

ENERGY DISTANCE AND EVOLUTION PROBLEMS: A PROMISING TOOL FOR KINETIC EQUATIONS

GENNARO AURICCHIO AND GIUSEPPE TOSCANI

ABSTRACT. We study the rate of convergence to equilibrium of the solutions to Fokker–Planck type equations with linear drift by means of Cramér and Energy distances, which have been recently widely used in problems related to AI, in particular for tasks related to machine learning. In all cases in which the Fokker–Planck type equations can be treated through these distances, it is shown that the rate of decay is improved with respect to known results which are based on the decay of relative entropy.

1. INTRODUCTION

In recent years, Cramér’s distance [9], also known as the continuous ranked probability score, has proven to be a useful tool in problems related to AI, and in particular to machine learning [4], especially for tasks involving distribution estimation, regression, and image generation. As explained in [4], the main reason behind the rediscovery of Cramér’s distance, which is a metric for comparing probability distributions, lies in its properties, in particular that of having unbiased sample gradients, making it easier to optimize using stochastic gradient descent.

Unfortunately, Cramér’s distance can only be successfully applied to problems involving univariate probability distributions. A closely related distance, which shares the same properties as Cramér’s distance [1], but is designed to also compare multivariate probability distributions, is the Energy distance introduced by Székely [25, 26, 27]. In [1], the relationships between the Energy distance and other well-known metrics—such as the Wasserstein distance [32] and the Fourier-based metrics introduced in [14], which are widely used in kinetic theory—were thoroughly analyzed. These metrics are of particular interest due to their applicability in the field of artificial intelligence [2, 3, 10, 23].

In this paper, we show that these distances are particularly adapted to study the rate of convergence to equilibrium of the solutions to the linear diffusion equation and to Fokker–Planck type equations characterized by a linear drift, including the one-dimensional Fokker–Planck type equations modeling socio-economic phenomena studied in [12], providing in all cases a decay better than exponential.

A further advantage of this analysis is related to the numerical approximation of evolution equations of Fokker–Planck type. In reason of the simple form of the Energy distance, which is essentially based on the computation of moments, the numerical study of the evolution of solutions towards a steady state could be greatly simplified.

Key words and phrases. Kinetic models; Fokker–Planck equations; Cramér distance; Energy distance; Large-time behavior.

In more details, in Section 2 we briefly report the definitions and the main properties of both Cramér and Energy distances, as well as their connections. The equivalence between these two distances in the one-dimensional case will be shown through a proof which is closely related to an analogous one that can be found in the literature devoted to the Gini index [31]. Then, in Section 3 we will show how these distances can be fruitfully employed to study convergence to equilibrium of solutions to one-dimensional Fokker–Planck equations with linear drift and general diffusion term, and to recover exponential rate of convergence in time. In Section 4 we will show that the Energy distance can be used to recover exponential convergence to equilibrium in the classical linear Fokker–Planck equation in any dimension. Last, Section 5 will be devoted to an application of the Energy distance to the linear diffusion equation. There, the rate of decay of the energy distance between two solutions of the linear diffusion is shown to improve with the increasing of the value of the order α characterizing the distance. Some conclusions will follow in Section 6.

2. BASIC FACTS ABOUT CRAMÉR AND ENERGY DISTANCES

Let F be a probability measure over $\mathbb{R}$. When F is continuous, we will assume it has density f . We will associate F to a random variable X , such that for a subset of the reals $A \subseteq \mathbb{R}$, we have

$$F(X \in A) = \int_A dF(x).$$

The (cumulative) distribution function of X is then

$$(2.1) \quad F(x) = P(X \leq x) = \int_{-\infty}^x dF(x).$$

Given $s > 0$, we denote by $P_s(\mathbb{R})$ the class of all probability measures F over $\mathbb{R}$ such that

$$m_s(F) = \int_{\mathbb{R}} |x|^s dF(x) < +\infty.$$

If $s = 1$, the quantity $m_1(F)$ defines the (absolute) mean value of X , say $E(|X|)$. In this relevant case we can express the standard mean value $E(X)$ directly in terms of the cumulative function, through the formula

$$(2.2) \quad E(X) = - \int_{-\infty}^0 F(x) dx + \int_0^{+\infty} (1 - F(x)) dx.$$

Given two random variables X and Y of cumulative distribution functions F (respectively G) the Cramér distance between X and Y is given by [9]

$$(2.3) \quad d(F, G) = \int_{\mathbb{R}} [F(x) - G(x)]^2 dx.$$

Note that, as written, Cramér distance is not a metric, while its square root is.

