SECOND-ORDER ASYMPTOTICS FOR THE FAST-DIFFUSION EQUATION[*]

ROBERT J. MCCANN[†] AND DEJAN SLEPČEV[‡]

ABSTRACT. In many dissipative settings, initial disturbances will gradually disappear and all but their crudest features — such as size and location — will eventually be forgotten. Quantifying the rate at which this information is lost is sometimes a question of central interest. Here this question is addressed for the fast-diffusion equation

$$\frac{\partial v}{\partial \tau} = \Delta(v^{1-\frac{2}{n+p}}), \qquad \tau, v(\mathbf{y}, \tau) \geq 0, \quad \mathbf{y} \in \mathbf{R}^n,$$

in the range $p \geq n \geq 2 \neq p$ of nonlinearities. We use displacement convexity and dissipation of entropy to show any two solutions starting with finite p-th moments and the same total mass converge at rate

$$\|v(\cdot, \tau) - \tilde{v}(\cdot, \tau)\|_{L^1(\mathbf{R}^n)} = \frac{a}{\tau^\alpha}|\mathbf{z}_0 - \tilde{\mathbf{z}}_0| + O\left(\frac{1}{\tau^{1-\delta}}\right) \text{ as } \tau \to \infty,$$

for any $\delta > 0$. Here $\alpha = \frac{1}{2}(1 + \frac{n}{p})$, a depends only on n, p, and $\|v_0\|_{L^1(\mathbf{R}^n)}$, while $\mathbf{z}_0$ denotes the center of mass of $v_0(\cdot) = v(\cdot, 0)$. In contrast, for $|p| < n$ we show the entropy is not displacement semiconvex, even near equilibrium.

1. INTRODUCTION

This investigation concerns long time asymptotics of the nonlinear diffusion equation

(1)
$$\frac{\partial v}{\partial \tau} = \Delta(v^m) \qquad \text{in} \quad \mathbf{R}^n \times]0, \infty[,$$
$$v(\cdot, 0) = v_0(\cdot) \geq 0 \qquad \text{on} \quad \mathbf{R}^n,$$

governing a density $v(\mathbf{y}, \tau) \geq 0$ in the fast-diffusion regime $\frac{n}{n+2} < m < 1$. In the more familiar regime $m > 1$ and with suitable boundary conditions, this *porous medium equation* models population spreading [29], groundwater flows [8] [30] [6]

---

and thermalization in plasmas [43]; vanishing of the diffusion coefficient mv^{m-1} at the free boundary ensures compactly supported initial data v_0 retain compact support at each subsequent time. By contrast, the *fast-diffusion* equation $m < 1$ models curvature-driven evolution [40] [35] and avalanches in sandpiles [9] [14], where divergence of the same coefficient ensures that compactly supported initial data instantaneously develop thick tails: $v(\mathbf{y}, \tau) \sim c_{m,\tau}|\mathbf{y}|^{n+p}$ for $\mathbf{y}$ large. The number $p := -n + \frac{2}{1-m}$ of spatial moments these solutions possess frequently serves as a convenient proxy for the exponent $m = 1 - \frac{2}{n+p}$ [18].

Although most of our conclusions require $p \geq n \geq 2 \neq p$, Herrero & Pierre [21] showed for all $p > \max\{2 - n, 0\}$ that (1) is well-posed among nonnegative $L^1(\mathbf{R}^n)$ initial data and has smooth, positive solutions for $\tau > 0$; (more general initial data $v_0 \geq 0$ were treated in [34] [16]). The smoothness follows from Aronson & Bénilan's estimate [3], while Bénilan & Crandall [7] deduced that the mass of these solutions remains constant if $p > 0$. The equations possess a family of *scaling* solutions

$$(2) \qquad V(\mathbf{y}, \tau) = (\tfrac{\alpha}{\tau})^{n\alpha} \rho((\tfrac{\alpha}{\tau})^{\alpha}\mathbf{y})$$

where $\alpha = \frac{1}{2}(1 + \frac{n}{p})$ and ρ is the *Barenblatt-Pattle self-similar* profile [5] [33]:

$$\rho(\mathbf{y}) = \left(C + \frac{|\mathbf{y}|^2}{n + p - 2} \right)_+^{-\frac{n+p}{2}}$$

with C a positive constant and $(r)_+ = \max\{r, 0\}$. For $p > 0$ an appropriate choice of C will yield a scaling solution with any desired finite positive mass. An elementary but tedious calculation [22, Lemmas 3–4] shows that two scaling solutions of the same mass differ by

$$(3) \qquad \|V(\cdot - \mathbf{y}_0, \tau - \tau_0) - V(\cdot, \tau)\|_{L^1(\mathbf{R}^n)} = a\frac{|\mathbf{y}_0|}{\tau^{\alpha}} + b\frac{\tau_0}{\tau} + O\left(\frac{1}{\tau^{2\alpha}} + \frac{1}{\tau^2}\right)$$

as $\tau \to \infty$, where

$$(4) \qquad a = \alpha^{\alpha} \left\|\frac{\partial \rho}{\partial y_1}\right\|_{L^1(\mathbf{R}^n)} \quad \text{and} \quad b = \begin{cases} \alpha\|\nabla \cdot (\mathbf{y}\rho)\|_{L^1(\mathbf{R}^n)} & \text{if } \mathbf{y}_0 = \mathbf{0} \\ 0 & \text{otherwise.} \end{cases}$$

Our task will be to show that the difference $\|v(\cdot, \tau) - \tilde{v}(\cdot, \tau)\|_{L^1(\mathbf{R}^n)}$ between two more general solutions with the same mass as ρ, shares the asymptotic expansion (3)–(4) up to order $O(1/\tau^{1-\delta})$ for any $\delta > 0$, with $\mathbf{y}_0 = \mathbf{z}_0 - \tilde{\mathbf{z}}_0$ denoting the displacement between the centers of mass of v_0 and $\tilde{v}_0$. We are successful, assuming $p \geq n \geq 2 \neq p$ and sufficiently small tails of v_0 and $\tilde{v}_0$.

As time increases the solutions of (1) are known to spread out, while their L^∞ norm decays to zero. But as they spread out, their shape becomes more and more like that of the scaling solution, which itself spreads while preserving its shape. This was

shown by Friedman & Kamin [20] in the L^1 sense for $p \notin [-n, 0]$, and more recently by Vazquez [39] in the *relative* L^∞ sense

$$(5) \qquad \left\| \frac{v(\cdot, \tau)}{V(\cdot, \tau)} - 1 \right\|_{L^\infty(\mathbf{R}^n)} = o(1) \quad \text{as} \quad \tau \to \infty$$

for $p > 0$. Much recent interest has been devoted to quantifying the rate of this attraction to the self-similar shape. In one-dimension, a large-time asymptotic expansion was predicted by Zel'dovich & Barenblatt [42] for the porous medium regime $0 > p + n = p + 1$, and subsequently amended and verified by Angenent [2]. For $|p| > n$ in higher dimensions, the sharp nonlinear rate of convergence $1/\tau^\alpha$ was found simultaneously and independently by Dolbeault & del Pino [19] and Otto [32] (and by Carrillo & Toscani [10] as well in the porous medium regime): it corresponds to the rate of convergence of two translations of the scaling solution. For $p \in {]0, n[}$ the sharp rate of convergence has proved more elusive: Carrillo & Vazquez showed this rate was bounded below by $1/\tau^{1/2}$ by combining the Aronson-Benilan inequality [3] with a spectral gap estimate of Carrillo, Lederman, Markowich & Toscani [12], but this bound is not sharp. Only for radially symmetric initial data were they able to obtain the correct bound $O(1/\tau)$, by using integral-comparison techniques. Independently, the complete linearized evolution around the self-similar profile was computed for $p > 0$ by Denzler & McCann [17] [18]. In particular, they too discovered an eigenvalue crossing at $p = n$, where dilations replace translations as the slowest mode to converge, and the subsequent disappearance of this lowest eigenvalue into continuous spectrum when $p \in {]0, 2]}$. This spectrum predicts the rate of convergence in 2-Wasserstein distance. However, in the fastest conservative range of nonlinearities $0 < p \leq 2$, Kim & McCann [23] [22] prove the sharp rate of convergence to be $O(1/\tau)$ with respect to the L^1 and the relative L^∞ norm (5). This result can be extended to a larger range of nonlinearities $p > 0$, but only at the expense of assuming more and more restrictive moment conditions on the initial data, which appears to be an artifact of their Newtonian potential comparison technique.

