right.\\ &\qquad\qquad \left. + \int_{\mathbb{R}^d} \Phi(x + \tau\varphi(x)) \, d\nu(x)\right] \\ &=: I_1(\tau) + I_2(\tau),\end{aligned}$$

and the integrands of $I_i(\tau), i \in \{1,2\}$, are differentiable in $\tau \in (-\varepsilon,\varepsilon)$ for all $x \in \mathbb{R}^d$ with derivative uniformly $L^1(\mathbb{R}^d)$-, and $L^1(\mathbb{R}^d;\nu)$-bounded in τ, respectively. Indeed, for I_2, this follows immediately from $\Phi \in C^1(\mathbb{R}^d)$ and the boundedness of $\nabla\Phi$ and φ. For I_1, the claim follows by Lemma 3.3, $v \in L^\infty(\mathbb{R}^d)$ and the local boundedness of η'. Therefore, we obtain

$$
\begin{aligned}
\mathrm{diff}E_\nu(\varphi) &= \frac{d}{d\tau}E(\mu_\tau^{\varphi,\nu})_{|\tau=0} = \frac{d}{d\tau}\bigg[\int_{\mathbb{R}^d}\eta\bigg(\frac{d\mu_\tau^{\varphi,\nu}}{dx}(x)\bigg)\,dx + \int_{\mathbb{R}^d}\Phi(x)\frac{d\mu_\tau^{\varphi,\nu}}{dx}(x)dx\bigg]_{|\tau=0}\\
&= \frac{d}{d\tau}\bigg[\int_{\mathbb{R}^d}\eta\Big(v(x)|\det D(\mathrm{Id}+\tau\varphi)(x+\tau\varphi(x))|^{-1}\Big)|\det D(\mathrm{Id}+\tau\varphi)(x)|\,dx\\
&\qquad + \int_{\mathbb{R}^d}\Phi(x+\tau\varphi(x))\,d\nu(x)\bigg]_{|\tau=0}.
\end{aligned} \tag{3.13}
$$

First note

$$
\frac{d}{d\tau}\bigg[\int_{\mathbb{R}^d}\Phi(x+\tau\varphi(x))\,d\nu(x)\bigg]_{|\tau=0} = \int_{\mathbb{R}^d}\nabla\Phi\cdot\varphi(x)\,d\nu(x) = -\int_{\mathbb{R}^d}D(x)\cdot\varphi(x)\,d\nu(x).
$$

Concerning the other summand in (3.13), we find

$$
\begin{aligned}
&\frac{d}{d\tau}\bigg[\int_{\mathbb{R}^d}\eta\Big(v(x)|\det D(\mathrm{Id}+\tau\varphi)(x+\tau\varphi(x))|^{-1}\Big)|\det D(\mathrm{Id}+\tau\varphi)(x)|\,dx\bigg]_{|\tau=0}\\
&= \int_{\mathbb{R}^d}\eta'(v(x))\bigg(\frac{d}{d\tau}\Big|\det D(\mathrm{Id}+\tau\varphi)(x+\tau\varphi(x))\Big|^{-1}\bigg)_{|\tau=0}v(x)\\
&\qquad + \eta(v(x))\frac{d}{d\tau}|\det D(\mathrm{Id}+\tau\varphi)(x)|_{|\tau=0}\,dx\\
&= \int_{\mathbb{R}^d}\Big(\eta(v(x)) - \eta'(v(x))v(x)\Big)\,\mathrm{div}\,\varphi(x)\,dx,
\end{aligned}
$$

where we used Lemma 3.3 for the final equality. Since $v \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ and since $\eta(r) - \eta'(r)r$ is differentiable with derivative $= -\frac{\beta'(r)}{b(r)}$, by the divergence theorem we obtain

$$
\int_{\mathbb{R}^d}\Big(\eta(v(x)) - \eta'(v(x))v(x)\Big)\,\mathrm{div}\,\varphi(x)\,dx = \int_{\mathbb{R}^d}\frac{\beta'(v(x))\nabla v(x)}{b(v(x))}\cdot\varphi\,dx = \int_{\mathbb{R}^d}\frac{\beta'(v(x))\nabla v(x)}{v(x)b(v(x))}\cdot\varphi\,d\nu(x).
$$

