Information geometry of *q*-Gaussian densities and behaviors of solutions to related diffusion equations

Information geometry of *q*-Gaussian densities and behaviors of solutions to related diffusion equations[*]

Atsumi Ohara[1] and Tatsuaki Wada[2]

[1] Department of Systems Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
[2] Department of Electrical and Electronic Engineering, Ibaraki University, Hitachi, Ibaraki 316-8511, Japan

E-mail: ohara@sys.es.osaka-u.ac.jp and wada@mx.ibaraki.ac.jp

Abstract

This paper presents new geometric aspects of the behaviors of solutions to the porous medium equation (PME) and its associated equation. First we discuss thermostatistical structure with information geometry on a manifold of generalized exponential densities. A dualistic relation between the two existing formalisms is elucidated. Next by equipping the manifold of *q-Gaussian densities* with such a structure, we derive several physically and geometrically interesting properties of the solutions. The manifold is proved invariant and attracting for the evolving solutions, which play crucial roles in our analysis. We demonstrate that the moment-conserving projection of a solution coincides with a geodesic curve on the manifold. Further, the evolutional velocities of the second moments and the convergence rate to the manifold are evaluated in terms of the *Bregman divergence*. Finally we show that the self-similar solution is geometrically special in the sense that it is simultaneously geodesic with respect to the mutually dual two affine connections.

PACS numbers: 89.70.Cf, 05.90.+m, 02.40.Hw

1. Introduction

Let $u(x, t)$ and $p(x, \tau)$ on $\mathbf{R}^n \times \mathbf{R}_+$ be, respectively, the solutions of the following nonlinear diffusion equation, which is called the *porous medium equation* (PME):

$$\frac{\partial u}{\partial t} = \Delta u^m, \qquad 1 < m \in \mathbf{R}, \tag{1}$$

[*] Preliminary forms of several results in this paper will appear in [50] without proofs.

with nonnegative initial data $0 \leqslant u(x, 0) = u_0(x) \in L^1(\mathbf{R}^n)$, and the associated nonlinear Fokker–Planck equation (NFPE)

$$\frac{\partial p}{\partial \tau} = \nabla \cdot (\beta x p + D \nabla p^m), \qquad 0 < \beta \in \mathbf{R}, \tag{2}$$

with nonnegative initial data $0 \leqslant p(x, 0) = p_0(x) \in L^1(\mathbf{R}^n)$. Here, D is a real symmetric positive definite matrix, which represents the diffusion coefficients. As is widely known [16, 17] and shown later, one solution is obtained from a simple transformation from the other, and vice versa.

The PME and NFPE with $m > 1$ represent the *slow diffusion* phenomena, which naturally arise in many physical problems including percolation of fluid through porous media, intensive thermal waves and so on. Classical results can be found in [1–5] and the references therein. Hence, the behaviors of their solutions have been studied analytically and thermostatistically in the literature [6–17], just to name a few. Further, these equations have been found to have close relations with optimal transport, Wasserstein metric and important inequalities in several branches of mathematics, such as functional analysis, probability theory or differential geometry [14–17] and [21–27]. Many research results are summarized in recent monographs [18–21, 24, 27].

For a real number q, consider the *q-Gaussian probability density* function [28, 29] defined by

$$G_q(x; \theta, \Theta) := \exp_q(\theta^T x + x^T \Theta x - \psi(\theta, \Theta)),$$
$$\theta = (\theta^i) \in \mathbf{R}^n, \qquad \Theta = (\theta^{ij}) \in \mathbf{R}^{n \times n}, \tag{3}$$

where $\exp_q t := [1 + (1-q)t]_+^{1/(1-q)}$, Θ is a real symmetric negative definite matrix and $\psi(\theta, \Theta)$ is a normalizing constant. The symbol $\cdot^T$ denotes the transpose of a vector or matrix and $[a]_+$ for a real a indicates $\max\{0, a\}$. Let $\mathcal{M}$ be the family of q-Gaussian densities specified by the parameters (θ, Θ), i.e.

$$\mathcal{M} := \{G_q(x; \theta, \Theta) | \theta \in \mathbf{R}^n, 0 > \Theta = \Theta^T \in \mathbf{R}^{n \times n}\}. \tag{4}$$

The main purpose of the present paper is to study how solutions of the PME and NFPE behave relatively to the q-Gaussian family $\mathcal{M}$. There are two major reasons for this novel viewpoint in the behavioral analysis. First, $\mathcal{M}$ is proved to be an invariant manifold which all the solutions of the PME and NFPE asymptotically approach. This implies that $\mathcal{M}$ analogously plays a central role in the analysis of the *self-similar (Barenblatt–Pattle) solution* of the PME [16] or the asymptotically stable equilibria of the NFPE. Hence, like the classical convergence analysis of the above two special solutions, we can expect to derive new interesting properties from this viewpoint. Secondly $\mathcal{M}$ admits *information geometry* [30–33] compatible with the Legendre structure of *generalized thermostatistics* [34, 35]. The geometry supplies to us several concepts such as projections or geodesics, which are useful tools to characterize a certain geometrical aspect of those solutions. They also give us clear physical interpretations, like evolutions of moments or the maximization of entropy. Consequently, we derive several new and interesting geometrical properties and physical information of solutions to the PME and NFPE. In mathematical statistics several families of multivariate distributions have been studied via information geometry [38–40] because of their importance in many applications. Further, Riemannian geometries of the family of q-Gaussians $\mathcal{M}$ are discussed with various Riemannian metrics in [41].

In section 2 we introduce and review Naudts' generalized thermostatistical theory [34–37] from the standpoint of information geometry, in particular, via recent work by Eguchi *et al* [32, 33]. See also [42, 43] for another context. As a by-product the dualistic relation

between two types of Bregman divergence is obtained, which clearly connects two formalisms proposed by Naudts and Eguchi *et al*. Further, we define projections to the generalized exponential family and notions of geodesics, which are naturally induced from information geometric structure. Among them the *m-projection* and *m-geodesic*, as well as the Bregman divergence, are our important tools to study the behavior. In section 3 we demonstrate the main results on behaviors of solutions in terms of introduced geometric concepts on the q-Gaussian family $\mathcal{M}$. We first prove that $\mathcal{M}$ is an invariant manifold for the PME and NFPE. Next, utilizing the convenient property that the m-projection of a density to $\mathcal{M}$ conserves its first and second moment, we study the behavior of the solutions to the PME and NFPE. Consequently, evolutions of the second moments and the convergence rate of solutions to the manifold are characterized by the divergence. Further, the trajectory of the m-projection for a solution is proved to be an m-geodesic curve on $\mathcal{M}$. Finally, we discuss a special geometric feature of the self-similar solution.

2. Legendre structure on the generalized exponential family

2.1. Generalized entropy and Bregman divergence

Following [34–37], we introduce generalized entropy and the Bregman divergence on the space of probability density functions. By bringing results derived from the *U-divergence* by Eguchi *et al* [32, 33] within a scope, we show remarks that clarify the relation between their formalisms.

For a fixed strictly increasing and positive function $\phi(s)$ on $(0, \infty)$, define a *generalized logarithmic function* as follows:

$$\ln_\phi(t) := \int_1^t \frac{1}{\phi(s)} \, \mathrm{d}s, \qquad t > 0.$$

Note that $\ln_\phi$ is concave and strictly increasing and satisfies $\ln_\phi(1) = 0$. A *generalized exponential function* denoted by $\exp_\phi$ is defined as the inverse function of $\ln_\phi$, which can be extended on $\mathbf{R}$ by respectively putting 0 or $+\infty$ on the smaller or larger outside of the range $\ln_\phi$. The function $\exp_\phi$ is confirmed to be strictly increasing and convex. For two probability density functions $p(x)$ and $r(x)$, we define the *Bregman divergence* as follows:

$$\mathcal{D}_\phi[p\|r] := \int F_\phi(p(x)) - F_\phi(r(x)) - \ln_\phi r(x)(p(x) - r(x)) \, \mathrm{d}x, \qquad (5)$$

where $F_\phi(s)$ is defined for $s > 0$ by

$$F_\phi(s) := \int_1^s \ln_\phi(t) \, \mathrm{d}t, \qquad F_\phi(0) := \lim_{s \to 0_+} F_\phi(s) < +\infty \text{ :assumed.} \qquad (6)$$

Note that F_ϕ is convex because $\ln_\phi$ is monotone increasing. The divergence, if it exists, is positive except $p(x) = r(x)$ (a.e.).

Introduce a generalized entropy functional defined by

$$\mathcal{I}_\phi[p] := \int -F_\phi(p(x)) + (1 - p(x))F_\phi(0) \, \mathrm{d}x, \qquad (7)$$

which is positive and concave with respect to p because F_ϕ is convex and $F_\phi(1) = 0$. We omit the justification of definition (7) as the generalized entropy (see for [34, 35]) because it needs arguments for duality of the generalized logarithmic functions.