Apart from one-dimension [2], these results address only the leading order rate of decay of $\|v(\cdot, \tau) - \tilde{v}(\cdot, \tau)\|_{L^1(\mathbf{R}^n)}$. The purpose of the present manuscript is to specify the coefficient of this leading order term, and moreover its asymptotic accuracy $O(1/\tau^{1-\delta})$ to within an arbitrarily small error $\delta > 0$ of the second order correction. We achieve this, by showing the second distinct eigenvalue found by Denzler & McCann correctly predicts the rate of convergence of the nonlinear evolution to the scaling solution in the $p \geq n \geq 2 \neq p$ range of fast-diffusion exponents, when one mods out translations by centering the initial data

$$(6) \qquad \int_{\mathbf{R}^n} \mathbf{y} \, v_0(\mathbf{y}) \, d\mathbf{y} = \mathbf{0}.$$

This centering allows us to obtain a faster rate of convergence than that found by Dolbeault, del Pino and Otto. Up to $\delta > 0$, it matches Kim & McCann's rate from the range of exponents $p \in \,]0, 2]$ where centering is unnecessary.

The potential for such improvements was also explored in the $n = 1 < -p$ study of Witelski and Bernoff [41]. Although we believe the same rate should hold in the range $p \in \,]2, n[$, this range represents a gap in our understanding when $n \geq 3$. For the range of exponents $p \in \,]0, n[$, we give a negative result which sheds some light on the difficulty: we show despite the fact that the second derivatives of the entropy (11) are bounded below at the fixed point $u = \rho$ of the rescaled dynamics, no such bound extends to any neighborhood of ρ in the obvious topologies. This gives some indication of why entropy dissipation methods have met with only limited success [12] [11] in the regime $p \in \,]0, n[$.

2. STRATEGY

As time increases the solutions of (1) spread out, tending to vanish in L^∞ norm. But as they spread out they approach the shape of the self-similar solution (which itself is spreading). To capture this asymptotic behavior it is traditional to rescale the solutions in a time-dependent manner so that Barenblatt solutions (shifted by unit time) become stationary solutions. Therefore, we introduce the *similarity variables*

$$(7) \qquad \mathbf{x} = (1 + \frac{\tau}{\alpha})^{-\alpha}\mathbf{y} \quad \text{and} \quad t = \alpha \ln(1 + \frac{\tau}{\alpha})$$

and

$$u(\mathbf{x}, t) := e^{nt} v\left(e^t \mathbf{x}, \alpha e^{t/\alpha} - \alpha\right),$$

so that

$$(8) \qquad v(\mathbf{y}, \tau) = (1 + \frac{\tau}{\alpha})^{-n\alpha} u\left((1 + \frac{\tau}{\alpha})^{-\alpha}\mathbf{y}, \alpha \ln(1 + \frac{\tau}{\alpha})\right).$$

The rescaled quantity u satisfies the following nonlinear Fokker-Planck equation

$$(9) \qquad \frac{\partial u}{\partial t} = \Delta(u^{1 - \frac{2}{n+p}}) + \nabla \cdot (\mathbf{x}u), \qquad (\mathbf{x}, t) \in \mathbf{R}^n \times \,]0, \infty[,$$

with the same initial data as v. Note that ρ is a stationary solution of the equation.

We begin by obtaining the rate of convergence of $u(t)$ towards ρ by estimating the decay of the relative entropy

$$(10) \qquad E(u|\rho) = E(u) - E(\rho),$$

corresponding to the Erdös-Renyi entropy

$$(11) \qquad E(u) = \int_{\mathbf{R}^n} \frac{1}{2}|\mathbf{x}|^2 u(\mathbf{x}) - \frac{n+p}{2} u(\mathbf{x})^{1 - \frac{2}{n+p}} \, d\mathbf{x}.$$

If $p \in \,]0, 2]$, one defines $E(u|\rho)$ directly by (27) which is equivalent to (10) except when the integrals comprising $E(\rho)$ diverge [26]. We follow Carrillo & Vazquez [11] in adopting an Bakry-Emery style entropy–entropy dissipation approach [4],

although the rates of convergence we obtain might also be derived using the implementation of Otto [32].

As Newman [31] and Ralston [36] recognized, the entropy $E(u)$ plays an important role in understanding nonlinear diffusion dynamics. Notice the self-similar profile ρ minimizes $E(u)$ uniquely among functions with given mass and finite variance:

$$\mathcal{M}_2 := \left\{ 0 \leq u \in L^1(\mathbf{R}^n) \mid \int_{\mathbf{R}^n} u(\mathbf{x}) \, d\mathbf{x} = 1 \text{ and } \int_{\mathbf{R}^n} |\mathbf{x}|^2 u(\mathbf{x}) < \infty \right\}.$$

Subsequently, Otto [32] discovered equation (9) can be formally viewed as the gradient flow of the entropy E on the set $\mathcal{M}_2$, endowed with the 2-Wasserstein metric. More precisely, he demonstrated that this metric arises as the arclength distance corresponding to the Hilbert norm

$$\|\Psi\|^2_{W^{1,2}_u} = \int_{\mathbf{R}^n} |\nabla \Psi(\mathbf{x})|^2 \, u(\mathbf{x}) \, d\mathbf{x},$$

on the tangent space $T_u \mathcal{M}_2 = W^{1,2}_u$ given by the Hilbert space completion of the smooth functions $\Psi : \mathbf{R}^n \longrightarrow \mathbf{R}$ with compact support. The evolution (9) gives steepest descent of E with respect to this metric. Local coordinates $\exp_u : T_u \mathcal{M}_2 \longrightarrow \mathcal{M}_2$ on the manifold are defined by the exponential map $v = \exp_u s\Psi$, where

(12) $$v(\mathbf{x} + s\nabla\Psi(\mathbf{x})) := u(\mathbf{x})/\det[I + sD^2\Psi(\mathbf{x})].$$

Taking two derivatives $d^2 E(v_s)/ds^2$ along the geodesic curve (12), Otto computed the Hessian of the entropy to be the following quadratic form on the tangent space $W^{1,2}_u$:

(13) $$\mathrm{Hess}_u E(\Psi, \Psi) = \int_{\mathbf{R}^n} |\nabla\Psi|^2 u + \int_{\mathbf{R}^n} \left(|\mathrm{Hess}\,\Psi|_2^2 - \frac{2}{n+p}(\Delta\Psi)^2 \right) u^{1-\frac{2}{n+p}},$$

where $|\mathrm{Hess}\,\Psi(\mathbf{x})|_2^2 := \sum_{i,j=1}^n (\partial^2\Psi/\partial x_i \partial x_j)^2$.

Our result on the rate of convergence follows from an interplay between the convergence in relative entropy and in the relative L^∞ sense (5). The convergence rate in relative entropy is controlled by the modulus of positivity for $\mathrm{Hess}_u E$ on the manifold $\mathcal{M}_2$ endowed with the Wasserstein metric. To be more precise let us denote a lower bound for the Hessian by

$$\Lambda(u) := \inf_{\|\Psi\|_{W^{1,2}_u}=1} \mathrm{Hess}_u E(\Psi, \Psi),$$

where the infimum encompasses all $\|\Psi\|_{W^{1,2}_u} = 1$ for which $\mathrm{Hess}_u E(\Psi, \Psi)$ takes an unambiguous value in $\mathbf{R} \cup \{\pm\infty\}$. If one can show that $\Lambda(u) \geq \Lambda_0 > 0$ uniformly in u (thus quantifying uniform convexity of E), exponential convergence of $u(\tau)$ towards the steady state at rate Λ_0 follows. This is indeed the case if both $p \geq n$ and $p > 2$, as was established by Otto [32]. An independent refinement of the

same strategy by different groups of researchers —Ambrosio, Gigli & Savaré [1], Carrillo, McCann & Villani [13], Sturm & von Renessee [38] and Sturm [37], shows the dynamics contract *globally* with respect to 2-Wasserstein distance.