Hence, since $\frac{\beta'(v)\nabla v}{v} \in L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$, the map

$$
\varphi \mapsto \mathrm{diff}E_\nu(\varphi) = \int_{\mathbb{R}^d}\bigg(\frac{\beta'(v(x))\nabla v(x)}{v(x)b(v(x))} - D(x)\bigg)\cdot\varphi(x)\,d\nu(x)
$$

is linear and continuous on $C^1_c(\mathbb{R}^d,\mathbb{R}^d)$ with respect to the $L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$-topology, Thus we have shown (3.11)-(3.12) and the proof is complete. □

Now assume all of Hypothesis 1. In this case, (gPME) has a unique bounded solution $t \mapsto u(t,x)dx$ for every initial datum $\mu_0 \in L^\infty(\mathbb{R}^d)\cap D(E)$, and, as we will show, this solution belongs to $D'_{\mathrm{nice}}\cap D(E)$. Consequently, Theorem 3.6 applies. More precisely, we have Theorem 3.7 below. We point out that the assumption $\beta' \le \gamma_1$ from (i) of Hypothesis 1 is not needed here, since $\beta'(u)$ is bounded due to the boundedness of u, even if β' is only locally bounded. Moreover, in this case we obtain a uniqueness result for the gradient flow, comparable to Theorem 3.2, since for solutions in $L^1_{\mathrm{loc}}([0,\infty)\times\mathbb{R}^d)$ all integrability conditions from Definition 3.1 are fulfilled.

Theorem 3.7. *Suppose Hypothesis 1 is satisfied and let $\mu_0 \in L^\infty(\mathbb{R}^d)\cap D(E)$.*

(i) There is a unique solution to (gPME) *in $L^\infty([0,\infty)\times\mathbb{R}^d$ with initial datum μ_0, and it solves* ($\mathcal{P}$-GF) *with E from* (3.7) *and weighted metric tensor $\langle\cdot,\cdot\rangle_b$. We have $D'_{nice}\cap D(E) \subseteq D(\mathit{diff}E)$, and the formula for $\nabla^{\mathcal{P}}_b E_\nu$ for $\nu \in D'_{nice}\cap D(E)$ is as in* (3.10).

(ii) *Moreover, the solution from (i) is the unique solution to this gradient flow in* $L^\infty([0,\infty)\times\mathbb{R}^d)\cap D'_{nice}\cap D(E)$ *with initial datum* μ_0.

Proof. (i) The existence and uniqueness assertion follows from [9, Prop.2.2] (there it is proven that such solutions exist and are unique for every $\mu_0 \in L^1(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)$). Hence it remains to prove $\mu_t \in D'_{\text{nice}}\cap D(E)$ dt-a.s. for this solution $t\mapsto \mu_t = u(t,x)dx$. First, $\mu_t\in D(E)\cap W^{1,1}_{\text{loc}}(\mathbb{R}^d)$ dt-a.s. follows from [9, Thm.4.1]. Furthermore, by [9, Thm.4.1.,Eq.(4.7)] and since $b(u(t))D\in L^\infty(\mathbb{R}^d,\mathbb{R}^d)$, the triangle inequality and $u(t)\in L^1(\mathbb{R}^d)$ yield $\beta'(u(t))\nabla u(t)(u(t))^{-1}\in L^2(\mathbb{R}^d,\mathbb{R}^d;\mu_t)$ dt-a.s.

(ii) It is easily seen that any such solution to the gradient flow is a solution to the FPK equation. Hence the claim follows from the restricted uniqueness result in (i).

□

Remark 3.8 (Stationary solutions to (gPME))**.** *If in addition to Hypothesis 1 also the "balance condition"*

$$\gamma_1\Delta\Phi(x) - b_0|\nabla\Phi(x)|^2 \le 0, \quad dx-a.s., \tag{3.14}$$

holds, it was shown in [9] that E from (3.7) *is a Lyapunov function for the solution* $t\mapsto\mu_t$*, i.e.* $E(\mu_t)\le E(\mu_s)$ *for all* $0\le s\le t$*. Moreover, it is proven there that* μ_t *converges to an equilibrium* u_∞ *in* $L^1(\mathbb{R}^d)$ *as* $t\to\infty$*, which is a stationary solution to* (gPME) *and can be calculated from E via*