One of the interesting properties of Cramér's distance is that it can be written equivalently in different ways. It is immediate to show that it can be written in terms

of Fourier transforms, where, as usual, the Fourier transform of a probability density f in $P_s(\mathbb{R})$ is defined by

$$\widehat{f}(\xi) = \int_{\mathbb{R}} e^{-i\xi x} dF(x).$$

Indeed, for any given pair of probability measures $F, G \in P_s(\mathbb{R})$, Parseval formula implies

$$(2.4) \quad \int_{\mathbb{R}} [F(x) - G(x)]^2 dx = \frac{1}{2\pi} \int_{\mathbb{R}} |\widehat{F}(\xi) - \widehat{G}(\xi)|^2 d\xi,$$

where $\widehat{F}$ and $\widehat{G}$ are the Fourier transforms of the probability measures F, G . On the other hand, if

$$\widehat{f}(\xi) = \int_{\mathbb{R}} e^{-i\xi x} dF(x), \quad \widehat{g}(\xi) = \int_{\mathbb{R}} e^{-i\xi x} dG(x),$$

integration by parts gives

$$\widehat{F}(\xi) - \widehat{G}(\xi) = \frac{\widehat{f}(\xi) - \widehat{g}(\xi)}{i\xi}.$$

Consequently, Cramér distance (2.3) can be equivalently written as

$$(2.5) \quad d(F, G) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{|\widehat{f}(\xi) - \widehat{g}(\xi)|^2}{\xi^2} d\xi.$$

Also, assuming that the random variables X and Y are mutually independent, and that X' and Y' are independent copies of X (respectively Y), Cramér's distance can be expressed as

$$(2.6) \quad d(F, G) = E(|X - Y|) - \frac{1}{2}E(|X - X'|) - \frac{1}{2}E(|Y - Y'|).$$

Note that (2.6) implies that the Cramér distance of F and G is bounded as soon as $F, G \in P_1(\mathbb{R})$.

This unexpected way of writing Cramér's distance, which has been shown in [27], can be more easily derived by resorting to an older result related to Gini inequality index [16, 17]. For a random variable $X \geq 0$ distributed according to F , and given two independent copies X and X' , the quantity

$$(2.7) \quad G(X) = \frac{1}{2E(X)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x - y| dF(x) dF(y) = \frac{1}{2E(X)} E(|X - X'|) \leq 1$$

defines the well-known Gini index. Gini index can be expressed in different ways [31]. Following the approach of Kendall and Stuart [18], Dorfman [11] proposed for the Gini index of a continuous income distribution the equivalent expression

$$(2.8) \quad G(X) = 1 - \frac{1}{2E(X)} \int_0^{+\infty} (1 - F(x))^2 dx.$$

A proof of this equivalence, that was omitted in [11] and in a subsequent paper on the same subject [15], can be found in [31], and relies on a simple remark, namely that, for given real numbers x and y

$$|x - y| = x + y - 2 \min\{x, y\},$$

so that

$$E(|X - Y|) = E(X) + E(Y) - 2E(\min\{X, Y\}).$$

It is immediate to show that, for a given pair of independent random variables X and Y , the cumulative function $H(x)$ of $\min\{X, Y\}$ takes the form

$$(2.9) \quad H(x) = P[\min\{X, Y\} \leq x] = 1 - P[(X > x) \cap (Y > x)] = 1 - [1 - F(x)][1 - G(x)].$$

Hence, owing to (2.9) and to formula (2.2) we get

$$\begin{aligned} E(|X - Y|) &= E(X) + E(Y) + 2 \int_{-\infty}^0 H(x) dx - 2 \int_0^{+\infty} (1 - H(x)) dx \\ &= E(X) + E(Y) + 2 \int_{-\infty}^0 [F(x) + G(x) - F(x)G(x)] dx + \\ &\quad - 2 \int_0^{+\infty} [1 - F(x) - G(x) + F(x)G(x)] dx. \end{aligned}$$

Likewise, given two independent copies X and X' distributed according to F we obtain

$$E(|X - X'|) = 2E(X) + 2 \int_{-\infty}^0 [2F(x) - F^2(x)] dx - 2 \int_0^{+\infty} [1 - F(x)]^2 dx,$$

and, given two independent copies Y and Y' distributed according to G

$$E(|Y - Y'|) = 2E(Y) + 2 \int_{-\infty}^0 [2G(x) - G^2(x)] dx - 2 \int_0^{+\infty} [1 - G(x)]^2 dx.$$

These identities clearly imply (2.6).