Here we should make two remarks. First, defining a function on $\mathbf{R}$ by

$$U_\phi(t) := \int_0^t \exp_\phi u \, \mathrm{d}u$$

and integrating the right-hand side of (6) by part, we have

$$
F_\phi(s) = s \ln_\phi s - \int_1^s t \frac{\mathrm{d} \ln_\phi t}{\mathrm{d}t} \mathrm{d}t
$$

$$
= s \ln_\phi s - \int_0^{\ln_\phi s} \exp_\phi u \, \mathrm{d}u
$$

$$
= s \ln_\phi s - U_\phi(\ln_\phi s),
$$

or equivalently,

$$
U_\phi(t) = t \exp_\phi t - F_\phi(\exp_\phi t). \tag{8}
$$

Thus, U_ϕ is regarded together with the relation (6), as the Legendre conjugate of F_ϕ, and hence, is convex. Then we have the dual expression of the divergence.

Proposition 1. *The Bregman divergence (5) is expressed in the dual form by*

$$
\mathcal{D}_\phi[p\|r] = \int U_\phi(\ln_\phi r(x)) - U_\phi(\ln_\phi p(x)) - p(\ln_\phi r(x) - \ln_\phi p(x)) \, \mathrm{d}x. \tag{9}
$$

This form of the divergence is called the *U-divergence* [32] and have been studied in the fields of statistics [33] because it is quite convenient in statistical inference from empirical data. In this paper we mainly use this form to discuss the geometry of generalized exponential family.

Second we see that the divergence is represented, using the entropy functional, by

$$
\mathcal{D}_\phi[p\|r] = \mathcal{I}_\phi[r] - \mathcal{I}_\phi[p] - \int (p(x) - r(x)) \ln_\phi r(x) \, \mathrm{d}x
$$

$$
= \Phi_\phi[r] - \mathcal{I}_\phi[p] - \int p(x) \ln_\phi r(x) \, \mathrm{d}x.
$$

Here the functional $\Phi_\phi[p]$ is defined by

$$
\Phi_\phi[p] := \int p(x) \ln_\phi p(x) \, \mathrm{d}x + \mathcal{I}_\phi[p]
$$

$$
= \int U_\phi(\ln_\phi p(x)) + (1 - p(x)) F_\phi(0) \, \mathrm{d}x.
$$

As is seen below, Φ_ϕ vanishes for the standard case $\phi(u) = u$.

Example (*q-logarithmic and exponential functions*). The results calculated in this example are used in section 3. Set $\phi(u) = u^q$, $q > 0$, $q \neq 1$; then we have the *q*-logarithmic and exponential functions [35, 37]:

$$
\ln_\phi t = \ln_q t := (t^{1-q} - 1)/(1-q), \qquad \exp_\phi t = \exp_q t := [1 + (1-q)t]_+^{1/(1-q)},
$$

where $[x]_+ = \max\{x, 0\}$ for $x \in \mathbf{R}$. Note that when $q \to 1$, they recover natural logarithmic and exponential functions. If $2 - q > 0$, we have $F_\phi(s)$ and the constant $F_\phi(0)$, respectively,

$$
F_\phi(s) = \int_1^s \frac{t^{1-q} - 1}{1-q} \, \mathrm{d}t = \frac{1}{1-q} \left(\frac{s^{2-q}}{2-q} - s \right) + \frac{1}{2-q}, \qquad F_\phi(0) = \frac{1}{2-q}.
$$

Consequently it reproduces the corresponding generalized entropy $\mathcal{I}_\phi$:

$$
\mathcal{I}_\phi[p] = \frac{1}{2-q} \int \frac{p(x)^{2-q} - p(x)}{q-1} \, \mathrm{d}x = \frac{1}{2-q} \mathcal{S}_{2-q}[p], \tag{10}
$$

where $\mathcal{S}_q$ is called the *Tsallis entropy* [47] defined by

$$
\mathcal{S}_q[p] = \int \frac{p(x)^q - p(x)}{1-q} \, \mathrm{d}x.
$$

Note that $\mathcal{S}_{2-q}$ is represented by

$$\mathcal{S}_{2-q}[p] = -\int p(x) \ln_q p(x)\, dx = \int p(x) \ln_{2-q}\left(\frac{1}{p(x)}\right) dx.$$

Consider the case $1 - q > 0$, which is used in section 3; then for $s \geqslant -1/(1-q)$ we have

$$U_\phi(s) = \frac{1}{2-q}\{1 + (1-q)s\}^{(2-q)/(1-q)} - \frac{1}{2-q}.$$

Since $\ln_q t \geqslant -1/(1-q)$ for $t \geqslant 0$, we see that

$$U_\phi(\ln_\phi s) = \frac{1}{2-q}(s^{2-q} - 1), \qquad s \geqslant 0.$$

Then the corresponding Bregman divergence of the form (9) is

$$\mathcal{D}_\phi[p\|r] = \int \frac{r(x)^{2-q} - p(x)^{2-q}}{2-q} - p(x)\frac{r(x)^{1-q} - p(x)^{1-q}}{1-q}\, dx, \qquad (11)$$

which is called the *β-divergence* and applied to robust estimation in statistics or machine learning [44–46, 33]. Finally, we have

$$\Phi_\phi[p] = \frac{1}{2-q}\int p(x)^{2-q} - p(x)\, dx.$$

This functional disappears when $q \to 1$.

2.2. Generalized exponential family and its geometry

Let us consider the following finite-dimensional statistical model called *the generalized exponential family* [37, 42] or *U-statistical model* [32], which is defined by

$$\mathcal{M}_\phi := \{p_\theta(x) = \exp_\phi(\theta^T h(x) - \psi_\phi(\theta))|\theta \in \Omega \subset \mathbf{R}^d\} \subset L^1(\mathbf{R}^n).$$

Here $h(x) = (h_i(x)), i = 1, \ldots, d$ is a certain vector-valued function and $\psi_\phi(\theta)$ is a normalizing factor of p_θ, i.e.

$$\int \exp_\phi(\theta^T h(x) - \psi_\phi(\theta))\, dx = 1. \qquad (12)$$

On the open domain Ω we assume that the regularity condition holds, i.e. all the formal calculus necessary below such as convergence of integrations, differentiability of the map from θ to p_θ and so on are valid, so that we can regard $\mathcal{M}_\phi$ as a differentiable manifold. Since the parameter $\theta = (\theta^i)$ specifies a density function in $\mathcal{M}_\phi$, it plays a role of the coordinate system for $\mathcal{M}_\phi$. Differentiating (12) by θ^i (we denote basis tangent vectors by $\partial_i := \partial/\partial\theta^i$), we have

$$\int \exp'_\phi(\theta^T h(x) - \psi_\phi(\theta))(h_i(x) - \partial_i \psi_\phi(\theta))\, dx = 0, \qquad (13)$$

which will be used later.

One of the simplest ways to define information geometric structure [30, 31] on $\mathcal{M}_\phi$, which is natural with the generalized entropy $\mathcal{I}_\phi$, is invoking the following *potential function*:

$$\Psi_\phi(\theta) := \Phi_\phi[p_\theta] + \psi_\phi(\theta) = \int U_\phi(\ln_\phi p_\theta) + (1 - p_\theta(x))F_\phi(0)\, dx + \psi_\phi(\theta).$$

Note that in the standard case $\phi(u) = u$, we have $\Psi_\phi(\theta) = \psi_\phi(\theta) = \log \int \exp \theta^T h(x)\, dx$ because Φ_ϕ vanishes as is seen in the example of the previous section. It follows from the relation $\exp_\phi = U'_\phi$ that

$$\eta_i(\theta) := \partial_i \Psi_\phi(\theta) = \int h_i(x)p_\theta(x)\, dx = \mathbf{E}_{p_\theta}[h_i(x)], \qquad (14)$$

where we denote by $\mathbf{E}_p[\cdot]$ the expectation operator for the density p. Using (13) we have

$$\partial_i \partial_j \Psi_\phi(\theta) = \int h_i(x) \exp'_\phi(\theta^T h(x) - \psi_\phi(\theta))(h_j(x) - \partial_j \psi_\phi(\theta)) \, dx$$

$$= \int \tilde{h}_i(x) \exp'_\phi(\theta^T h(x) - \psi_\phi(\theta)) \tilde{h}_j(x) \, dx, \tag{15}$$

where $\tilde{h}_i(x) := h_i(x) - \partial_i \psi_\phi(\theta)$. Thus, the Hesse matrix of $\Psi_\phi(\theta)$ is expressed by

$$(\partial_i \partial_j \Psi_\phi) = \int \tilde{h}(x) \tilde{h}^T(x) \exp'_\phi(\theta^T h(x) - \psi_\phi(\theta)) \, dx,$$

$$\text{where} \quad \tilde{h}(x) = (\tilde{h}_1(x) \cdots \tilde{h}_n(x))^T,$$

and we see that it is positive semidefinite because $\exp'_\phi$ is nonnegative, and hence, Ψ_ϕ is a convex function of θ. In the following, we assume that $(\partial_i \partial_j \Psi_\phi) = (\partial \eta_j / \partial \theta^i)$ is positive definite for $\forall \theta \in \Omega$. Hence, $\eta = (\eta_j)$ is locally bijective to θ and we call (η_i) the *expectation coordinate system* for $\mathcal{M}_\phi$. By relation (14) the Legendre conjugate of $\Psi_\phi(\theta)$ is the sign-reversed generalized entropy of $p_\theta \in \mathcal{M}_\phi$, i.e

$$\Psi_\phi^*(\eta) := \theta^T \eta - \Psi_\phi(\theta) = \int p_\theta \log_\phi p_\theta - U_\phi(\ln_\phi p_\theta) - (1 - p_\theta) F_\phi(0) \, dx$$

$$= -\mathcal{I}_\phi[p_\theta]. \tag{16}$$

Hence, Ψ_ϕ can be physically interpreted as the *generalized Massieu potential* [48, 49], and hence our Riemmanian metric $(\partial_i \partial_j \Psi_\phi) = (\partial \eta_j / \partial \theta^i)$ defined below is regarded as a *susceptance* matrix.