$$u_\infty(x) = g^{-1}\big(-\Phi + c\big),$$

where $g(r)=\eta'(r)$ *with* η *from* (3.6) *and* $c\in\mathbb{R}$ *is a uniquely determined constant. A similar result for degenerate diffusivities* β *was obtained in [8]. That* $t\mapsto E(\mu_t)$ *is decreasing is, at least heuristically, also implied by Remark 2.4. This suggests that our identification of the energy E provides an ansatz to find Lyapunov functions — and hence stationary solutions — in more general cases than those covered in [9].*

Remark 3.9. *We would like to point out that if* $\beta(r)=\sigma r$*,* $\sigma\in(0,\infty)$ *and* $b(r)=b_0\in(0,\infty)$*, then* $\eta(r)=\frac{\sigma}{b_0}r(\log r-1)$*,* $r\ge 0$*, i.e. in this case E in* (3.7) *is the classical Boltzmann entropy function.*

To conclude this subsection, we want to prove that in the situation of Theorem 3.6 for each fixed $\nu = v(x)dx\in D'_{\text{nice}}$, the gradient field $\nabla^{\mathcal{P}}_b E_\nu$ can be approximated by weighted smooth gradient fields in $L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$. To this end, we define

$$G^\nu := \text{closure of } \big\{b(v)\nabla\zeta \,\big|\, \zeta\in C_c^\infty(\mathbb{R}^d)\big\} \text{ in } (L^2(\mathbb{R}^d,\mathbb{R}^d;\nu),\langle\cdot,\cdot\rangle_{b,\nu}) \tag{3.15}$$

(see (3.9)). Then $T_\nu\mathcal{P} = L^2(\mathbb{R}^d,\mathbb{R}^d;\nu) = G^\nu\oplus(G^\nu)^{\perp,\langle\cdot,\cdot\rangle_{b,\nu}}$, where $(G^\nu)^{\perp,\langle\cdot,\cdot\rangle_{b,\nu}}$ denotes the orthogonal complement of G^ν with respect to $\langle\cdot,\cdot\rangle_{b,\nu}$.

Proposition 3.10. *Let* $\nu = v(x)dx\in D'_{nice}$*. Then* $\nabla^{\mathcal{P}}_b E_\nu\in G^\nu$*.*

Proof. We know by Theorem 3.6 that $\nabla^{\mathcal{P}}_b E_\nu = b(v)\nabla\big(g(v)+\Phi\big)$, with $\nabla\big(g(v)+\Phi\big)\in L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$. By Hypothesis 1 (i) we may assume $b\equiv 1$. Set $w := g(v)+\Phi$ and $w_N := (w\wedge N)\vee(-N)$, $N\in\mathbb{N}$. Then, as $N\to\infty$,

$$\nabla w_N = \mathbb{1}_{\{-N\le w\le N\}}\nabla w_N \xrightarrow{N\to\infty} \nabla w \text{ in } L^2(\mathbb{R}^d,\mathbb{R}^d;\nu).$$

Hence, since $\nu\in D'_{\text{nice}}$, by [22, Thm.3.1], there exist $\zeta_n\in C_c^\infty(\mathbb{R}^d)$, $n\in\mathbb{N}$, such that

$$\zeta_n \xrightarrow{n\to\infty} w_N \text{ in } L^2(\mathbb{R}^d;\nu),$$

$$\nabla\zeta_n \xrightarrow{n\to\infty} \nabla w_N \text{ in } L^2(\mathbb{R}^d,\mathbb{R}^d;\nu),$$

and the assertion follows. □

3.3 More general divergence-type equations with x-dependent drift

Theorem 3.6 can be extended to the more general divergence-type equation

$$\partial_t u = \operatorname{div}\big(A(u)\nabla u\big) - \operatorname{div}\big(B(u)D(x)u\big), \tag{3.16}$$

where $A, B : \mathbb{R} \to \mathbb{R}^{d\times d}$ are matrix-valued. More precisely, consider the following assumptions.

Hypothesis 2.