In the general case $p > 0$, Denzler & McCann computed the precise value of $\Lambda(\rho)$ —among other things —where ρ is the steady state. Although $\Lambda(\rho)$ is positive, $\Lambda(u)$ can be arbitrarily negative when $p \in]0, n[$, so the dynamics no longer form a contraction globally. One might still hope for a positive lower bound on $\Lambda(u)$ locally, near the steady state ρ, but this is not the case either: the next section shows there exist functions $u \in \mathcal{M}^p$ arbitrarily close to ρ in Wasserstein metric (as well as in relative entropy, L^1 norm, and L^∞ norm) at which $\Lambda(u)$ is arbitrarily negative. In particular $\Lambda(u)$ is not a lower semicontinuous function of u in the Wasserstein distance. This constitutes our negative result. In a final section, we obtain our positive result by exploiting continuity of both $\Lambda(u)$ and the next higher eigenvalue when ρ is subject to a perturbation which is small in the relative L^∞ sense.

3. LOCAL SEMICONVEXITY OF THE ENTROPY LANDSCAPE FAILS FOR $|p| < n$

For $|p| \geq n$, Otto's gradient flow formulation of the dynamics takes place in a *convex* landscape, since the entropy $E(u)$ is *displacement convex* [28]. The present section explains why, for $|p| < n$, the analogous landscape fails even to be semi-convex in the displacement sense of McCann, i.e. to admit lower second derivative bounds locally. This is demonstrated by showing that $\Lambda(u)$ is not locally bounded below. We begin with an elementary observation:

Lemma 1. *If $|p| < n$ and $u \in \mathcal{M}_2$ then $\Lambda(u) < 1$ and $-\infty = \lim_{a \downarrow 0} \Lambda(au)$.*

Proof. Choosing the constant $c > 0$ to normalize the test function $\Psi(\mathbf{x}) = c|\mathbf{x}|^2/2$, we find $\mathrm{Hess}_u E(\Psi, \Psi) = 1 + (p - n)\frac{nc^2}{n+p} \int_{\mathbf{R}^n} u^{1-\frac{2}{n+p}}$, which is less than one if $|p| < n$. This proves the first claim: $\Lambda(u) < 1$. Now $\Lambda(au) - 1 \leq a^{-2/(n+p)}(\Lambda(u) - 1)$ decreases to minus infinity as $a > 0$ vanishes, showing the second claim follows from the first. $\square$

We demonstrate the local unboundedness of $\Lambda(u)$ using the construction depicted in Figure 1. The figure displays a function of a single variable, but it is easy to adapt this construction to higher dimensions by visualizing regions A and B as spherical shells around the ball C.

Proposition 2 (Lack of semiconvexity of the entropy landscape). *If $|p| < n$ and $u \in \mathcal{M}_2$ then both the 2-Wasserstein and L^1 limits $\liminf_{\tilde{u} \to u} \Lambda(\tilde{u}) = -\infty$ diverge.*

Proof. Let $\varepsilon > 0$ and $u \in \mathcal{M}_2$. We will construct $\tilde{u} \in \mathcal{M}_2$ arbitrarily close to u in 2-Wasserstein distance such that $\Lambda(\tilde{u})$ is arbitrarily negative. That is, such that $d_2(u, \tilde{u}) < \varepsilon$ and $\Lambda(\tilde{u}) < -1/\varepsilon$. To construct the example we perturb the function u in a small ball II centered at a point where u has a positive Lebesgue density. We
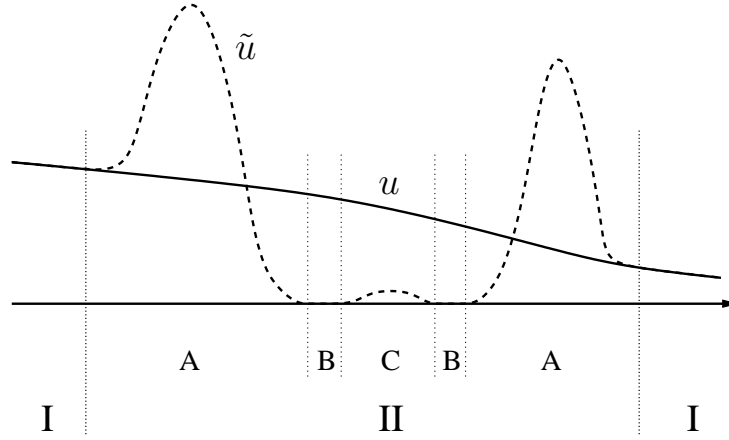
Figure 1: Illustrating the discontinuity of $\Lambda(u)$

will displace most of the mass from the inner part of the ball (letters B and C) to the spherical shell denoted by A. The Wasserstein distance between u and $\tilde{u}$ will be less then the square of the diameter of II times the mass of u in II and hence can be made less then ε either by choosing region II to be small or by choosing a region where u is small (as can be done if $u = \rho$ by perturbing ρ far from the origin). Notice that such a perturbation is also small in L^1, but to be small in L^∞ we would need to center the ball at a Lebesgue point where ε exceeds the density of u.

The proof of Lemma 1 yields a smooth function v, of arbitrarily small mass, supported in region C such that $\mathrm{Hess}_v E(\mathbf{x}^2, \mathbf{x}^2) < -\|\mathbf{x}^2\|^2_{W_v^{1,2}}/\varepsilon$. In particular $\Lambda(v) < -1/\varepsilon$. By choosing v so that its mass is less then the mass of u in region C we construct $\tilde{u}$ by letting it be equal to v in the region C, 0 in the region B, and by removing all of the excess mass to the region A.

Choose a test function $\Psi(\mathbf{x}) := c\chi(\mathbf{x})\mathbf{x}^2$, where χ is a cut-off function equal to 1 in C and equal to 0 outside of $B \cup C$, and c is the constant ensuring $\|\Psi\|_{W_v^{1,2}} = 1$. It follows that $\Psi \in W_{\tilde{u}}^{1,2}$ and that

$$\mathrm{Hess}_{\tilde{u}}(\Psi, \Psi) = \mathrm{Hess}_v(\Psi, \Psi) < -\frac{1}{\varepsilon} \quad \text{and} \quad \|\Psi\|_{W_{\tilde{u}}^{1,2}} = \|\Psi\|_{W_v^{1,2}} = 1.$$

Thus $\Lambda(\tilde{u}) < -\frac{1}{\varepsilon}$ as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

As mentioned in the proof, the same proposition holds for L^∞ limits, provided u has Lebesgue points of arbitrarily small density. We remark that it is not necessary to make $\tilde{u}$ equal to 0 in the region B. Making $\tilde{u}$ sufficiently small would have sufficed.

4. Nearly optimal second order asymptotics for $p \geq n \geq 2$

For $p \geq n$ on the other hand, $\Lambda(u)$ will prove continuous with respect to u near ρ in the relative L^∞ sense. Combining this observation with Vazquez' convergence of $u(t)$ towards ρ in relative L^∞ (5), yields sufficient control on the Hessian to obtain an improved rate of convergence for the relative entropy of centered solutions. Dilations show the rate of Theorem 3 to be optimal, apart from the parameter $\delta > 0$ which may be taken arbitrarily small. Unfortunately, this near optimal rate is obtained only in the relative entropy and L^1 senses; if we were able to obtain it in the relative L^∞ setting we could eliminate the parameter $\delta > 0$ by bootstrapping our way to optimality.

We shall work with initial data which belong to

$$\mathcal{M}_0^p := \left\{ u_0 \in \mathcal{M}_2 \,\Big|\, \int_{\mathbf{R}^n} \mathbf{x} u_0(\mathbf{x})\, d\mathbf{x} = 0 \text{ and } \limsup_{|\mathbf{x}| \to \infty} |\mathbf{x}|^{n+p} u_0(\mathbf{x}) < +\infty \right\},$$

or satisfy either of the weaker hypotheses (16)–(17). Let $\mathcal{M}^p$ denote the set of all translates of measures in $\mathcal{M}_0^p$. Note that for $p > 2$ all functions u in $\mathcal{M}^p$ have finite energy $E(u)$.