In one dimension, the Energy distance introduced by Szelély [25, 26, 27] is nothing but twice the Cramér distance, as defined by (2.6). The interest in Energy distance is related to the fact that it can be expressed in any dimension, and that it allows for interesting generalizations.

In what follows we refer to the Energy distance between two multidimensional random variables $\mathbf{X}$ and $\mathbf{Y}$ with densities $f(\mathbf{x})$, and, respectively, $g(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^n$, $n \geq 1$. The Energy distance of order α , with $0 < \alpha < 2$ is given by

$$\begin{aligned} E_\alpha(f, g) &= 2 \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha f(\mathbf{x})g(\mathbf{y}) d\mathbf{x}d\mathbf{y} \\ (2.10) \quad &- \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha f(\mathbf{x})f(\mathbf{y}) d\mathbf{x}d\mathbf{y} - \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha g(\mathbf{x})g(\mathbf{y}) d\mathbf{x}d\mathbf{y} \\ &= - \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{x}) - g(\mathbf{x})][f(\mathbf{y}) - g(\mathbf{y})] d\mathbf{x}d\mathbf{y}. \end{aligned}$$

As for the Cramér distance, the Energy distance of order α can be equivalently expressed in Fourier transform. In this case it holds [1, 27]

$$(2.11) \quad E_\alpha(f, g) = c_{n,\alpha} \int_{\mathbb{R}^n} \frac{|\widehat{f}(\boldsymbol{\xi}) - \widehat{g}(\boldsymbol{\xi})|^2}{|\boldsymbol{\xi}|^{n+\alpha}} d\boldsymbol{\xi},$$

where

$$(2.12) \quad c_{n,\alpha} = \frac{\alpha 2^\alpha \Gamma\left(\frac{n+\alpha}{2}\right)}{2\pi^{n/2} \Gamma\left(\frac{2-\alpha}{2}\right)}.$$

The passage to Fourier transform is the easiest way to verify that right-hand side in (2.10) is nonnegative, so that the mixed integral dominates the sum of the last two. Further, expression (2.11) allows one to relate in a precise way the energy distance of order α to homogeneous Sobolev spaces of fractional order.

3. CONVERGENCE TO EQUILIBRIUM FOR ONE-DIMENSIONAL FOKKER–PLANCK EQUATIONS

To better appreciate how Cramér and Energy distances work in various problems related to kinetic equations, we will apply Cramér’s distance to investigate the large time behavior of the solution $f(x, t)$ to the linear one-dimensional Fokker–Planck equation with constant diffusion $\sigma > 0$, i.e.

$$(3.1) \quad \frac{\partial}{\partial t} f(x, t) = \frac{\partial}{\partial x} \left[\sigma \frac{\partial}{\partial x} f(x, t) + x f(x, t) \right],$$

when the initial value $f_0(x)$ is a probability density. Since equation (3.1) is mass preserving, the solution remains a probability density for all subsequent times. Consequently, equation (3.1) can be equivalently written in terms of the cumulative function $F(x, t)$, as defined in (2.1). Taking the integrals on both sides with respect to the spatial variable from $-\infty$ to x we obtain that $F(x, t)$ satisfies

$$(3.2) \quad \frac{\partial}{\partial t} F(x, t) = \sigma \frac{\partial f(x, t)}{\partial x} + x \frac{\partial F(x, t)}{\partial x},$$

with initial value $F_0(x)$, the cumulative function of the initial probability density $f_0(x)$.