(i) $B \in L^\infty(\mathbb{R}, \mathbb{R}^{d\times d})$, $|B| \geq b_0 > 0$, $B(r)$ is invertible for all $r \in \mathbb{R}_+$, and $B(r)^{-1}A(r) = \Psi(r)\,\mathrm{Id}$, where $\Psi : \mathbb{R} \to \mathbb{R}$, $\Psi \in C(\mathbb{R})$ and $c_0 \leq \Psi \leq c_1$ for $c_i > 0, i \in \{1, 2\}$.

(ii) D satisfies (iii) of Hypothesis 1.

The proof of the following result is analogous to the proof of Theorem 3.6. Here, we define

$$D''_{\text{nice}} := \Big\{\nu = v(x)dx \in \mathcal{P} \,\Big|\, v \in L^\infty(\mathbb{R}^d) \cap W^{1,1}_{\text{loc}}(\mathbb{R}^d),\, \frac{A(v)\nabla v}{v} \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu)\Big\}.$$

Proposition 3.11. *Let Hypothesis 2 be fulfilled. Then any solution $t \mapsto \mu_t = u(t, x)dx$ to (3.16) such that $u(t) \in L^\infty(\mathbb{R}^d) \cap W^{1,1}_{loc}(\mathbb{R}^d)$, $\frac{A(u(t,\cdot))}{u(t,\cdot)}\nabla u(t, \cdot) \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu_t)$ and $\mu_t \in D(E)$ dt-a.s. is a solution to the gradient flow on $\mathcal{P}$ with E,*

$$E : D(E) \subseteq \mathcal{P}_a \to \mathbb{R}, \quad E(udx) := \int_{\mathbb{R}^d} \eta(u(x))dx - \int_{\mathbb{R}^d} \Phi(x)u(x)dx,$$
$$D(E) := \{\nu \in \mathcal{P}_a \,|\, \nu = v(x)dx, \Phi \in L^1(\mathbb{R}^d; \nu), v \log v \in L^1(\mathbb{R}^d)\},$$

where

$$\eta(r) := \int_0^r \int_1^s \frac{\Psi(w)}{w} dw\, ds, \quad r \in \mathbb{R}, \tag{3.17}$$

and with metric tensor

$$\langle\cdot,\cdot\rangle_B : \nu \mapsto \langle\cdot,\cdot\rangle_{B,\nu} := \langle B(v)^{-1}\cdot,\cdot\rangle_\nu. \tag{3.18}$$

The weighted gradient for $\nu = v(x)dx$ from $D''_{nice} \cap D(E)$ is

$$\nabla^{\mathcal{P}}_B E_\nu = \frac{A(v)\nabla v}{v} - B(v)D \tag{3.19}$$

in $\big(L^2(\mathbb{R}^d, \mathbb{R}^d; \nu), \langle\cdot,\cdot\rangle_{B,\nu}\big)$.

Existence of solutions to (3.16) with the properties from Proposition 3.11 seems an open question. In the simpler case that D is independent of x, well-posedness of entropic solutions u to (3.16) with $u(t) \in L^\infty(\mathbb{R}^d) \cap W^{1,1}_{\text{loc}}$ has, e.g., been obtained in [10], but we do not know of a general result ensuring the additional assumptions of Proposition3.11.

Appendix Differential geometry on $\mathcal{P}$ via natural charts

In order to further convince the reader that the differential geometry from Section 2.2 is natural, here we present its deduction from rigorous differential geometric principles. More precisely, starting with a natural chart on $\mathcal{P}$, we obtain our geometry in analogy to classical Riemannian case. Since we consider $\mathcal{P}$ as an infinite-dimensional manifold with the topology of weak convergence of probability measures, i.e. the initial topology of $\nu \mapsto \nu(\zeta)$, $\zeta \in C_c^\infty(\mathbb{R}^d)$, a natural global chart $\pi : \mathcal{P} \to \mathbb{R}^\infty$ is

$$\pi : \mathcal{P} \to \pi(\mathcal{P}) \subseteq \mathbb{R}^\infty, \pi(\nu) := (\nu(g_i))_{i\in\mathbb{N}},$$

where $\{g_i, i \in \mathbb{N}\} \subseteq C_c^\infty(\mathbb{R}^d)$ is dense. The coordinate maps are $\pi^{(i)}(\nu) := \nu(g_i)$, $i \in \mathbb{N}$, and we also set $\pi_k(\nu) := (\nu(g_1), \ldots, \nu(g_k))$. There is no natural choice for $\{g_i, i \in \mathbb{N}\}$, i.e. the coordinates $\pi^{(i)}$ do not have intrinsic geometric meaning, and ultimately we will not rely on such a choice.