As a *Riemannian metric* $g = (g_{ij})$ on $\mathcal{M}_\phi$, which is an inner product for tangent vectors, we use the Hesse matrix of Ψ_ϕ. Note that we can alternatively express (15) as

$$g_{ij}(\theta) = g(\partial_i, \partial_j) := \partial_i \partial_j \Psi_\phi = \int \partial_i p_\theta(x) \partial_j \ln_\phi p_\theta(x) \, dx.$$

For an explicit formula of g in the case of q-Gaussian densities $\mathcal{M}$, see, for example, [41]. Further we define *generalized mixture connection* $\nabla^{(m)}$ and *exponential connection* $\nabla^{(e)}$ by their components

$$\Gamma_{ij,k}^{(m)}(\theta) = g\big(\nabla_{\partial_i}^{(m)} \partial_j, \partial_k\big) := \int \partial_i \partial_j p_\theta(x) \partial_k \ln_\phi p_\theta(x) \, dx,$$

$$\Gamma_{ij,k}^{(e)}(\theta) = g\big(\nabla_{\partial_i}^{(e)} \partial_j, \partial_k\big) := \int \partial_k p_\theta(x) \partial_i \partial_j \ln_\phi p_\theta(x) \, dx. \tag{17}$$

Then the *duality relation of the connections* [30, 31] $\partial_i g_{jk} = \Gamma_{ij,k}^{(m)} + \Gamma_{ik,j}^{(e)}$ holds. From the definition we see that the coordinate system (θ^i) is special in the sense that $\Gamma_{ij,k}^{(e)}$ actually vanishes, i.e.

$$\Gamma_{ij,k}^{(e)}(\theta) = -\int \partial_k p_\theta(x) \partial_i \partial_j \psi_\phi(\theta) \, dx = -\partial_i \partial_j \psi_\phi(\theta) \partial_k \int p_\theta(x) \, dx = 0.$$

Hence, $\nabla^{(e)}$ is a flat connection on $\mathcal{M}_\phi$ and (θ^i) is an *affine coordinate* with respect to $\nabla^{(e)}$, in other words, each ∂_i is parallel with respect to $\nabla^{(e)}$. Note that we have $\Gamma_{ij,k}^{(m)} = \partial_i \partial_j \partial_k \Psi_\phi$. Similarly, we can show that $\nabla^{(m)}$ is also flat on $\mathcal{M}_\phi$ and (η_j) is affine with respect to $\nabla^{(m)}$ via formal argument based on the duality [30, 31].

Thus, we have obtained *dually flat* [30, 31] structure $(g, \nabla^{(m)}, \nabla^{(e)})$ on $\mathcal{M}_\phi$ introduced by the derivatives of Ψ_ϕ. Note that if $\phi(u) = u$, then $\mathcal{M}_\phi$, g, $\nabla^{(m)}$ and $\nabla^{(e)}$ respectively reduce to the *exponential family*, the *Fisher information matrix*, the usual *mixture and exponential*

connections. We shall find in the following that the structure offers useful tools to us for not only the statistical inference but also the analysis of the PME or NFPE.

The next result immediately follows from the fact that the coordinates are affine with respect to the flat connections, and is frequently used in this paper.

Proposition 2. *Let $\mathcal{C}$ be a one-dimensional submanifold in $\mathcal{M}_\phi$. Each coordinate θ^i of $\mathcal{C}$ is a linear function of a common scalar variable, i.e. $\mathcal{C}$ is expressed as a straight line in the coordinate system $\theta = (\theta^i)$ if and only if $\mathcal{C}$ coincides with a $\nabla^{(e)}$-geodesic (e-geodesic, in short) curve. Similarly, $\mathcal{C}$ is expressed as a straight line in the coordinate system $\eta = (\eta_i)$ if and only if $\mathcal{C}$ coincides with a $\nabla^{(m)}$-geodesic (m-geodesic) curve.*

In the rest of this subsection we assume that $\ln_\phi \circ \exp_\phi(t) = t$ holds for all $t \in \mathbf{R}$. For densities $p_\theta(x) \in \mathcal{M}_\phi$ and $p(x)$, we have

$$
\begin{aligned}
\mathcal{D}_\phi[p \| p_\theta] &= \int U_\phi(\ln_\phi p_\theta(x)) - U_\phi(\ln_\phi p(x)) \\
&\quad - p(x)[\theta^T h(x) - \psi_\phi(\theta) - \ln_\phi p(x)] \, \mathrm{d}x \\
&= \Psi_\phi(\theta) + \int F_\phi(p(x)) - (1 - p(x)) F_\phi(0) \, \mathrm{d}x - \theta^T \mathbf{E}_p[h(x)] \\
&= \Psi_\phi(\theta) - \mathcal{I}_\phi[p] - \theta^T \mathbf{E}_p[h(x)].
\end{aligned}
\tag{18}
$$

Further, for p_{θ_1} and p_{θ_2} both in $\mathcal{M}_\phi$, the Bregman divergence is represented by

$$
\begin{aligned}
\mathcal{D}_\phi[p_{\theta_1} \| p_{\theta_2}] &= \Psi_\phi(\theta_2) + \Psi_\phi^*(\eta_1) - \eta_1^T \theta_2 \\
&= \Psi_\phi(\theta_2) - \Psi_\phi(\theta_1) - \eta_1^T(\theta_2 - \theta_1),
\end{aligned}
\tag{19}
$$

where η_1 and η_2 are the expectation coordinates for p_{θ_1} and p_{θ_2}, respectively. Using the above, we introduce the notion of *m-projection*, which is geometrically related to the m-geodesic and orthogonality [30, 31], and prove important properties essential in the behavioral analysis.

Definition 1. *Let $p(x)$ be a given density. If there exists a minimizing density $\hat{p}_\theta(x)$ for the variational problem $\min_{p_\theta \in \mathcal{M}_\phi} \mathcal{D}_\phi[p \| p_\theta]$, or equivalently, a minimizing parameter $\hat{\theta}$ for the problem $\min_{\theta \in \Omega} \mathcal{D}_\phi[p \| p_\theta]$ exists, we call $\hat{p}_\theta(x) = p_{\hat{\theta}}(x)$ the m-projection of $p(x)$ to $\mathcal{M}_\phi$.*

Proposition 3. *Let $\hat{p}_\theta \in \mathcal{M}_\phi$ be the m-projection of p. Then the following properties hold.*

(i) *The m-projection conserves the expectation of $h(x)$, i.e. $\mathbf{E}_p[h(x)] = \mathbf{E}_{\hat{p}_\theta}[h(x)]$,*
(ii) *The following triangular equality holds: $\mathcal{D}_\phi[p \| p_\theta] = \mathcal{D}_\phi[p \| \hat{p}_\theta] + \mathcal{D}_\phi[\hat{p}_\theta \| p_\theta]$ for all $p_\theta \in \mathcal{M}_\phi$.*

Proof. Consider the optimality condition for the convex optimization problem $\min_{p_\theta \in \mathcal{M}} \mathcal{D}_\phi[p \| p_\theta]$. Since the second term of the right-hand side of (18) is constant, we have

$$
\frac{\partial \Psi_\phi}{\partial \theta^i}(\theta) - \mathbf{E}_p[h_i(x)] = 0, \qquad i = 1, \dots, d.
\tag{20}
$$

Let $\hat{\theta}$ be the solution for the above optimality condition, i.e. $\hat{p}_\theta = p_{\hat{\theta}}$. Then, from (14) and (20), statement (i) holds.