Theorem 3 (Near optimal decay rate for relative entropy of centered solutions)**.** *Fix $p \geq n \geq 2 \neq p$, and let $u(\mathbf{x}, t)$ solve (9) smoothly for $t > 0$ subject to an initial condition*

$$(14) \qquad 0 = \lim_{t \to 0} \| u(\cdot, t) - u_0(\cdot) \|_{L^1(\mathbf{R}^n)}.$$

with $u_0 \in \mathcal{M}_0^p$. For each $\delta > 0$ there exists $T > 0$ such that for all $t > T$,

$$(15) \qquad E(u(t)|\rho) \leq E(u_0|\rho)\, e^{-2(\alpha^{-1} - \delta)(t - T)}.$$

Remark 4. *The class of initial data for which the theorem holds can be extended. In particular Vazquez [39] has shown that if u_0 satisfies the following condition*

$$(16) \qquad |\mathbf{x}|^p \int_{B(\mathbf{x}, |\mathbf{x}|/2)} |u_0(\tilde{\mathbf{x}})|\, d\tilde{\mathbf{x}} < const$$

for all $\mathbf{x}$ large enough then $u(\,\cdot\,, t) \in \mathcal{M}^p$ for all positive times. Thus the conclusion of the theorem holds for centered initial data satisfying (16).

Remark 5. *Since (16) is satisfied for all $u \in \mathcal{M}^p$ we conclude that if $u_0 \in \mathcal{M}^p$ then for all times $u(t) \in \mathcal{M}^p$. Thus $\mathcal{M}^p$ is closed under the evolution.*

Remark 6. *We remark that any function with p-th moments satisfies condition (16). That is, (16) holds if*

$$(17) \qquad \int_{\mathbf{R}^n} |\mathbf{x}|^p u_0(\mathbf{x}) d\mathbf{x} < \infty.$$

Proof. (of Remark 6). To show the contrapositive, assume $u_0 \geq 0$ fails to satisfy (16). Then there exists a sequence $\mathbf{x}_k \to \infty$ such that $|\mathbf{x}_{k+1}| > 2|\mathbf{x}_k|$ and $|\mathbf{x}_k|^p \int_{B(\mathbf{x}_k, |\mathbf{x}_k|/2)} u_0(\mathbf{x}) \, d\mathbf{x} > 1$. But since the balls $B(\mathbf{x}_k, |\mathbf{x}_k|/2)$ are disjoint

$$
\begin{aligned}
\int_{\mathbf{R}^n} |\mathbf{x}|^p u_0(\mathbf{x}) \, d\mathbf{x} &\geq \sum_{k=1}^\infty \int_{B(\mathbf{x}_k, \frac{|\mathbf{x}_k|}{2})} |\mathbf{x}|^p u_0(\mathbf{x}) \, d\mathbf{x} \\
&\geq \sum_{k=1}^\infty \frac{|\mathbf{x}_k|^p}{2^p} \int_{B(\mathbf{x}_k, \frac{|\mathbf{x}_k|}{2})} u_0(\mathbf{x}) \, d\mathbf{x} \\
&= +\infty
\end{aligned}
$$

which precludes (17). $\qquad\square$

Proof. Let us first prove Theorem 3 under an additional assumption on $u_0 \in \mathcal{M}_0^p$:

Case 1°. Assume that for some number $a > 0$

$$
\lim_{|\mathbf{x}| \to \infty} u_0(\mathbf{x}) |\mathbf{x}|^{n+p} = a.
$$

Let $\varepsilon(t) := \|(u(t) - \rho)/\rho\|_{L^\infty(\mathbf{R}^n)}$ and let $h(t) := E(u(t)|\rho)$. Vazquez [39] showed (5) for $v_0 = u_0 \in \mathcal{M}_0^p$. But note that $|u(t) - \rho|/\rho = |v(\tau) - V(\tau)|/V(\tau)$ where t and τ are related by (7). Therefore $\varepsilon(t)$ converges to zero as t goes to infinity. In particular there exists T such that $\varepsilon(t) < \min\left\{ \frac{n+p}{2(n+p-1)} \alpha\delta, 1 \right\}$ for all $t > T$.

Under the additional condition of Case 1°, Lee & Vazquez [27, Lemma 6.1] and Carrillo & Vazquez [11, Lemma 5.1] have shown that it is easy to obtain the decay of derivatives of $u(t)$. In particular, for any multi-index β, locally uniformly in time around $t > 0$,

$$
(18) \qquad D^\beta u(\mathbf{x}, t) = O(|\mathbf{x}|^{-n-p-|\beta|}) \quad \text{as } |\mathbf{x}| \to \infty;
$$

see Kim & McCann [22, Corollary 9]. As Carrillo & Vazquez also showed, this is sufficient for h to be a C^2 function, and they computed

$$
\begin{aligned}
h'(t) &= -\int_{\mathbf{R}^n} u \left| \mathbf{x} - \frac{n+p-2}{2} \nabla u^{-\frac{2}{n+p}} \right|^2 d\mathbf{x} \\
&= -\int_{\mathbf{R}^n} u |\nabla \Psi|^2 \, d\mathbf{x}
\end{aligned}
$$

where $\Psi(\mathbf{x}) = \frac{|\mathbf{x}|^2}{2} - (\frac{n+p}{2} - 1) u^{-\frac{2}{n+p}}$, and

$$
(19) \quad h''(t) = 2 \int_{\mathbf{R}^n} u |\nabla \Psi|^2 \, d\mathbf{x} + 2 \int_{\mathbf{R}^n} u^m \left(|\operatorname{Hess} \Psi|_2^2 - \frac{2}{n+p} (\Delta \Psi)^2 \right) d\mathbf{x}.
$$

Thus $h'(t) = -\|\Psi\|_{W_u^{1,2}}^2$ and $h''(t) = 2 \operatorname{Hess}_u E(\Psi, \Psi)$.

For $p \geq n$, the Cauchy-Schwartz inequality $Tr[A^+B] \leq \sqrt{Tr[A^+A]Tr[B^+B]}$ with $A = I$ and $B = D^2\Psi$ shows the second integrand in (19) is non-negative. This is Otto's conclusion $\mathrm{Hess}_u E(\Psi, \Psi) \geq \|\Psi\|^2_{W^{1,2}_u}$ [32].

Moreover, Denzler & McCann [17] [18] showed that for Ψ orthogonal in $W^{1,2}_\rho$ to the functions $X = \{x_1, x_2, \ldots, x_n\}$ which generate translations, this inequality can be improved by a factor $\alpha^{-1} = 2p/(n+p)$. In particular,

$$\text{(20)} \qquad \mathrm{Hess}_\rho E(\tilde{\Psi}, \tilde{\Psi}) \geq \frac{1}{\alpha}\|\tilde{\Psi}\|^2_{W^{1,2}_\rho}$$

for $\tilde{\Psi} \in W^{1,2}_\rho$ such that $\langle \tilde{\Psi}, x_i \rangle_\rho := \int_{\mathbf{R}^n} \rho(\mathbf{x})\nabla\tilde{\Psi} \cdot \hat{\mathbf{e}}_i \, d\mathbf{x} = 0$ for all $i = 1, \ldots, n$, where $\hat{\mathbf{e}}_i$ are coordinate unit vectors.