If now $g(x, t)$ is another solution to (3.1) starting from an initial probability density $g_0(x)$, in reason of the linearity, the difference between the two cumulative functions $F(x, t)$ and $G(x, t)$ solves the equation

$$(3.3) \quad \frac{\partial}{\partial t} [F(x, t) - G(x, t)] = \sigma \frac{\partial [f(x, t) - g(x, t)]}{\partial x} + x \frac{\partial [F(x, t) - G(x, t)]}{\partial x}.$$

Therefore

$$(3.4) \quad \begin{aligned} \frac{d}{dt} \int_{\mathbb{R}} [F(x, t) - G(x, t)]^2 dx &= 2 \int_{\mathbb{R}} [F(x, t) - G(x, t)] \frac{\partial}{\partial t} [F(x, t) - G(x, t)] dx \\ &= 2\sigma \int_{\mathbb{R}} [F(x, t) - G(x, t)] \frac{\partial}{\partial x} [f(x, t) - g(x, t)] dx + \int_{\mathbb{R}} x \frac{\partial}{\partial x} [(F(x, t) - G(x, t))^2] dx \\ &= -2\sigma \int_{\mathbb{R}} [f(x, t) - g(x, t)]^2 dx - \int_{\mathbb{R}} [F(x, t) - G(x, t)]^2 dx. \end{aligned}$$

Note that the last line in (3.4) simply follows from integration by parts.

These simple computations show that the Cramér distance between two solutions of the linear one-dimensional Fokker–Planck equation decays to zero at exponential rate, and that each term in it contributes to the decay. In particular, the linear drift is

responsible of an exponential decay at rate $\exp\{-t\}$, while the diffusion term induces a decay that can be easily evaluated by resorting to some well-known inequality, in this case Nash inequality. We will not dwell further on these mathematical tools.

The important aspect to be outlined is that this result of decay continues to hold in the case in which $G(x, t)$ coincides with the equilibrium solution $F_\infty(x)$ of the Fokker–Planck equation, so that the Cramér distance provides (in few lines) a proof of (better than) exponential convergence towards equilibrium of the solution to equation (3.1).

An interesting aspect of the previous computations is related to the fact that in the one-dimensional case it can be applied to a variety of Fokker–Planck type equations, characterized by a linear drift but different diffusion terms. Consider for example the case of the Fokker–Planck equation

$$(3.5) \quad \frac{\partial}{\partial t} f(x, t) = \frac{\partial}{\partial x} \left[\frac{\partial}{\partial x} f^p(x, t) + x f(x, t) \right],$$

that, for $p > 1$, is characterized by the Barenblatt equilibrium solution

$$(3.6) \quad f_\infty(x) = (C^2 - x^2)^{1/(p-1)},$$

and the constant C is chosen so that the mass of f_∞ is equal to 1. Then the previous analysis applies with the main difference related to the fact that now the decay of Cramér’s distance between two solutions is given by

$$(3.7) \quad \begin{aligned} & \frac{d}{dt} \int_{\mathbb{R}} [F(x, t) - G(x, t)]^2 dx \\ &= -2 \int_{\mathbb{R}} [f(x, t) - g(x, t)] [f^p(x, t) - g^p(x, t)] dx - \int_{\mathbb{R}} [F(x, t) - G(x, t)]^2 dx. \end{aligned}$$

Since in (3.7) the contribution of the diffusion term is negative, this results in a (better than) exponential decay to zero of Cramér distance at rate $\exp\{-t\}$. It is remarkable that a related improvement of the exponential decay for equation (3.5) has been observed before in [7] by resorting to the decay of the relative Renyi entropies.

Maybe the most interesting results one can obtain through Cramér distance are related to one-dimensional Fokker–Planck equations of interest in socio-economic problems [12, 20, 21]. As far as the economic situation in a western society is concerned, a very popular mathematical model suitable to describe the evolution of the distribution of wealth $f(x, t)$, $x \in \mathbb{R}_+$, is the Fokker–Planck equation [5, 8]

$$(3.8) \quad \frac{\partial f(x, t)}{\partial t} = \frac{\sigma}{2} \frac{\partial^2}{\partial x^2} (x^2 f(x, t)) + \lambda \frac{\partial}{\partial x} ((x - 1) f(x, t)).$$

The key feature of equation (3.8) is that, for a given initial probability density $f_0(x)$ of mean value equal to 1, and in presence of suitable boundary conditions at the point $x = 0$, the solution preserves both mass and momentum and approaches in time a unique stationary solution of unitary mass [28]. This stationary state is given by the inverse Gamma distribution

$$(3.9) \quad f_\infty(w) = \frac{(\mu - 1)^\mu \exp\left(-\frac{\mu-1}{w}\right)}{\Gamma(\mu) w^{1+\mu}},$$

where the positive constant $\mu > 1$ is given by

$$\mu = 1 + 2\frac{\lambda}{\sigma}.$$

As predicted by the Italian economist Vilfredo Pareto [22], the equilibrium solution (3.9) exhibits a power-law tail for large values of the wealth variable.