Test functions. The natural test functions on $(\mathcal{P},\pi)$ are cylindrical C_b^1-maps $f\circ\pi_k$, $k\in\mathbb{N}$, $f\in C_b^1(\mathbb{R}^k)$, i.e. we are led to consider the class

$$\tilde{\mathcal{F}}C_b^1 := \big\{F:\mathcal{P}\to\mathbb{R}\,|\,F(\nu)=f\big(\nu(g_1),\dots,\nu(g_k)\big), k\geq 1, g_i \text{ as above }, f\in C_b^1(\mathbb{R}^k)\big\}.$$

However, since the choice of $\{g_i\}_{i\geq 1}$ was arbitrary, it is reasonable to replace g_i within the definition of $\tilde{\mathcal{F}}C_b^1$ by arbitrary $h_i\in C_c^\infty(\mathbb{R}^d)$. Since our differential structure is of at most second order, we even allow $h_i\in C_c^2(\mathbb{R}^d)$, which leads to the test function class $\mathcal{F}C_b^2$, see (2.3).

Tangent spaces. In conjunction with the test function class $\tilde{\mathcal{F}}C_b^1$, a curve $(-\varepsilon,\varepsilon)\ni t\mapsto\mu_t$ on $(\mathcal{P},\pi)$ is differentiable, if each $t\mapsto\pi^{(i)}\circ\mu_t$, $i\in\mathbb{N}$, is differentiable. For a differentiable curve with $\mu_0=\nu$ and for $F\in\tilde{\mathcal{F}}C_b^1$, set

$$\frac{d}{dt}\mu_{t|t=0}(F) := \frac{d}{dt}F(\mu_t)_{|t=0} = (\nabla f)(\nu(g_1),\dots,\nu(g_k))\cdot\frac{d}{dt}(\pi_k\circ\mu_t)_{|t=0},$$

and note

$$\frac{d}{dt}\mu_{t|t=0}(FG) = F(\nu)\frac{d}{dt}\mu_{t|t=0}(G)+G(\nu)\frac{d}{dt}\mu_{t|t=0}(F),\quad \forall F,G\in\tilde{\mathcal{F}}C_b^1,$$

i.e. $\frac{d}{dt}\mu_{t|t=0}$ is a derivation. The equivalence class of differentiable curves $\tilde{\mu}$ with $\frac{d}{dt}(\pi_k\circ\tilde{\mu}_t)_{|t=0} = \frac{d}{dt}(\pi_k\circ\mu_t)_{|t=0}$ for all $k\in\mathbb{N}$ is denoted by $[\frac{d}{dt}\mu_t]$, and the tangent space $\tilde{T}_\nu\mathcal{P}$ at ν is the set of all such equivalence classes. As usual, the associated tangent bundle is $\tilde{T}\mathcal{P} := \bigsqcup_{\mu\in\mathcal{P}}\tilde{T}_\mu\mathcal{P}$. In general, it is hard to further characterize elements of $\tilde{T}_\nu\mathcal{P}$, and it is not clear whether these spaces are Hilbert.

Therefore, in order to obtain Hilbert tangent spaces, we restrict to suitable sub-tangent spaces $T_\nu\mathcal{P}\subseteq\tilde{T}_\nu\mathcal{P}$ as follows. For $\nu\in\mathcal{P}$ and $\varphi\in L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$, the curve

$$(-1,1)\ni t\mapsto\mu_t^{\varphi,\nu} := \nu\circ(\mathrm{Id}+t\varphi)^{-1}$$

is differentiable with $\mu_0^{\varphi,\nu}=\nu$, and

$$\frac{d}{dt}(\pi^{(i)}\circ\mu_t^{\varphi,\nu}) = \frac{d}{dt}\int_{\mathbb{R}^d} g_i(\mathrm{Id}+t\varphi)\,d\nu = \int_{\mathbb{R}^d}\nabla g_i(\mathrm{Id}+t\varphi)\cdot\varphi\,d\nu. \tag{A.1}$$