For statement (ii), we use (18) and (19). Since statement (i) implies that $\hat{\eta} := \mathbf{E}_{\hat{p}_\theta}[h(x)] = \mathbf{E}_p[h(x)]$, and the identity $\Psi_\phi(\hat{\theta}) + \Psi_\phi^*(\hat{\eta}) - \hat{\theta}^T \hat{\eta} = 0$ holds from (16), we can modify the right-hand side of the triangular equality as

$$
\Psi_\phi(\hat{\theta}) - \mathcal{I}_\phi[p] - \hat{\theta}^T \mathbf{E}_p[h(x)] + \Psi_\phi(\theta) + \Psi_\phi^*(\hat{\eta}) - \theta^T \hat{\eta} = \Psi_\phi(\theta) - \mathcal{I}_\phi[p] - \theta^T \mathbf{E}_p[h(x)],
\tag{21}
$$

which is equal to $\mathcal{D}_\phi[p \| p_\theta]$.   □

Remark 1. From statement (i) the m-projection $\hat{p}_\theta$ is characterized as the density in $\mathcal{M}_\phi$ with the expectation coordinate equal to $\mathbf{E}_p[h(x)]$. In other words, for any p satisfying $\mathbf{E}_p[h(x)] = \mathbf{E}_{\hat{p}_\theta}[h(x)] = \hat{\eta}$, we have

$$\mathcal{D}_\phi[p\|\hat{p}_\theta] = \Psi_\phi(\hat{\theta}) - \mathcal{I}_\phi[p] - \hat{\theta}^T \mathbf{E}_p[h(x)]$$
$$= \Psi_\phi(\hat{\theta}) - \hat{\theta}^T \hat{\eta} - \mathcal{I}_\phi[p] = \mathcal{I}_\phi[\hat{p}_\theta] - \mathcal{I}_\phi[p] \geqslant 0.$$

Thus, $\hat{p}_\theta$ achieves the *maximum entropy* among densities p with the equal expectation of $h(x)$.

3. Several geometric properties of the porous medium and the associated Fokker–Planck equation

3.1. Preliminaries

In this section, we study the Cauchy problems of PME (1) and NFPE (2) from a viewpoint of information geometry on the q-Gaussian family $\mathcal{M}$ given in (4). In other words, we apply the general argument in the previous section to the case where $\phi(u) = u^q$, $q > 0$ (cf the example in section 2) and $h(x)$ is quadratic, i.e. $\theta^T h(x)$ is actually $\Theta \cdot (xx^T) + \theta \cdot x$. Here, $\cdot$ used for matrices A and B denotes their inner product, i.e. $A \cdot B = \text{trace}(A^T B)$.

In the following we fix the relation between the exponents of the PME and the parameter of the q-exponential function by $m = 2 - q$. Hence, we consider the case $1 < m < 2$, or equivalently, $0 < q < 1$. For the brevity, we omit the subscripts ϕ. Further, for $G_q(x; \theta, \Theta) \in \mathcal{M}$ and an arbitrary density $p(x)$, we modify the divergence (9) as

$$\mathcal{D}[p(x)\|G_q(x; \theta, \Theta)] := \int U(\ln_q G_q(x; \theta, \Theta)) - U(\ln_q p(x))$$
$$- p(x)[\theta^T x + x^T \Theta x - \psi(\theta, \Theta) - \ln_q p(x)\, dx].$$

The modified divergence also satisfies positivity, convexity in θ and Proposition 3. In NFPE (2) we can always choose β as an arbitrary constant by a suitable linear scaling of τ and D. Hence, we set β and introduce another constant α for notational simplicity as follows:

$$\beta := \frac{1}{n(m-1)+2}, \qquad \alpha := n\beta.$$

For the q-Gaussian family $\mathcal{M}$, we can regard (θ, Θ) as the canonical coordinates. On the other hand, the expectation coordinates are nothing but the first moment vector and second moment matrix (η, H) defined by

$$\eta = \int x G_q(x; \theta, \Theta)\, dx, \qquad H = \int xx^T G_q(x; \theta, \Theta)\, dx.$$

Note that the dimension of $\mathcal{M}$ is $N := n + n(n+1)/2 = n(n+3)/2$.

We assume that $u(x, 0)$ and $p(x, 0)$, which respectively denote the initial data of the PME and NFPE, are nonnegative and integrable function with finite second moments. Under these assumptions, it is proved that there exists a unique nonnegative weak solution if $m > 0$, and that the mass $M = \int u(x, t)\, dx$ is conserved for all $t > 0$ if $m \geqslant (n-2)/n$. See [16] for details and additional properties. When we consider the solutions, we restrict their initial masses to be normalized to one without loss of generalities.

In this subsection we show three fundamental facts, of which the last one might be new. First of all, we review how the solutions of the PME and NFPE relate each other (cf [16, 17]). Because of this fact the properties of the solution of PME (1) are important to investigate those of NFPE (2) and vise versa.

Proposition 4. *Let $u(x, t)$ be a solution of PME (1) with initial data $u(x, 0) = u_0(x) \in L^1(\mathbf{R}^n)$. Define*

$$p(z, \tau) := (t + 1)^{\alpha} u(x, t), \qquad z := (t + 1)^{-\beta} Rx, \qquad \tau := \ln(t + 1);$$

then $p(z, \tau)$ is a solution of (2) with $\nabla = \nabla_z$, $D = RR^T$ and initial data $p(z, 0) = u_0(R^{-1}z)$.

Proof. It is known [16] that

$$p(y, \tau) := (t + 1)^{\alpha} u(x, t), \qquad y := x(t + 1)^{-\beta}, \qquad \tau := \ln(t + 1) \quad (22)$$

is a solution of the following NFPE:

$$\frac{\partial p}{\partial \tau} = \nabla_y \cdot (\beta y p + \nabla_y p^m), \tag{23}$$

where $\nabla_y = (\partial/\partial y^1 \cdots \partial/\partial y^n)^T$. Since it holds, for the transformation $z = Ry$, that

$$\nabla_z = \left(\frac{\partial}{\partial z^1} \cdots \frac{\partial}{\partial z^n} \right)^T = R^{-T} \nabla_y,$$

we see that $p(y, \tau)$ is the solution of (23) if and only if $p(R^{-1}z, \tau)$ is the solution of

$$\frac{\partial p(R^{-1}z, \tau)}{\partial \tau} = \nabla_z \cdot (\beta z p(R^{-1}z, \tau) + D\nabla_z p(R^{-1}z, \tau)^m).$$

Note that the drift vector is invariant. Thus, the statement follows. $\qquad \square$

Next, the equilibrium density for NFPE (2) is solved using generalized thermostatistical concept and Lyapunov approach. To analyze the behavior of (2) let us define a *generalized free energy*:

$$\begin{aligned}
\mathcal{F}[p] &:= \int \frac{\beta}{2m} x^T D^{-1} x p(x) \, \mathrm{d}x - \mathcal{I}[p] \\
&= \frac{1}{m} \int \frac{\beta}{2} x^T D^{-1} x p(x) + p(x) \ln_q p(x) \, \mathrm{d}x.
\end{aligned} \tag{24}$$

This type of functional was first introduced in [8, 9] and developed by many researchers [13, 16, 17] to discuss convergence of the PME and NFPE. Note that when $n = 1$, it reduces to $U/D - S_{2-q}$ up to constant, with the average energy $U = \mathbf{E}_p[\beta x^2/2]$ for a drift vector βx. Hence, the diffusion coefficient D can be interpreted as the temperature in the thermodynamical argument.

Note that (2) can be rewritten as

$$\frac{\partial p(x, \tau)}{\partial \tau} = (R\nabla) \cdot [\beta R^{-1} x p(x, \tau) + R\nabla p(x, \tau)^{2-q}]$$

using a symmetric matrix R satisfying $R^2 = D$. Recalling the relation $q = 2 - m$, we have

$$\frac{\delta \mathcal{F}}{\delta p} = \frac{1}{m} \left\{ \frac{\beta}{2} (R^{-1} x) \cdot (R^{-1} x) - \frac{m p(x, \tau)^{m-1} - 1}{1 - m} \right\}.$$

Hence, it holds that

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{F}[p(x, \tau)]}{\mathrm{d}\tau} &= \int \frac{\delta \mathcal{F}}{\delta p} \frac{\partial p}{\partial \tau} \, \mathrm{d}x = \int \frac{\delta \mathcal{F}}{\delta p} (R\nabla) \cdot (\beta R^{-1} x p + R\nabla p^m) \, \mathrm{d}x \\
&= -\int \left((R\nabla) \frac{\delta \mathcal{F}}{\delta p} \right) \cdot (\beta R^{-1} x p + R\nabla p^m) \, \mathrm{d}x \\
&= -\frac{1}{m} \int p \| \beta R^{-1} x + m p^{m-2} R\nabla p \|^2 \, \mathrm{d}x \leqslant 0. \tag{25}
\end{aligned}$$

Thus, $\mathcal{F}[p(x, \tau)]$ serves as a Lyapunov functional for (2) and the equilibrium density $p_\infty(x)$ is determined from (25) as a q-Gaussian:

$$
\begin{aligned}
p_\infty(x) = G_q(x; 0, \Theta_\infty) &= \exp_q(x^T \Theta_\infty x - \psi(0, \Theta_\infty)), \\
&= \exp_q(\Theta_\infty \cdot (xx^T) - \psi(0, \Theta_\infty)),
\end{aligned} \tag{26}
$$

where the parameters are given by

$$
\theta_\infty = 0, \qquad \Theta_\infty = -\frac{\beta}{2m} D^{-1}.
$$

From (16) the generalized Massieu potential on the q-Gaussian family $\mathcal{M}$ is represented by

$$
\Psi(\theta, \Theta) = \theta^T \mathbf{E}_{G_q(\theta, \Theta)}[x] + \Theta \cdot \mathbf{E}_{G_q(\theta, \Theta)}[xx^T] + \mathcal{I}[G_q(\theta, \Theta)].
$$

Since the generalized free energy is written as

$$
\mathcal{F}[p] = -\Theta_\infty \cdot \mathbf{E}_p[xx^T] - \mathcal{I}[p],
$$

we can express the difference of the free energies at $p_\infty(x) \in \mathcal{M}$ and $p(x)$ by the divergence via (18):

$$
\begin{aligned}
\mathcal{D}[p\|p_\infty] &= \Psi(0, \Theta_\infty) - \mathcal{I}[p] - \Theta_\infty \cdot \mathbf{E}_p[xx^T] \\
&= \mathcal{F}[p] - \mathcal{F}[p_\infty].
\end{aligned}
$$

Thus, the minimization of $\mathcal{F}[\cdot]$ is equivalent to that of $\mathcal{D}[\cdot \| p_\infty]$, which can be interpreted as the *H-theorem* in statistical physics.