We plan use this estimate for vectors $\Psi \in W^{1,2}_u$ tangent to the trajectory $u(t)$, namely $\Psi(\mathbf{x}) = |\mathbf{x}|^2/2 + \frac{m}{m-1}u^{m-1}$, when u is near ρ in the relative L^∞ sense. Note that $\varepsilon(t) < 1$ implies $W^{1,2}_u = W^{1,2}_\rho$. The decay of derivatives (18) is suffi cient to guarantee that the evolution preserves the center of mass for positive times. Additionally, as we will show for more general initial data considered in Case 2°, the center of mass is a continuous function of time at $t = 0$. Thus using the center of mass condition (6), and the decay (18) of u to integrate by parts yields

$$\text{(21)} \qquad \langle \Psi, x_i \rangle_u = \int_{\mathbf{R}^n} (x_i u + \frac{\partial u^m}{\partial x_i}) \, d\mathbf{x} = 0.$$

However, $\langle \Psi, x_i \rangle_\rho \neq 0$ generally. We must therefore project Ψ onto the orthogonal complement of the set of functions $X := \{x_1, x_2, \ldots x_n\}$ in $W^{1,2}_\rho$. If u is suffi ciently close to ρ, neither the norm in $W^{1,2}_\rho$ nor the Hessian will be much changed by this projection, since the set X consists precisely of the lowest eigenfuctions for both Hessians, Hess_u and Hess_ρ. We estimate these changes precisely below, denoting

$$\begin{aligned}
\tilde{\Psi} \quad &:= \quad \Pi_{X^\perp}\Psi \\
&= \quad \Psi - \sum_{i=1}^n \frac{\langle \Psi, x_i \rangle_\rho}{\|x_i\|^2_{W^{1,2}_\rho}} x_i \\
&= \quad \Psi - \sum_{i=1}^n \frac{\int_{\mathbf{R}^n} \rho\nabla\Psi \cdot \hat{\mathbf{e}}_i \, d\mathbf{x}}{\int_{\mathbf{R}^n} \rho \, d\mathbf{x}} x_i.
\end{aligned}$$

For $i = 1, \ldots, n$, $|u - \rho| \leq \varepsilon(t)\rho$ combines with (21) to imply

$$\begin{aligned}
|\langle \Psi, x_i \rangle_\rho|^2 \quad &= \quad \left| \int_{\mathbf{R}^n} (\rho - u)\nabla\Psi \cdot \hat{\mathbf{e}}_i \, d\mathbf{x} \right|^2 \\
&\leq \quad \varepsilon(t)^2 \left(\int_{\mathbf{R}^n} \rho \left| \frac{\partial\Psi}{\partial x_i} \right| d\mathbf{x} \right)^2 \\
&\leq \quad \varepsilon(t)^2 \int_{\mathbf{R}^n} \rho \left| \frac{\partial\Psi}{\partial x_i} \right|^2 d\mathbf{x},
\end{aligned}$$

by Jensen's inequality. Thus

$$
\begin{aligned}
\|\tilde{\Psi}\|^2_{W^{1,2}_\rho} &= \|\Psi\|^2_{W^{1,2}_\rho} - \sum_{i=1}^{n} |\langle \Psi, x_i\rangle_\rho|^2 \\
&\geq (1 - \varepsilon(t)^2)\|\Psi\|^2_{W^{1,2}_\rho} \\
&\geq (1 - \varepsilon(t))\|\Psi\|^2_{W^{1,2}_u}.
\end{aligned}
$$

(22)

Now

$$
\begin{aligned}
-\frac{h''(t)}{2h'(t)} &= 1 + \frac{1}{\|\Psi\|^2_{W^{1,2}_u}} \int_{\mathbf{R}^n} u^m \left((m-1)(\Delta\Psi)^2 + |\operatorname{Hess}\Psi|^2_2\right) \\
&\geq 1 + \frac{(1-\varepsilon(t))^m(1-\varepsilon(t))}{\|\tilde{\Psi}\|^2_{W^{1,2}_\rho}} \int_{\mathbf{R}^n} \rho^m \left((m-1)(\Delta\Psi)^2 + |\operatorname{Hess}\Psi|^2_2\right) \\
&\geq 1 + \frac{(1-m\varepsilon(t))(1-\varepsilon(t))}{\|\tilde{\Psi}\|^2_{W^{1,2}_\rho}} \int_{\mathbf{R}^n} \rho^m \left((m-1)(\Delta\tilde{\Psi})^2 + |\operatorname{Hess}\tilde{\Psi}|^2_2\right) \\
&\geq [1 - (m+1)\varepsilon(t)]\alpha^{-1},
\end{aligned}
$$

where we have used (22) to obtain the first inequality and (20) to obtain the last. For all $t > T$, this means

$$
\begin{aligned}
h''(t) &\geq -2\alpha^{-1}\left[1 - \tfrac{n+p-1}{n+p}2\varepsilon(t)\right]h'(t) \\
&\geq -2[\alpha^{-1} - \delta]h'(t),
\end{aligned}
$$

(23)

since $\frac{m+1}{2} = \frac{n+p-1}{n+p}$. It was also proven, e.g. by Carrillo & Vazquez, that $h(t) \to 0$ as $t \to \infty$. Hence, since h' is nonpositive, $h'(t) \to 0$ as $t \to \infty$. Integrating in (23) from t to ∞ yields

$$
-h'(t) \geq 2(\alpha^{-1} - \delta)h(t) \quad \text{for all } t > T.
$$

The conclusion of the theorem now follows by dividing by $h(t)$, integrating from T to t, and noting that $E(u(T)|\rho) \leq E(u_0|\rho)$,

Case 2°. Now let us consider the general case $u_0 \in \mathcal{M}^p_0$. Then

$$
u_0(\mathbf{x}) = O(|\mathbf{x}|^{-n-p})
$$

The strategy we employ is to approximate u_0 by functions whose tails satisfy the condition of Case 1°. The obstacle to carrying out this approximation is that the results of Case 1° depend on the initial data through the quantity T produced by Vazquez' relative-L^∞ convergence result (5). The remainder of the proof is concerned with showing that T can be chosen uniformly with respect to the approximations we need.

Recall that Vazquez' argument is partly based a comparison of the solution of interest with scaling solutions of different total mass [39]. For that reason we will write ρ_C and $V(\cdot, \cdot, C)$ when we want to specify what constant C was used in

defi ning (2). We furthermore introduce the function $U(\mathbf{x}, t, C, t_0)$ which solves (9) with initial condition $V(\mathbf{x}, \alpha t_0, C)$. This solution can be written explicitly:

(24)

$$
\begin{aligned}
U(\mathbf{x}, t, C, t_0) &= e^{nt} V(e^t \mathbf{x}, \alpha(e^{t/\alpha} - 1 + t_0), C) \\
&= (1 + (t_0 - 1)e^{-t/\alpha})^{-n\alpha} \rho_C \left((1 + (t_0 - 1)e^{-t\alpha})^{-\alpha} \mathbf{x} \right) \\
&= \eta^{\frac{n+p}{2}} \left(C\eta^{2\alpha} + \frac{|\mathbf{x}|^2}{n + p - 2} \right)^{-\frac{n+p}{2}} \\
&= \left(\frac{1}{n + p - 2} \right)^{-\frac{n+p}{2}} \eta^{\frac{n+p}{2}} |\mathbf{x}|^{-(n+p)} + o(|\mathbf{x}|^{-(n+p)}) \quad \text{as } |\mathbf{x}| \to \infty.
\end{aligned}
$$

Here $\eta := 1 + (t_0 - 1)e^{-t/\alpha}$ lies between 1 and t_0 for all $t > 0$, from which one sees the smallness of the error term is $o(|\mathbf{x}|^{-(n+p)})$ uniformly in time. Observe that the limit

$$
\begin{aligned}
A(t, t_0) &:= \lim_{|\mathbf{x}| \to \infty} U(\mathbf{x}, t, C, t_0) |\mathbf{x}|^{n+p} \\
&= [(n + p - 2)\eta]^{\frac{n+p}{2}}
\end{aligned}
$$

is fi nite and independent of C for all nonnegative t and t_0, meaning tail thickness depends only on the age of the solution, and not on its mass. Additionally note that $A(0, 0) = 0$ and $\lim_{t_0 \to \infty} A(t, t_0) = \infty$.

Upper bound of tails by scaling solutions. Our fi rst step is to show that the tails of the solution u are for all times bounded by the tails of a similarity solution with negative C. Let

$$
\overline{a}(t) := \limsup_{|\mathbf{x}| \to \infty} u(\mathbf{x}, t) |\mathbf{x}|^{n+p}.
$$

Note that $\overline{a}(t)$ is fi nite for all t by Remark 5. Note that $U(\mathbf{x}, t, -1, t_0)$ is a solution of equation (9) on an exterior domain with moving boundary. In particular at time t the domain is the exterior of ball of radius $\sqrt{n + p - 2}\, \eta^\alpha$. The value of U grows to ∞ as $\mathbf{x}$ approaches the boundary of the domain. Observe that for t_0 large enough $U(\mathbf{x}, 0, -1, t_0) > u_0(\mathbf{x})$ for all $|\mathbf{x}| > \sqrt{n + p - 2}\, t_0^\alpha$. Therefore by comparison principle on the exterior domain $U(\,\cdot\,, t, -1, t_0) > u(\,\cdot\,, t)$. This implies that the function $\overline{a}(t)$ is bounded on fi nite time intervals.