We set

$$T_\nu\mathcal{P} := \left\{\frac{d}{dt}\mu_t^{\varphi,\nu}, \varphi\in L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)\right\},$$

and thus $T_\nu\mathcal{P}\subseteq\tilde{T}_\nu\mathcal{P}$ in the sense that $\big[\frac{d}{dt}\mu_t^{\varphi,\nu}\big]\in\tilde{T}_\mu\mathcal{P}$ for all $\varphi\in L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$. $T_\nu\mathcal{P}$ is isomorphic to $L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$, thus we identify $T_\nu\mathcal{P}$ and $L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$ and endow $T_\nu\mathcal{P}$ with the standard $L^2(\mathbb{R}^d,\mathbb{R}^d;\nu)$-inner product $\langle\cdot,\cdot\rangle_\nu$. Since (A.1) still holds when one replaces $\pi^{(i)}\circ\mu_t^{\varphi,\nu}$ by $\mu_t^{\varphi,\nu}(h)$ for any $h\in C_c^2(\mathbb{R}^d)$, elements in $T_\nu\mathcal{P}$ act as derivations not only on $\tilde{\mathcal{F}}C^1$, but on $\mathcal{F}C_b^2$. Hence we obtain the tangent spaces $T_\nu\mathcal{P}$, tangent bundle $T\mathcal{P}$ and metric tensor $\langle\cdot,\cdot\rangle:\nu\mapsto\langle\cdot,\cdot\rangle_\nu$ from (2.4)-(2.5).

Differential and Gradient. Following Riemannian differential geometry, for $\nu\in\mathcal{P}$ and $F\in\tilde{\mathcal{F}}C^1$, the differential of F at ν, $\widetilde{\mathrm{diff}}F_\nu:\tilde{T}_\nu\mathcal{P}\to\mathbb{R}$, is a cotangent element with action on $[\frac{d}{dt}\mu_t]\in\tilde{T}_\nu\mathcal{P}$ via

$$\widetilde{\mathrm{diff}}F_\nu\left(\big[\frac{d}{dt}\mu_t\big]\right) := \frac{d}{dt}\mu_t(F) = \frac{d}{dt}F(\mu_t)_{|t=0}.$$

For $\varphi\in T_\mu\mathcal{P}\subseteq\tilde{T}_\mu\mathcal{P}$ and $F(\nu)=f(\nu(g_1),\dots,\nu(g_k))$, we have

$$\widetilde{\mathrm{diff}}F_\nu(\varphi) = \frac{d}{dt}F(\mu_t^{\varphi,\nu})_{|t=0} = \sum_{i=1}^k\partial_i f(\nu(g_1),\dots,\nu(g_k))\langle\nabla g_i,\varphi\rangle_\nu,$$

and the same formula holds for $F \in \mathcal{F}C_b^2$, i.e. when the coordinates $g_1, \dots, g_k$ are replaced by arbitrary $h_1, \dots, h_k \in C_c^2(\mathbb{R}^d)$. Hence we define $\mathrm{diff}F_\nu$ for every $(\nu, F) \in \mathcal{P} \times \mathcal{F}C_b^2$ as in (2.6). Since our sub-tangent spaces $T_\mu\mathcal{P}$ are Hilbert, there is a natural notion of gradient: For $F \in \mathcal{F}C_b^2$, its gradient $\nabla^{\mathcal{P}} F : \nu \mapsto \nabla^{\mathcal{P}} F_\nu \in T_\nu\mathcal{P}$ is the unique section in the tangent bundle $T\mathcal{P}$ such that

$$\left\langle \nabla^{\mathcal{P}} F_\nu, \varphi \right\rangle_\nu = \mathrm{diff}F_\nu(\varphi), \quad \forall \varphi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \nu),$$

which is the gradient from (2.7)-(2.8). This characterization of the gradient is, of course, based on the Hilbert space structure of $T_\nu\mathcal{P}$, which is one main reason for restricting to these sub-tangent spaces.

Acknowledgement. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1283/2 2021 - 317210226.