Finally, we show that the q-Gaussian family $\mathcal{M}$ is invariant and attracting for the PME and NFPE, i.e. a solution belongs to $\mathcal{M}$ for all future time if its initial density is in $\mathcal{M}$, and converges to $\mathcal{M}$ otherwise. Hence, together with the previous fact on the equilibrium, analysis with respect to $\mathcal{M}$ is expected to add basic knowledge about the behaviors of the solutions.

Proposition 5. *The q-Gaussian family $\mathcal{M}$ is an invariant and attracting manifold for the PME and NFPE.*

Proof. We prove that $\Delta G_q(x; \theta, \Theta)^m$ belongs to the tangent space of $\mathcal{M}$ at each $G_q(x; \theta, \Theta)$. For the q-Gaussian density $G_q(x; \theta, \Theta)$ defined by (3), we see that $\Delta G_q(x; \theta, \Theta)^m$ is of the form

$$
\Delta G_q(x; \theta, \Theta)^m = \begin{cases} Q(x; \theta, \Theta) G_q(x; \theta, \Theta)^{2-m}, & x \in \operatorname{supp} G_q(x; \theta, \Theta), \\ 0, & x \notin \operatorname{supp} G_q(x; \theta, \Theta), \end{cases}
$$

with a certain quadratic function of x, i.e $Q(x; \theta, \Theta) = x^T A x + b^T x + c$, where the coefficients $A = (a_{ij})$, $b = (b_i)$ and c depend on $\theta = (\theta^i)$ and $\Theta = (\theta^{ij})$. Note that it holds that

$$
\int \Delta G_q(x; \theta, \Theta)^m \, d\mu = 0 \quad \forall \theta, \quad \forall \Theta
$$

from the mass conservation property of the PME. Hence, A, b and c have a linear constraint and the scalar coefficient c is determined by A and b. On the other hand, the natural tangent basis vectors of $\mathcal{M}$ are calculated by

$$
\frac{\partial G_q}{\partial \theta^i}(x; \theta, \Theta) = \begin{cases} \left(x_i + \dfrac{\partial \psi}{\partial \theta^i}\right) G_q(x; \theta, \Theta)^{2-m}, & x \in \operatorname{supp} G_q(x; \theta, \Theta), \\ 0, & x \notin \operatorname{supp} G_q(x; \theta, \Theta), \end{cases}
$$

$$
\frac{\partial G_q}{\partial \theta^{ij}}(x; \theta, \Theta) = \begin{cases} \left((2 - \delta_{ij})x_i x_j + \dfrac{\partial \psi}{\partial \theta^{ij}}\right) G_q(x; \theta, \Theta)^{2-m}, & x \in \operatorname{supp} G_q(x; \theta, \Theta), \\ 0, & x \notin \operatorname{supp} G_q(x; \theta, \Theta), \end{cases}
$$

for $i, j = 1, \ldots, n$ with $i \leqslant j$. Here, δ_{ij} is Kronecker's delta. From the definition of $\mathcal{M}$ they also conserve the mass, i.e.

$$\int \frac{\partial G_q}{\partial \theta^i}(x; \theta, \Theta) \, dx = 0, \qquad \int \frac{\partial G_q}{\partial \theta^{ij}}(x; \theta, \Theta) \, dx = 0 \quad \forall \theta, \quad \forall \Theta.$$

Consider the following linear combination of these tangent vectors with the coefficients a_{ij} and b_i:

$$v := \sum_{i,j} a_{ij} \frac{\partial G_q}{\partial \theta^{ij}}(x; \theta, \Theta) + \sum_i b_i \frac{\partial G_q}{\partial \theta^i}(x; \theta, \Theta).$$

Then v is equal to $\Delta G_q(x; \theta, \Theta)^m$ because both satisfy the mass conservation constraint and the remaining coefficient c should be consequently represented by

$$c = \sum_{i,j} a_{ij} \frac{\partial \psi}{\partial \theta^{ij}} + \sum_i b_i \frac{\partial \psi}{\partial \theta^i} \qquad \forall \theta, \quad \forall \Theta.$$

Thus, $\Delta G_q(x; \theta, \Theta)^m$ belongs to the tangent space of $\mathcal{M}$ at $G_q(x; \theta, \Theta)$. The attractivity is straightforward from the well-known fact that all the solutions of the PME asymptotically converge to the self-similar (Barenblatt–Pattle) solution $u^{\mathrm{BP}}(x, t)$ lying on $\mathcal{M}$. (cf subsection 3.4.) The invariance and attractivity of $\mathcal{M}$ for the NFPE follows from this result and the transformation in proposition 4. □

3.2. Trajectories of m-projections

First we study the behavior of a solution $u(x, t)$ of the PME in terms of its m-projection to $\mathcal{M}$ denoted by $\hat{u}(x, t)$. Owing to the properties of the divergence described in section 2, this is equivalent to consider the first and second moments of $u(x, t)$. Let $\eta^{\mathrm{PM}} = (\eta_i^{\mathrm{PM}})$ and $H^{\mathrm{PM}} = (\eta_{ij}^{\mathrm{PM}})$ be, respectively, the first moment vector and the second moment matrix of u;

$$\eta_i^{\mathrm{PM}}(t) := \mathbf{E}_u[x_i], \qquad \eta_{ij}^{\mathrm{PM}}(t) := \mathbf{E}_u[x_i x_j].$$

Theorem 1. *Consider solutions of the PME with the common initial first and second moments. Then their m-projections to $\mathcal{M}$ evolve monotonically along with the common m-geodesic curve that starts from the density determined by the initial moments.*

Proof. Differentiating η_{ij}^{PM} by t, we have

$$\dot{\eta}_{ij}^{\mathrm{PM}} = \int \frac{\partial u}{\partial t} x_i x_j \, dx = \int \Delta u^m x_i x_j \, dx = -\int \nabla u^m \cdot \nabla(x_i x_j) \, dx$$

$$= \int u^m \Delta(x_i x_j) \, dx = 2\delta_{ij} \int u^m \, dx.$$

Hence, the second moment evolves as follows:

$$\eta_{ij}^{\mathrm{PM}}(t) = \eta_{ij}^{\mathrm{PM}}(0) + \delta_{ij} \sigma_u^{\mathrm{PM}}(t), \qquad \sigma_u^{\mathrm{PM}}(t) := 2 \int_0^t dt' \int u(x, t')^m \, dx.$$

Note that $\sigma_u^{\mathrm{PM}}(t)$ is positive and monotone increasing on $t > 0$. By similar argument we see that $\dot{\eta}^{\mathrm{PM}} = 0$, i.e. the first moment vector is invariant. Thus, the evolution of $(\eta^{\mathrm{PM}}(t), H^{\mathrm{PM}}(t))$ for every solution $u(x, t)$ is represented as a straight line. Recalling that the m-projection conserves these moments from proposition 3, we see that $(\eta^{\mathrm{PM}}(t), H^{\mathrm{PM}}(t))$ is just the expectation coordinates of $\hat{u}(x, t)$. Thus, the trajectory of the m-projection of $u(x, t)$ is an m-geodesic curve by proposition 2. □

Remark 2. (i) From the argument for the NFPE, we will see that $\sigma_u^{\mathrm{PM}}(t) = O(t^{2\beta})$ as $t \to \infty$. (ii) Theorem implies that the trajectories of m-projections on $\mathcal{M}$ for all the PME solutions $u(x, t)$ are parallelized in the expectation coordinates, i.e.

$$\eta^{\mathrm{PM}}(t) = \eta^{\mathrm{PM}}(0), \tag{27}$$

$$H^{\mathrm{PM}}(t) = H^{\mathrm{PM}}(0) + \sigma_u^{\mathrm{PM}}(t)I, \tag{28}$$

where I is the n by n identity matrix. In other words, the PME has the following $N(=\dim \mathcal{M})$ *constants of motion* including the mass M:

$$J_0(=M) := \int u(x, t) \, \mathrm{d}x,$$

$$J_i\left(=\eta_i^{\mathrm{PM}}\right) := \int x_i u(x, t) \, \mathrm{d}x, \qquad i = 1, \ldots, n,$$

$$J_{ij}\left(=\eta_{ij}^{\mathrm{PM}}\right) := \int x_i x_j u(x, t) \, \mathrm{d}x, \qquad i = 1, \ldots, n, \qquad j = 1, \ldots, n, \quad i \neq j,$$

$$J_{kk} := \sum_{i=1}^{n} e_i^{(k)} \left(\int x_i^2 u(x, t) \, \mathrm{d}x - \eta_{ii}^{\mathrm{PM}}(0) \right) \equiv 0, \qquad k = 1, \ldots, n - 1,$$

where $e^{(k)} = \left(e_1^{(k)} \cdots e_n^{(k)}\right)$, $k = 1, \ldots, n - 1$ are a set of $n - 1$ basis vectors of the hyperplane $\mathcal{H} = \left\{x \in \mathbf{R}^n \,\middle|\, \sum_{i=1}^{n} x_i = 0\right\}$. Particularly, a solution on the N-dimensional manifold $\mathcal{M}$ has $N - 1$ constants of motion except the trivial one J_0. This implies the possibility that solutions on $\mathcal{M}$ may be explicitly solved by quadratures.