This uniform control on the tails, combined with (14), implies that the center of mass is a continuous function of time.

Approximations with uniform tails. Let $\phi \in C^\infty([0, \infty))$ be nondecreasing cut-off function such that $\phi = 0$ on $[0, 1]$ and $\phi = 1$ on $[2, \infty)$. Let

$$
u_{0,k}(\mathbf{x}) := \left(1 - \phi\left(\frac{\mathbf{x} - \mathbf{x}_k}{k} \right) \right) u_0(\mathbf{x} - \mathbf{x}_k) + \gamma_k\, \phi\left(\frac{\mathbf{x}}{k} \right) |\mathbf{x}|^{-(n+p)}
$$

for each k, where $\gamma_k > 0$ and $\mathbf{x}_k \in \mathbf{R}^n$ are chosen to ensure that $u_{0,k}$ has the same unit mass and zero center of mass as $u_0 \in \mathcal{M}_0^p$. From the definition of $\overline{a}$ follows that $\limsup_{k\to\infty} \gamma_k \leq \overline{a}(0)$. Note that $\mathbf{x}_k \to 0$ as $k \to \infty$. Let $u_k(\mathbf{x}, t)$ be the solution of (9) with initial data $u_{k,0}$. Recall that the solutions u_k preserve the center of mass. Since the evolution is an L^1 contraction and both $u_k(t)$ and $u(t)$ decay as $|\mathbf{x}|^{-n-p}$ for $|\mathbf{x}|$ large it follows that centers of mass of $u_k(t)$ converge towards the center of mass of $u(t)$. Therefore the center of mass of $u(t)$ is 0 for all times.

Bounds by scaling solutions. We now show that, after a short time, our solution $u(\mathbf{x}, t)$ and all its approximants will be bounded below by a scaling solution which is sufficiently young and sufficiently light, and bounded above by a scaling solution which is sufficiently old and sufficiently massive. Let

$$\underline{a} := \liminf_{|\mathbf{x}|\to\infty} u_0(\mathbf{x})|\mathbf{x}|^{n+p}$$

We can assume without the loss of generality that $\underline{a} > 0$ For if u_0 does not have this property, all it takes is to wait some time $\tilde{t}$ and consider the problem with initial data $u(\,\cdot\,, \tilde{t})$. To verify that consider $u(\mathbf{x}, t)$ and $\tilde{u}(\mathbf{x}, t) := U(\mathbf{x}, t - 1, C, 0)$ as solutions of (9) on $(\mathbf{R}^n \backslash \mathbf{B}_1(\mathbf{0})) \times [1, \infty[$. Since u is positive for $t > 0$ while $\tilde{u}(\,\cdot\,, 1) = 0$ on the complement of the origin, $u(\,\cdot\,, t) > \tilde{u}(\,\cdot\,, t)$ on $\partial\mathbf{B}_1(\mathbf{0})$ for all $t \in [1, \tilde{t}]$ for some $\tilde{t} > 1$. Therefore by comparison principle $u \geq \tilde{u}$ on $(\mathbf{R}^n \backslash \mathbf{B}_1(\mathbf{0})) \times [1, \tilde{t}]$. Since $\tilde{u}(\,\cdot\,, \tilde{t})$ has the desired lower bound on the decay of its tails so does $u(\,\cdot\,, \tilde{t})$. Furthermore, since $\|\tilde{u}(\,\cdot\,, \tilde{t})\|_{L^\infty(\mathbf{B}_1(\mathbf{0}))} \to 0$ as $C \to \infty$ the inequality between $u(\,\cdot\,, \tilde{t})$ and the scaling solution can be extended to all of $\mathbf{R}^n$. In particular we can also assume that $u_0(\,\cdot\,) > U(\,\cdot\,, 0, \underline{C}, \underline{t})$ on $\mathbf{R}^n$ for some $\underline{C} > 0$ and $\underline{t}$.

Similarly, we can obtain an upper bound on $\mathbf{R}^n$. Since $a(1)$ is finite and $u(\,\cdot\,, 1)$ is a bounded function, by choosing t_0 large and $C = \eta^{-2\alpha}$ one has $U(\,\cdot\,, 1, C, t_0) > u(\,\cdot\,, 1)$ on $\mathbf{R}^n$. Therefore, by shifting time by one, we can also assume that for some $\overline{t}$ and $\overline{C}$, $u_0(\,\cdot\,) < U(\,\cdot\,, 0, \overline{C}, \overline{t})$.

By the definition of $\underline{a}$ and $\overline{a}$:

$$0 < \underline{a} \leq \liminf_{k\to\infty} \gamma_k \leq \limsup_{k\to\infty} \gamma_k \leq \overline{a}(0) < \infty.$$

Using this uniformity of tails of approximating solutions we can assume (adjusting the choices of $\underline{C}$, $\underline{t}$, $\overline{C}$, and $\overline{t}$ slightly if necessary) that for some k_0 and all $k > k_0$,

$$U(\,\cdot\,, 0, \underline{C}, \underline{t}) \leq u_{k,0}(\,\cdot\,) \leq U(\,\cdot\,, 0, \overline{C}, \overline{t}).$$

By comparison principle it follows that

(25) $$U(\,\cdot\,, t, \underline{C}, \underline{t}) \leq u_k(\,\cdot\,, t) \leq U(\,\cdot\,, t, \overline{C}, \overline{t}).$$

for all $t \geq 0$ and all $k > k_0$.

We claim that there exists $T > 0$ such that for all $k > k_0$, $\|u_k(\,\cdot\,, t)/\rho - 1\|_{L^\infty} < \delta_1 := \min\left\{\frac{n+p}{2(n+p-1)}\alpha\delta, 1\right\}$ for each $t > T$. In other words we claim that the

relative L^∞ convergence (5) obtained by Vazquez occurs uniformly in k. Following Vazquez' proof of convergence we split $\mathbf{R}^n$ in two parts.

Tail analysis. From the expression (24) for the tail of U, it follows that there exist $T_1 > 0$ and $R > 0$ such that

$$1 - \delta_1 < \frac{U(\mathbf{x}, t, \underline{C}, \underline{t})}{\rho(\mathbf{x})} < \frac{U(\mathbf{x}, t, \overline{C}, \overline{t})}{\rho(\mathbf{x})} < 1 + \delta_1$$

for all $t \geq T_1$ and all $|\mathbf{x}| \geq R$. In this region, the desired estimates for $u_k(\mathbf{x}, t)$ now follow from (25).

Analysis on $\overline{\mathbf{B}}_R(\mathbf{0})$. It is enough to show that there exists T_2 such that for all $t \geq T_2$ and all $\mathbf{x} \in \overline{\mathbf{B}}_R(\mathbf{0})$, $|u(\mathbf{x}, t) - \rho(\mathbf{x})| < \delta_2 := \delta_1 \min\{\rho(\mathbf{z}) \mid \mathbf{z} \in \overline{\mathbf{B}}_R(\mathbf{0})\}$.

Let $\lambda := \inf\{U(\mathbf{x}, t, \underline{C}, \underline{t}) \mid |\mathbf{x}| \leq 2R, \ t \geq 0\}$ and $\Lambda := \sup\{U(\mathbf{x}, t, \overline{C}, \overline{t}) \mid |\mathbf{x}| \leq 2R, \ t \geq 0\}$. From (24) one can deduce that $0 < \lambda \leq \Lambda < \infty$. Therefore the equation (9) is uniformly parabolic along the trajectories u_k (for $k > k_0$) and u. Classical regularity results [25] yield the existence of some Hölder exponent $\theta > 0$ such that $u_k(\,\cdot\,, t)$ is in $C^\theta(\mathbf{B}_R(\mathbf{0}))$ for all $k > k_0$ and $t \geq 1$ with an upper bound on Hölder norm independent of k and t.