References

[1] S. Albeverio, A. Daletskii, Y. Kondratiev, E. Lytvynov. Laplace operators in deRham complexes associated with measures on configuration spaces, *J. Geom. Phys.*, 47(2):259-302, 2003.

[2] S. Albeverio, Y. Kondratiev, M. Röckner. Differential geometry of Poisson spaces, *C. R. Acad. Sci Paris Sér. I Math.*, 323:1129-1134, 1996.

[3] S. Albeverio, Y. Kondratiev, M. Röckner. Canonical Dirichlet operator and distorted Brownian motion on Poisson spaces, *C. R. Acad. Sci Paris Sér. I Math.*, 323:1179-1184, 1996.

[4] S. Albeverio, Y. Kondratiev, M. Röckner. Analysis and geometry on Configuration Spaces, *J. Funct. Anal.*, 154:444-500, 1998.

[5] S. Albeverio, Y. Kondratiev, M. Röckner. Analysis and geometry on configuration spaces: The Gibbsian Case, *J. Funct. Anal.*, 157:242-291, 1998.

[6] L. Ambrosio, N. Gigli, G. Savare. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures.* Birkhäuser Basel, 2005.

[7] V. Barbu, M. Röckner. Solutions for nonlinear Fokker–Planck equations with measures as initial data and McKean-Vlasov equations. *J. Funct. Anal.*, 280(7):108926, 2021.

[8] V. Barbu, M. Röckner. The invariance principle for nonlinear Fokker–Planck equations. *J. Differential Equations*, 315:200-221, 2022.

[9] V. Barbu, M. Röckner. The evolution to equilibrium of solutions to nonlinear Fokker–Planck equation. *Indiana Univ. Math. J.*, 72(1):89-131, 2023.

[10] G.-Q. Chen, B. Perthame. Well-posedness for non-isotropic degenerate parabolic-hyperbolic equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 20(4):645-668, 2003.

[11] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. H. Poincaré Probab. Statist.*, 46(1):1-23, 2010.

[12] A. Figalli, F. Glaudo. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows.* EMS textbooks in mathematics, EMS Press, 2021.

[13] R. Jordan, D. Kinderlehrer, F. Otto. The Variational Formulation of the Fokker–Planck Equation. *SIAM J. Math. Anal.*, 29(1):1-17, 1998.

[14] Z.-M. Ma, M. Röckner. Construction of diffusions on configuration spaces. *Osaka J. Math.*, 37(2):273-314, 2000.

[15] F. Otto. The Geometry of Dissipative Evolution Equations: The Porous Medium Equation. *Comm. Partial Differential Equations*, 26(1-2):101-174, 2001.

[16] L. Overbeck, M. Röckner. Geometric aspects of finite and infinite-dimensional Fleming–Viot processes. *Rand. Op. Stoch. Eq.* 5:35-58, 1997.

[17] L. Overbeck, M. Röckner, B. Schmuland. An Analytic Approach to Fleming-Viot Processes with Interactive Selection. *Ann. Probab.*, 23(1):1-36, 1995.

[18] M. Pierre. Uniqueness of the solutions of $u_t - \Delta\phi(u) = 0$ with initial datum a measure. *Nonlinear Anal.*, 6(2):175-187, 1982.

[19] P. Ren, M. Röckner, F.-Y. Wang. Linearization of nonlinear Fokker–Planck equations and applications. *J. Differential Equations*, 322:1-37, 2022.

[20] M. Röckner. *Stochastic analysis on configuration spaces: basic ideas and recent results.* New directions in Dirichlet forms (J. Jost et al., eds),157-231. Studies in Advanced Mathematics, International Press, 1998.

[21] M. Röckner. *Dirichlet Forms on Infinite-Dimensional 'Manifold-Like' State Spaces: A Survey of Recent Results and Some Prospects for the Future.* Probability Towards 2000,287-306. Springer New York, New York, 1998.

[22] M. Röckner, T. Zhang. Uniqueness of Generalized Schrödinger Operators, Part II. *J. Funct. Anal.*, 119:455-467, 1994.

[23] J. L. Vázquez. *The porous medium equation.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, Oxford, 2007.

[24] C. Villani. *Optimal Transport: Old and New.* Springer Berlin Heidelberg, 2008.