Note that the m-projection $\hat{u}(x, t)$ of a solution $u(x, t)$ satisfies PME (1) only when $\hat{u}(x, t) = u(x, t)$; in other words, $u(x, t)$ is a solution on the invariant manifold $\mathcal{M}$. This is because the evolutional velocity of each m-projection $\hat{u}(x, t)$ along a common m-geodesic curve depends on how far from $\mathcal{M}$ the corresponding original solution $u(x, t)$ evolves. This phenomenon is specific to the slow diffusion, and is quantitatively evaluated in terms of the expectation coordinates and the divergence as follows.

Let $\hat{f}_0(x) \in \mathcal{M}$ be the m-projection of the density $f_0(x)$. Consider two solutions $u_1(x, t)$ and $u_2(x, t)$ of the PME satisfying $u_1(x, t_0) = f_0(x)$ and $u_2(x, t_0) = \hat{f}_0(x)$ at a certain time $t = t_0$. Note that $u_2(x, t) \in \mathcal{M}$ for all t because $\mathcal{M}$ is invariant. From the moment conservation property of the m-projection stated in proposition 3, the second moment matrices $H_i^{\mathrm{PM}}(t)$ of $u_i(x, t)$ for $i = 1, 2$ satisfy $H_1^{\mathrm{PM}}(t_0) = H_2^{\mathrm{PM}}(t_0)$. However, their velocities at t_0 have the following relation:

$$\dot{H}_1^{\mathrm{PM}}(t_0) - \dot{H}_2^{\mathrm{PM}}(t_0) = 2 \int f_0^m(x) - \hat{f}_0^m(x) \, \mathrm{d}x \, I$$

$$= 2m(m - 1)(\mathcal{I}[\hat{f}_0(x)] - \mathcal{I}[f_0(x)])I$$

by (28) and the expression of the generalized entropy (10). Using the relation in remark 1, we have the following.

Corollary 1. *Let $\hat{f}_0(x) \in \mathcal{M}$ be the m-projection of $f_0(x)$ and assume that two solutions $u_1(x, t)$ and $u_2(x, t)$ of the PME satisfy the conditions $u_1(x, t_0) = f_0(x)$ and $u_2(x, t_0) = \hat{f}_0(x)$ at a certain time $t = t_0$. Then velocities of their respective second moment matrices at t_0 are related by*

$$\dot{H}_1^{\mathrm{PM}}(t_0) - \dot{H}_2^{\mathrm{PM}}(t_0) = 2m(m - 1)\mathcal{D}[f_0(x) \| \hat{f}_0(x)]I.$$
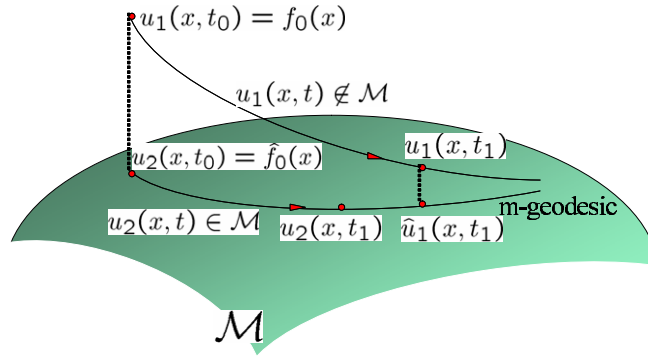
Figure 1. For two solutions $u_1(x, t) \notin \mathcal{M}$ and $u_2(x, t) \in \mathcal{M}$, the m-projection $\hat{u}_1(x, t)$ and $u_2(x, t)$ evolve along a common m-geodesic curve on $\mathcal{M}$ with different velocities.

(This figure is in colour only in the electronic version)

Thus, the m-projection $\hat{u}_1(x, t)$ of $u_1(x, t) \notin \mathcal{M}$, which has the common second moment matrix $H_1^{PM}(t)$ for all t, evolves faster than $u_2(x, t) \in \mathcal{M}$, while $\hat{u}_1(x, t)$ and $u_2(x, t)$ pass along a common m-geodesic curve on $\mathcal{M}$ by theorem 1. (See figure 1 and [50] for numerical experiments.) Finally the corollary suggests that by measuring the diagonal elements of $H_1^{PM}(t)$ we can estimate how far $u_1(x, t)$ is from $\mathcal{M}$ in terms of the divergence. Note that the difference of velocities vanishes when $m \to 1$. Hence, this is the specific property of the slow diffusions governed by the PME.

Next we study the behavior of the solutions $p(x, \tau)$ of NFPE (2). Recall the transformation from a solution $u(x, t)$ of the PME to $p(x, \tau)$ given in proposition 4. Since it holds that $dz = (t + 1)^{-\alpha} \det(R)\, dx$, we have the relations of the moments:

$$\int p(z, \tau)\, dz = \int (t + 1)^{\alpha} u(x, t)\, dz = \det(R) \int u(x, t)\, dx \tag{29}$$

$$\int z p(z, \tau)\, dz = (t + 1)^{-\beta} \det(R) R \int x u(x, t)\, dx \tag{30}$$

$$\int z z^T p(z, \tau)\, dz = (t + 1)^{-2\beta} \det(R) R \left(\int x x^T u(x, t)\, dx \right) R^T. \tag{31}$$

The first relation shows that the solution of the NFPE also conserves its mass. Let $\eta^{FP}(\tau)$ and $H^{FP}(\tau)$ be, respectively, the first and second moments of $p(x, \tau)$, i.e.

$$\eta^{FP} = \left(\eta_i^{FP} \right), \qquad H^{FP} = \left(\eta_{ij}^{FP} \right),$$

where

$$\eta_i^{FP} := \mathbf{E}_p[x_i], \qquad \eta_{ij}^{FP} := \mathbf{E}_p[x_i x_j].$$

From the behavior of the moments of the PME and the above relations, we have

$$\eta^{FP}(\tau) = (t + 1)^{-\beta} \det(R) R \eta^{PM}(t) = e^{-\beta \tau} \det(R) R \eta^{PM}(0)$$
$$= e^{-\beta \tau} \eta^{FP}(0),$$
$$H^{FP}(\tau) = (t + 1)^{-2\beta} \det(R) R H^{PM}(t) R^T$$
$$= e^{-2\beta \tau} \det(R) [R H^{PM}(0) R^T + \sigma_u^{PM}(e^{\tau} - 1) D]$$
$$= e^{-2\beta \tau} H^{FP}(0) + e^{-2\beta \tau} \sigma_p^{FP}(e^{\tau} - 1) D,$$

where the scaling $\tau = \ln(t+1)$ is assumed and $\sigma_p^{\mathrm{FP}}(t)$ is defined by

$$
\sigma_p^{\mathrm{FP}}(t) := \det(R)\sigma_u^{\mathrm{PM}}(t)
$$

$$
= 2\int_0^{\ln(1+t)} \mathrm{d}\tau'\, \mathrm{e}^{\tau'+\alpha(1-m)\tau'} \int p(x,\tau')^m \,\mathrm{d}x
$$

for a solution u of the PME and the corresponding solution p of the NFPE. Note that differentiating the above by t, we have

$$
(1+t)^{\alpha(1-m)} \int p(x,\tau)^m \,\mathrm{d}x = \det(R) \int u(x,t)^m \,\mathrm{d}x. \tag{32}
$$

For the limiting case $m \to 1$ (and accordingly $\beta \to 1/2$), we see that the above expressions recover the well-known linear Fokker–Plank case with a drift vector $x/2$:

$$
\eta^{\mathrm{FP}}(\tau) = \mathrm{e}^{-\tau/2}\eta^{\mathrm{FP}}(0), \qquad H^{\mathrm{FP}}(\tau) = \mathrm{e}^{-\tau}H^{\mathrm{FP}}(0) + 2(1-\mathrm{e}^{-\tau})D.
$$

Since we know that $p(x,\tau)$ converges to $p_\infty(x) \in \mathcal{M}$ in (26) and it holds that

$$
\lim_{\tau\to\infty} H^{\mathrm{FP}}(\tau) = \sqrt{\det D}\Big(\lim_{t\to\infty}(t+1)^{-2\beta}\sigma_u^{\mathrm{PM}}(t)\Big)D \tag{33}
$$

because $\det R = \sqrt{\det D}$, we conclude that the left-hand side of (33) exists and $\sigma_u^{\mathrm{PM}}(t) = O(t^{2\beta})$ as $t \to \infty$ (cf remark 2). Summing up the above with proposition 2, we obtain the following geometric property of the NFPE.