To derive a contradiction, assume no such T_2 exists. Then there exist sequences $t_j \to \infty$ as $j \to \infty$ and $k_j > k_0$ such that $\|u_{k_j}(\,\cdot\,, t_j) - \rho(\,\cdot\,)\|_{L^\infty(\mathbf{B}_R(\mathbf{0}))} \geq \delta_2$. But since the Hölder norm of functions $u_{k_j}(\,\cdot\,, t_j)$ is bounded from above they are equicontinuous, and hence by Arzelà-Ascoli lemma there exists an L^∞-convergent subsequence. For notational simplicity we assume that the subsequence is the whole sequence. So $u_{k_j}(\,\cdot\,, t_j) \to \tilde{\rho}$ in $L^\infty(\mathbf{B}_R(\mathbf{0}))$ as $j \to \infty$ for some function $\tilde{\rho}$. Note that this also implies the convergence in $L^1(\mathbf{B}_R(\mathbf{0}))$.

Now observe from (25) that $E(u_{k,0})$ is uniformly bounded. Looking into Carrillo and Vazquez' derivation of an L^1 rate of convergence [11, Sec. 4], this can be seen to imply uniformity of the limit

$$\lim_{t \to \infty} \sup_{k > k_0} \|u_k(\,\cdot\,, t) - \rho(\,\cdot\,)\|_{L^1} = 0$$

with respect to k. Thus $u_{k_j}(\,\cdot\,, t_j) \to \rho$ in L^1. Therefore on $\mathbf{B}_R(\mathbf{0})$, $\tilde{\rho} = \rho$ which contradicts the hypothesis that u_{k_j} remains distance δ_2 from ρ somewhere in this ball. The only conclusion can be that T_2 exists after all.

Let $T = \max\{T_1, T_2\}$. Since $u_{k,0}$ satisfy the condition of Case 1°, and T satisfies the desired condition used in the proof of Case 1°, we conclude that the decay rate (15) holds for u_k for all $k \geq k_0$. Since $E(u_{k,0}|\rho) \to E(u_0|\rho)$ as $k \to \infty$ and since $E(u(t)|\rho) \leq \liminf_{k \to \infty} E(u_k(t)|\rho)$ the theorem is established. $\qquad\square$

It is well-known [11] [32] that the relative entropy controls the distance in the L^1 norm by a Csiszár–Kullback [15] [24] type inequality

$$(26) \qquad \|u - \rho\|_{L^1(\mathbf{R}^n)} \leq \left(\frac{8}{m} \int_{\mathbf{R}^n} \rho^{2-m}\right)^{1/2} \sqrt{E(u|\rho)}.$$

In the present setting, the relative L^∞ convergence (5) of u to ρ yields similar inequalities for other L^q-spaces by combining Hölder's inequality with a second order Taylor expansion of the (convex) power law defining the entropy.

Lemma 7 (Csiszár-Kullback type interpolation inequality via Taylor expansion)**.** *Fix $p > 2$ and $u \in L^1(\mathbf{R}^n)$. Suppose $|(u(\mathbf{x}) - \rho(\mathbf{x}))/\rho(\mathbf{x})| < \varepsilon < 1$ for all $\mathbf{x} \in \mathbf{R}^n$. Then for all $1 < q < 1 + \frac{2(2-m)}{n(1-m)} = 2 + \frac{p+2}{n}$,*

$$\int_{\mathbf{R}^n} |u - \rho|^{2/q} \leq \left[\frac{2(1+\varepsilon)^{2-m}}{m} E(u|\rho) \right]^{1/q} \left[\int_{\mathbf{R}^n} \rho^{(2-m)/(q-1)} \right]^{(q-1)/q} < +\infty.$$

Proof. Let $\phi(r) = \frac{1}{m-1} r^m$. Note that

$$(27) \qquad E(u|\rho) := \int_{\mathbf{R}^n} \phi(u) - \phi(\rho) - \phi'(\rho)(u - \rho),$$

and all integrals converge due to the fact that $p > 2$ and $|u - \rho| < \rho$. Therefore

$$E(u|\rho) = \frac{1}{2} \int_{\mathbf{R}^n} \phi''(\rho + \theta(u - \rho)) (u - \rho)^2$$

for some function $0 \leq \theta \leq 1$. Using that $\phi''(r) = m\rho^{m-2}$ and that $-\varepsilon\rho < \theta(u-\rho) < \varepsilon\rho$ one obtains that

$$(28) \qquad E(u|\rho) \geq \frac{m}{2}(1+\varepsilon)^{m-2} \int_{\mathbf{R}^n} \rho^{m-2}(u - \rho)^2$$

Hölder's inequality yields

$$(29) \qquad \int_{\mathbf{R}^n} |u - \rho|^{2/q} \leq \left(\int_{\mathbf{R}^n} (u - \rho)^2 \rho^{m-2} \right)^{1/q} \left(\int_{\mathbf{R}^n} \rho^{(2-m)/(q-1)} \right)^{(q-1)/q}.$$

We now note that $1 < q < 1 + \frac{2(2-m)}{n(1-m)}$ implies $\frac{2-m}{q-1} > \frac{n(1-m)}{2}$. On the other hand it is easy to check that $\rho \in L^r$ precisely for $r > \frac{n}{n+p} = \frac{n(1-m)}{2}$. Combining the inequalities (28) and (29) now finishes the proof. $\square$

The Csiszár–Kullback type interpolation inequalities of Lemma 7 are now used to convert decay of the relative entropy from Theorem 3 into a nearly optimal L^r convergence rate.

Corollary 8 (L^r convergence rates in similarity variables)**.** *Fix $p \geq n \geq 2$ with $p > 2$ and $u_0 \in \mathcal{M}_0^p$. Let $u(\mathbf{x}, t)$ be the solution of (9) with initial data u_0. For any $r \in \left] \frac{2n}{2n+p+2}, 2 \right[$ and $\delta \in]0, \frac{r}{\alpha}[$ there exists t_0 such that for all $t > t_0$*

$$(30) \qquad \int_{\mathbf{R}^n} |u(\mathbf{x}, t) - \rho(\mathbf{x})|^r \, d\mathbf{x} < e^{-(\alpha^{-1}r - \delta)(t - t_0)}.$$

Proof. Recall $m = 1 - \frac{2}{n+p}$ and note $q := \frac{2}{r} \in \,]1, 2 + \frac{p+2}{n}[$. Take T large enough (depending on u_0) that Theorem 3 yields $h(t) := E(u(t)|\rho) \leq h(0)e^{-2(\alpha^{-1} - \delta/r)(t-T)}$. Convergence in relative L^∞ norm allows us to assume $\varepsilon(t) \leq 1/2$ for all $t > T$. Since $h(t) \to 0$, taking $t_0 > T$ so large that $\frac{2}{m}h(0)\, 2^{2-m}\|\rho^{2-m}\|_{L^{\frac{1}{q-1}}(\mathbf{R}^n)} < e^{2(\alpha^{-1} - \delta/r)(t_0 - T)}$ implies (30) via Lemma 7. $\qquad\square$

For comparison with the standard Csiszár-Kullback result, notice that $r = 1$ is an allowed value in (30). Transforming this estimate from similarity variables back to original quantities using (7) yields the following rate of convergence, in terms of v. Apart from $\delta > 0$, which can be chosen arbitrarily small, these decay results achieve the optimal power of τ, as example (3)–(4) reveals.

Corollary 9 (L^1 convergence rate in original variables). *Fix $p \geq n \geq 2$ with $p > 2$ and $v_0 \in \mathcal{M}_0^p$. Then for any $\delta > 0$ there exists $\tau_0 > 0$ such that for all $\tau > \tau_0$*

$$(31) \qquad \|v(\tau) - V(\tau + \alpha)\|_{L^1(\mathbf{R}^n)} \leq \left(\frac{\alpha + \tau_0}{\alpha + \tau} \right)^{1-\alpha\delta}.$$

Proof. Set $\tau_0 := \alpha(e^{t_0/\alpha} - 1)$. Then (30) becomes equivalent to (31) via the changes of variables (7)–(8) and definition (2). $\qquad\square$

Corollary 10 (Near optimal bound on second order L^1-asymptotics). *Fix $p \geq n \geq 2 \neq p$, centers $\mathbf{z}_0, \tilde{\mathbf{z}}_0 \in \mathbf{R}^n$, and a as in (4). Let v and $\tilde{v}$ solve the evolution (1) with centered initial data $v_0, \tilde{v}_0 \in \mathcal{M}_0^p$. Then for any $\delta > 0$,*

$$(32) \quad \|v(\cdot - \mathbf{z}_0, \tau) - \tilde{v}(\cdot - \mathbf{z}_0, \tau)\|_{L^1(\mathbf{R}^n)} = \frac{a}{\tau^\alpha}|\mathbf{z}_0 - \tilde{\mathbf{z}}_0| + O(\frac{1}{\tau^{1-\delta}}) \text{ as } \tau \to \infty.$$

Proof. Comparing v with V and $\tilde{v}$ with V separately, (32) follows from (3) and (31) via the triangle inequality. $\qquad\square$

REFERENCES

[1] L.A. Ambrosio, N. Gigli, and G. Savar´e. *Gradient flows in metric spaces and in the space of probability measures*. Lecture Notes in Mathematics ETH Z¨urich. Birkh¨auser Verlag, Basel, 2005.