Corollary 2. *Consider the solutions of the NFPE with the common initial first and second moments. Then their m-projections to $\mathcal{M}$ evolve along with the common m-geodesic curve from the density determined by the initial moments to the equilibrium $p_\infty(x)$.*

Note that the following relation holds with the scaling $\tau = \ln(t+1)$:

$$
\frac{\mathrm{d}}{\mathrm{d}\tau} H^{\mathrm{FP}}(\tau) = (t+1)^{-2\beta}\left(-2\beta H^{\mathrm{FP}}(0) - 2\beta\sigma_u^{\mathrm{PM}}(t)D + (t+1)\frac{\mathrm{d}\sigma_u^{\mathrm{PM}}(t)}{\mathrm{d}t}D \right). \tag{34}
$$

Hence, we cannot guarantee the monotonic behavior of the second moment matrix $H^{\mathrm{FP}}(\tau)$ unlike the linear Fokker–Plank equation. For example, if the initial density $p(x,0)$ is not on $\mathcal{M}$ but has the common second moments with the equilibrium density, we cannot expect that the right-hand side of (34) is zero and the second moment matrix possibly oscillates around its equilibrium.

3.3. Convergence rate of the solution of the PME to $\mathcal{M}$

Finally, we show that the triangle equality of the divergence is useful to estimate the convergence rate of the solution of the PME to $\mathcal{M}$. It is known [14, 15, 17] that a solution of the NFPE decays exponentially with respect to the divergence, i.e.

$$
\mathcal{D}[p(x,\tau)\|p_\infty(x)] = \mathcal{F}[p(x,\tau)] - \mathcal{F}[p_\infty(x)] \leqslant (\mathcal{D}[p(x,0)\|p_\infty(x)])\,\mathrm{e}^{-2\beta\tau}. \tag{35}
$$

Proposition 6. *Let $u(x,t)$ be a solution of the PME and $\hat{u}(x,t)$ be the m-projection of $u(x,t)$ to the q-Gaussian family $\mathcal{M}$ at each t. Then $u(x,t)$ asymptotically approaches $\mathcal{M}$ satisfying*

$$
\mathcal{D}[u(x,t)\|\hat{u}(x,t)] \leqslant \frac{C_0}{1+t},
$$

where C_0 is a constant depending on the initial function $u(x,0)$.

Proof. Owing to the triangular equality of the m-projection in proposition 3, it holds that

$$
\mathcal{D}[p(x,\tau)\|\hat{p}(x,\tau)] + \mathcal{D}[\hat{p}(x,\tau)\|p_\infty(x)] = \mathcal{D}[p(x,\tau)\|p_\infty(x)].
$$

Together with (35), we have

$$\mathcal{D}[p(x,\tau)\|\hat{p}(x,\tau)] \leqslant \mathcal{D}[p(x,0)\|p_\infty(x)]\,e^{-2\beta\tau}.$$

Let $u(x,t)$ and $\hat{u}(x,t)$ be functions defined from $p(x,\tau)$ and $\hat{p}(x,\tau)$, respectively, through the transformation in proposition 4. Then, it is easy to see that $\hat{u}(x,t) \in \mathcal{M}$ if and only if $\hat{p}(x,\tau) \in \mathcal{M}$ at each fixed t (and τ). Further, since the first and second moments of $p(x,\tau)$ and $\hat{p}(x,\tau)$ are equal, so are those of $u(x,t)$ and $\hat{u}(x,t)$ from (30) and (31). Thus, we conclude that $\hat{u}(x,t)$ is also the m-projection of $u(x,t)$. It holds from (32) that

$$\det(R)\int \hat{u}(x,t)^m - u(x,t)^m \,\mathrm{d}x = (1+t)^{\alpha(1-m)}\int \hat{p}(x,\tau)^m - p(x,\tau)^m \,\mathrm{d}x.$$

Hence, the relation in remark 1 shows that

$$\mathcal{D}[u(x,t)\|\hat{u}(x,t)] = (1+t)^{\alpha(1-m)}\mathcal{D}[p(x,\tau)\|\hat{p}(x,\tau)]/\det(R)$$
$$\leqslant (1+t)^{\alpha(1-m)-2\beta}C_0 = (1+t)^{-1}C_0$$

because $\alpha(1-m) - 2\beta = -1$. $\qquad\square$

Remark 3. From the above result we can conclude that the L^1 convergence of $u(x,t)$ to $\mathcal{M}$ is with the rate $1/\sqrt{1+t}$, using the Csiszar–Kullback inequality [14]. This implies that the convergence to $\mathcal{M}$ is faster than that to the self-similar solution of the PME (cf the next subsection), the rate of which is known to be $1/t^\beta$ when $1 < m \leqslant 2$ [16, 17]. This might be a new aspect of the convergence properties for solutions of the PME.

3.4. The trajectory of self-similar solution

For PME (1) there exists a special solution on $\mathcal{M}$ called the *self-similar solution* or *Barenblatt–Pattle solution* [51, 52]. The solution is expressed in terms of the q-Gaussian density for the case of unit mass by

$$u^{\mathrm{BP}}(x,t) := t^{-\alpha}\left(C - \frac{(m-1)\beta}{2m}t^{-2\beta}x^T x\right)^{1/(m-1)}$$
$$= t^{-\alpha}\exp_q(t^{-2\beta}x^T\Theta(1)x - \psi(0,\Theta(1)))$$
$$= \exp_q(x^T\Theta(t)x - \psi(0,\Theta(t)))$$
$$= G_q(x;0,\Theta(t)), \qquad \Theta(t) = -t^{-1}\frac{\beta}{2m}I, \qquad (36)$$

where C is a constant for $u^{\mathrm{BP}}(x,t)$ to have unit mass and the normalizing constant satisfies

$$\psi(0,\Theta(t)) = \frac{1 - Ct^{\alpha(1-m)}}{m-1}.$$

The self-similar solution plays an important role in analysis of the PME [16]. It is known that any solution for the PME with unit initial mass converges to $u^{\mathrm{BP}}(x,t)$, e.g. in L^1 norm $\lim_{t\to\infty}\|u(x,t) - u^{\mathrm{BP}}(x,t)\|_1 = 0$ [16, 17]. Geometrically, it is also special in the following sense.

Lemma 1. *The trajectory of the self-similar solution $u^{\mathrm{BP}}(x,t)$ is a curve on $\mathcal{M}$ that is simultaneously an m- and e-geodesic.*

Proof. In theorem 1 we have already proved that the trajectory is an m-geodesic. Since (36) shows that the trajectory is expressed as a straight line in the canonical coordinate system (θ,Θ), the statement follows from proposition 2. $\qquad\square$

Remark 4. The property of the self-similar solution stated in the above lemma is called *doubly autoparallel* [53]. From this we can readily conclude that the trajectory of the self-similar solution is also a geodesic with respect to the Levi–Civita connection.

As an application of this property, we have the following decomposition of the divergence.

Proposition 7. *Let $u(x, t)$ be a solution of the PME and $\hat{u}(x, t)$ be its m-projection to $\mathcal{M}$. For the trajectory of the self-similar solution $u^{\mathrm{BP}}(x, t)$ with $0 < t$ denoted by γ, define the m-projection of $\hat{u}(x, t)$ to γ by*

$$u^*(x, t) := \arg \min_{r(x) \in \gamma} \mathcal{D}[\hat{u}(x, t) \| r(x)].$$

Then it holds for all $t > 0$ that

$$\mathcal{D}[u(x, t) \| u^{\mathrm{BP}}(x, t)] = \mathcal{D}[u(x, t) \| \hat{u}(x, t)] + \mathcal{D}[\hat{u}(x, t) \| u^*(x, t)] + \mathcal{D}[u^*(x, t) \| u^{\mathrm{BP}}(x, t)].$$

Proof. From proposition 3, we have

$$\mathcal{D}[u(x, t) \| u^{\mathrm{BP}}(x, t)] = \mathcal{D}[u(x, t) \| \hat{u}(x, t)] + \mathcal{D}[\hat{u}(x, t) \| u^{\mathrm{BP}}(x, t)].$$

The decomposition

$$\mathcal{D}[\hat{u}(x, t) \| u^{\mathrm{BP}}(x, t)] = \mathcal{D}[\hat{u}(x, t) \| u^{\mathrm{BP}}(x, t)] + \mathcal{D}[u^*(x, t) \| u^{\mathrm{BP}}(x, t)]$$

follows from the standard argument of the Pythagorean relation in information geometry [30, 31]. $\square$

4. Conclusions

We have studied the behavior of the solutions to the PME and NFPE focusing on the q-Gaussian family $\mathcal{M}$. By proving that $\mathcal{M}$ is a stable invariant manifold of both equations, we have obtained several properties of the solutions, e.g. $N(=\dim \mathcal{M})$ constants of motions, the convergence rate to $\mathcal{M}$ and geometrical characterization of the self-similar solution of the PME. In particular, the dependence of the evolutional velocity on the divergence from $\mathcal{M}$ (corollary 1) would be a peculiar phenomenon to the slow diffusion.

Through the analysis, we see that the generalized concepts of statistical physics and the compatibly defined information geometric structure on $\mathcal{M}$ provide us with abundant and precise information on the behavior of solutions.

In [15], Otto reported that the PME can be regarded as a gradient system via Riemannian geometry with the Wasserstein metric. The relation with the framework in the present paper is left unclear. Another important future work would be to confirm how the obtained results are analogously extended to the other parameter ranges: $2 \leqslant m$ or $m < 1$ (fast diffusion), or the other type of nonlinear diffusion equation.

16

References

[1]   Muskat M 1937 *The Flow of Homogeneous Fluids Through Porous Media* (New York: McGraw-Hill)
[2]   Buckmaster J 1977 Viscous sheets advancing over dry beds *J. Fluid Mech.* **81** 735–56
[3]   Larsen E W and Pomraning G C 1980 Asymptotic analysis of nonlinear Marshak waves *SIAM J. Appl. Math.* **39** 201–12
[4]   Kath W L 1985 Waiting and propagating fronts in nonlinear diffusion *Physica* D **12** 375–81
[5]   Barenblatt G I 1996 *Scaling, Self-Similarity and Intermediate Asymptotics* (Cambridge: Cambridge University Press)
[6]   Friedman A and Kamin S 1980 The asymptotic behavior of gas in an N-dimensional porous medium *Trans. Am. Math. Soc.* **262** 551–63
[7]   Aronson D G 1985 *The Porous Medium Equation* (*Lecture Notes in Mathematics* vol 1224) (Berlin: Springer)
[8]   Newman W 1984 A Lyapunov functional for the evolution of solutions to the porous medium equation to self-similarity I *J. Math. Phys.* **25** 3124–27
[9]   Ralston J 1984 A Lyapunov functional for the evolution of solutions to the porous medium equation to self-similarity II *J. Math. Phys.* **25** 3120–23
[10] Plastino A R and Plastino A 1995 Non-extensive statistical mechanics and generalized Fokker–Planck equation *Physica* A **222** 347–54
[11] Shiino M 1998 H-theorem with generalized relative entropies and the Tsallis statistics *J. Phys. Soc. Japan* **67** 3658–60
[12] Frank T D 2001 Lyapunov and free energy functionals of generalized Fokker–Planck equations *Phys. Lett.* A **290** 93–100
[13] Frank T D 2002 Generalized Fokker–Planck equations derived from generalized linear nonequilibrium thermodynamics *Physica* A **310** 397–412
[14] Carrillo J A and Toscani G 2000 Asymptotic L1-decay of solutions of the porous medium equation to self-similarity *Indiana Univ. Math. J.* **49** 113–41
[15] Otto F 2001 The geometry of dissipative evolution equations: the porous medium equation *Commun. Partial Differ. Equ.* **26** 101–74
[16] Vázquez J L 2003 Asymptotic behaviour for the porous medium equation posed in the whole space *J. Evol. Equ.* **3** 67–118
[17] Toscani G 2005 A central limit theorem for solutions of the porous medium equation *J. Evol. Equ.* **5** 185–203
[18] Frank T D 2005 *Nonlinear Fokker–Planck Equations: Fundamentals and Applications* (Berlin: Springer)
[19] Vázquez J L 2006 *Smoothing and Decay Estimates for Nonlinear Diffusion Equations: Equations of Porous Medium Type* (Oxford: Oxford University Press)
[20] Vázquez J L 2007 *The Porous Medium Equation: Mathematical Theory* (Oxford: Oxford University Press)
[21] Ambrosio L A, Gigli N and Savaré G 2005 *Gradient Flows: In Metric Spaces and in the Space of Probability Measures* (Basel: Birkhäuser)
[22] Carrillo J A, Di Francesco M and Toscani G 2006 Intermediate asymptotics beyond homogeneity and self-similarity: long time behavior for $u_1 = \Delta\phi(u)$ *Arch. Ration. Mech. Anal.* **180** 127–49
[23] Dolbeault J and del Pino M 2002 Best constants for Gagliardo–Nirenberg inequalities and application to nonlinear diffusions *J. Math. Pures Appl.* **81** 847–75
[24] Villani C 2003 *Topics in Optimal Transportation* (Providence: American Mathematical Society)
[25] Carrillo J A, McCann R J and Villani C 2006 Contractions in the 2-Wasserstein length space and thermalization of granular media *Arch. Ration. Mech. Anal.* **179** 217–63
[26] Otto F and Westdickenberg M 2006 Eulerian calculus for the contraction in the Wasserstein distance *SIAM J. Math. Anal.* **37** 1227–55
[27] Villani C 2009 *Optimal Transport: Old and New* (Berlin: Springer)
[28] Tsallis C, Levy S V F, Souza A M C and Maynard R 1995 Statistical–mechanical foundation of the ubiquity of Levy distributions in nature *Phys. Rev. Lett.* **75** 3589–93
      Tsallis C, Levy S V F, Souza A M C and Maynard R 1996 Statistical–mechanical foundation of the ubiquity of Levy distributions in nature *Phys. Rev. Lett.* **77** 5442 (erratum)
[29] Suyari H and Tsukada M 2002 Law of errors in Tsallis statistics *IEEE Trans. Inf. Theory* **51** 753–7
[30] Amari S I 1985 *Differential-Geometrical Methods in Statistics* (*Lecture Notes in Statistics* vol 28) (New York: Springer)
[31] Amari S I and Nagaoka H 2000 *Methods of Information Geometry* Trans. Math. Monogr. vol 191 (Providence: American Mathematical Society)
[32] Eguchi S 2006 Information geometry and statistical pattern recognition *Sugaku Expositions* **19** 197–216 (originally 2004 Sūgaku **56** 380–99 *in Japanese*)

[33] Murata N, Takenouchi T, Kanamori T and Eguchi S 2004 Information geometry of U-Boost and Bregman divergence *Neural Comput.* **16** 1437–81

[34] Naudts J 2002 Deformed exponentials and logarithms in generalized thermodynamics *Physica* A **316** 323–34

[35] Naudts J 2004 Generalized thermodynamics based on deformed exponentials and logarithmic functions *Physica* A **340** 32–40

[36] Naudts J 2004 Continuity of a class of entropies and relative entropies *Rev. Math. Phys.* **16** 809–22

[37] Naudts J 2004 Estimators, escort probabilities, and ϕ-exponential families in statistical physics *J. Ineq. Pure Appl. Math.* **5** 102

[38] Burbea J 1986 Informative geometry of probability spaces *Expo. Math.* **4** 347–78

[39] Mitchell A F S 1989 The information matrix, skewness tensor and α-connections for the general multivariate elliptic distribution *Ann. Inst. Stat. Math.* **41** 289–304

[40] Oller J M and Corcuera J M 1995 Intrinsic analysis of statistical estimation *Ann. Stat.* **23** 1562–81

[41] Andai A 2008 On the geometry of generalized Gaussian distributions *J. Multivariate Anal.* **100** 777–93

[42] Grunwald P D and David A P 2004 Game theory, maximum entropy, minimum discrepancy and robust baysian decision theory *Ann. Stat.* **32** 1367–433

[43] David A P 2007 The geometry of proper scoring rules *Ann. Inst. Stat.* **59** 77–93

[44] Minami M and Eguchi S 2002 Robust blind source separation by beta-divergence *Neural Comput.* **14** 1859–86

[45] Higuchi I and Eguchi S 2004 Robust principal component analysis with adaptive selection and for Tuning Parameters *J. Mach. Learn. Res.* **5** 453–71

[46] Takenouchi T and Eguchi S 2004 Robustifying adaboost by adding the naive error rate *Neural Comput.* **16** 767–87

[47] Tsallis C 1998 Possible generalization of Boltzmann–Gibbs statistics *J. Stat. Phys.* **52** 479–87

[48] Wada T and Scarfone A M 2005 Connection between Tsallis' formalisms employing the standard linear average energy and ones employing the normalized q-average energy *Phys. Lett.* A **335** 351–62

[49] Callen H B 1985 *Thermodynamics and an Introduction to Thermostatistics* 2nd edn (New York: Wiley)

[50] Ohara A 2009 Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation *Eur. Phys. J.* B **70** 15–28

[51] Barenblatt G I 1952 On some unsteady motions of a liquid or a gas in a porous medium *Prikl. Mat. Mekh.* **16** 67–78 (in Russian)

[52] Pattle R E 1959 Diffusion from an instantaneous point source with concentration dependent coefficient *Q. J. Mech. Appl. Math.* **12** 407–09

[53] Uohashi K and Ohara A 2004 Jordan algebras and dual affine connections on symmetric cones *Positivity* **8** 369–78