[2] S.B. Angenent. Large time asymptotics for the porous medium equation. In *Nonlinear diffusion equations and their equilibrium states I*, volume 12 of *Math. Sci. Res. Inst. Publ.*, pages 21–34. Springer, New York, 1988.

[3] D.G. Aronson and P. B´enilan. R´egularit´e des solutions de l' ´equation des milieux poreux dans $\mathbf{R}^N$. *C.R. Acad. Sci. Paris Sér. I Math.*, 288:103–105, 1979.

[4] D. Bakry and M. Emery. Diffusions hypercontractives. In *Sém. Proba. XIX*, number 1123 in Lecture Notes in Math., pages 177–206. Springer, New York, 1985.

[5] G.I. Barenblatt. On some unsteady motions of a liquid or gas in a porous medium. *Akad. Nauk. SSSR. Prikl. Mat. Mekh.*, 16:67–78, 1952.

[6] G.I. Barenblatt, V.M. Entov and V.M. Ryzhik. *Flow of Fluids Through Natural Rocks*. Kluwer Academic Publishers, Dordrecht, 1990.

[7] P. B´enilan and M. G. Crandall. The continuous dependence on φ of solutions of $u_t - \Delta\varphi(u) = 0$. *Indiana Univ. Math. J.*, 30:161–177, 1981.

[8] J. Boussinesq. Recherches th´eoriques sur l' ´ecoulement des nappes d'eau infiltr´es dans le sol et sur le d´ebit de sources. *Comptes Rendus Acad. Sci. / J. Math. Pures Appl.*, 10:5–78, 1903/04.

[9] J.M. Carlson, J.T. Chayes, E.R. Grannan, G.H. Swindle. Self-organized criticality and singular diffusion. *Phys. Rev. Lett.*, 65:2547–2550, 1990.

[10] J.A. Carrillo and G. Toscani. Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, 49:113–141, 2000.

[11] J.A. Carrillo and J.L. V´azquez. Fine asymptotics for fast diffusion equations. *Comm. Partial Differential Equations*, 28:1023–1056, 2003.

[12] J.A. Carrillo, C. Lederman, P.A. Markowich and G. Toscani. Poincar´e inequalities for linearizations of very fast diffusion equations. *Nonlinearity*, 15:565–580, 2002.

[13] J.A. Carrillo, R.J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. To appear in *Arch. Rational Mech. Anal.*

[14] J.T. Chayes, S.J. Osher, J.V. Ralston. On singular diffusion equations with applications to self-organized criticality. *Comm. Pure Appl. Math.*, 46:1363–1377, 1993.

[15] I. Csisz´ar. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[16] B.E.J. Dahlberg and C. Kenig. Nonnegative solutions of fast diffusion equations. *Rev. Mat. Iberoamericana*, 4:11–29, 1988.

[17] J. Denzler and R.J. McCann. Phase transitions and symmetry breaking in singular diffusion. *Proc. Natl. Acad. Sci. USA*, 100:6922–6925, 2003.

[18] J. Denzler and R.J. McCann. Fast diffusion to self-similarity: complete spectrum, long time asymptotics, and numerology. *Arch. Rational Mech. Anal.*, 175:301–342, 2005.

[19] J. Dolbeault and M. del Pino. Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. *J. Math. Pures Appl.*, 81:847–875, 2002.

[20] A. Friedman and S. Kamin. The asymptotic behaviour of a gas in an n-dimensional porous medium. *Trans. Amer. Math. Soc.*, 262:551–563, 1980.

[21] M.A. Herrero and M. Pierre. The Cauchy problem for $u_t = \Delta u^m$ when $0 < m < 1$. *Trans. Amer. Math. Soc.*, 291:145–158, 1985.

[22] Y.J. Kim and R.J. McCann. Potential theory and optimal convergence rates in fast nonlinear diffusion. *Preprint.*

[23] Y.J. Kim and R.J. McCann. Sharp decay rates for the fastest conservative diffusions. To appear in *C.R. Acad. Sci. Paris S´er. I Math.*

[24] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*, 4:126–127, 1967.

[25] O.A. Lady\v{z}enskaja, V.A. Solonnikov, and N.N. Ural'ceva. *Linear and quasilinear equations of parabolic type*. American Mathematical Society, Providence, 1967.

[26] C. Lederman and P.A. Markowich. On fast-diffusion equations with infinite equilibrium entropy and finite equilibrium mass. *Comm. Partial Differential Equations*, 28:301–332, 2003.

[27] K.A. Lee and J.L.Vazquez. Geometrical properties of solutions of the porous medium equation for large times. *Indiana Univ. Math. J.*, 52:991–1016, 2003.

[28] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.

[29] J.D. Murray. *Mathematical Biology, 2nd edition*. Springer-Verlag, Berlin, 1993.

[30] M. Muskat. *The Flow of Homogeneous Fluids Through Porous Media*. McGraw-Hill, New York, 1937.

[31] W. Newman. A Lyapunov functional for the evolution of solutions to the porous medium equation to self-similarity. I. *J. Math. Phys.*, 25:3120–3123, 1984.

18

[32] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26:101–174, 2001.

[33] R.E. Pattle. Diffusion from an instantaneous point source with concentration dependent coefficient. *Quart. J. Mech. Appl. Math.*, 12:407–409, 1959.

[34] M. Pierre. Nonlinear fast diffusion with measures as data. *Pitman Research Notes in Math*, 149:179–188, 1987.

[35] M. del Pino and M. Saez. On the extinction profile for solutions of $u_t = \Delta u^{(N-2)/(N+2)}$. *Indiana Univ. Math. J.*, 50:612–628, 2001.

[36] J. Ralston. A Lyapunov functional for the evolution of solutions to the porous medium equation to self-similarity. II. *J. Math. Phys.*, 25:3124–3127, 1984.

[37] K.-T. Sturm. Convex functionals of probability measures and nonlinear diffusions on manifolds. *J. Math. Pures Appl.*, 84:149–168, 2005.

[38] K.-T. Sturm and M.-K. von Renessee. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure Appl. Math.*, 58:923–940, 2005.

[39] J.L. V´azquez. Asymptotic behaviour for the porous medium equation posed in the whole space. *J. Evol. Equ.*, 3:67–118, 2003.

[40] J.L. Vazquez, J.R. Esteban and A. Rodriguez. The fast diffusion equation with logarithmic nonlinearity and the evolution of conformal metrics in the plane. *Adv. Differential Equations*, 1:21–50, 1996.

[41] T.P. Witelski and A.J. Bernoff. Self-similar asymptotics for linear and nonlinear diffusion equations. *Stud. Appl. Math.*, 100:153–193, 1998.

[42] Ya.B. Zel'dovich and G.I. Barenblatt. The asymptotic properties of self-modelling solutions of the nonstationary gas filtration equations. *Sov. Phys. Doklady*, 3:44–47, 1958.

[43] Ya.B. Zel'dovich and A.S. Kompaneets. Theory of heat transfer with temperature dependent thermal conductivity. In *Collection in Honour of the 70th Birthday of Academician A.F. Ioffe*, pages 61–71. Izdvo. Akad. Nauk. SSSR, Moscow, 1950.

[†]DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, ONTARIO CANADA M5S 3G3
mccann@math.toronto.edu

[‡]DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES CA 90095
slepcev@math.ucla.edu