The rate of convergence to equilibrium of the solution to equation (3.8) was studied in [29] by expressing the linear kinetic equation in the Fourier space, that made possible to resort to a metric equivalent to the one-dimensional Energy distance of order α , by suitably selecting $0 < \alpha < 2$. As a matter of fact, if we apply to equation (3.8) the same procedure we applied to the classical Fokker–Planck equation (3.1) we end up with the equation

$$(3.10) \quad \begin{aligned} & \frac{d}{dt} \int_{\mathbb{R}_+} [F(x, t) - G(x, t)]^2 dx \\ &= -\sigma \int_{\mathbb{R}_+} x^2 [f(x, t) - g(x, t)]^2 dx - \lambda \int_{\mathbb{R}_+} [F(x, t) - G(x, t)]^2 dx, \end{aligned}$$

which guarantees exponential convergence to equilibrium in Cramér’s distance at rate $\exp\{-\lambda t\}$.

The last example we consider concerns with a mathematical model describing opinion formation [30]. There, the opinion variable is assumed to take values in the bounded interval $\mathcal{I} = (-1, 1)$, the values ± 1 denoting the extremal opinions. Among the various models introduced in [30], one Fokker–Planck type equation has to be distinguished in view of its equilibrium configurations, which are represented by Beta-type probability densities supported in the interval $(-1, 1)$. This Fokker–Planck equation for the evolution of opinion density $f(x, t)$, with $|x| < 1$, is given by

$$(3.11) \quad \frac{\partial f(x, t)}{\partial t} = \frac{\lambda}{2} \frac{\partial^2}{\partial x^2} ((1 - x^2)f(x, t)) + \frac{\partial}{\partial x} ((x - m)f(x, t)).$$

In (3.11), $\lambda/2 > 0$ is the diffusion coefficient, while m , with $-1 < m < 1$ is the mean opinion. Suitable boundary conditions at the boundary points $y = \pm 1$ then guarantee conservation of mass and momentum of the solution [12].

If the initial condition is a probability density, the steady state equals a probability density of Beta type, given by

$$(3.12) \quad f_{m,\lambda}(x) = C_{m,\lambda} (1 - x)^{-1+\frac{1-m}{\lambda}} (1 + x)^{-1+\frac{1+m}{\lambda}}.$$

In (3.12) the constant $C_{m,\lambda}$ is such that the mass of $f_{m,\lambda}$ is equal to one. Since $-1 < m < 1$, $f_{m,\lambda}$ is integrable on $\mathcal{I}$. Note that $f_{m,\lambda}$ is continuous on $\mathcal{I}$, and as soon as $\lambda > 1 + |m|$ tends to infinity as $y \rightarrow \pm 1$. Convergence to equilibrium in relative entropy for the solution to equation (3.11) has been studied in [13], where it was shown that entropy methods can produce exponential convergence in L_1 towards equilibrium with an explicit rate, for some range of the parameters λ and m only. Owing to the weaker Cramér distance, we can easily fill up the gap, by showing that exponential convergence to equilibrium holds for all admissible values of the parameters. Indeed,

the procedure we applied to the classical Fokker–Planck equation (3.1) gives now

$$\begin{aligned}
 (3.13) \quad & \frac{d}{dt} \int_{-1}^1 [F(x, t) - G(x, t)]^2 dx \\
 &= -\frac{\lambda}{2} \int_{-1}^1 (1 - x^2) [f(x, t) - g(x, t)]^2 dx - \int_{-1}^1 [F(x, t) - G(x, t)]^2 dx,
 \end{aligned}$$

thus showing that, at difference with the results in [13], the (better than) exponential convergence to equilibrium in Cramér distance holds independently of the values of the parameters.

Clearly, in view of identity (2.6), the same results can be obtained by resorting to the Energy distance of order $\alpha = 1$. It is important to remark that the choice of an Energy distance of order $\alpha \neq 1$ leads to more intricate computations. An example of this can be found in [28], where the results of exponential convergence in a certain range of the parameter α have been obtained by resorting to the Fourier expression of the distance. We will be back on this aspect in the next Section.

4. THE LINEAR FOKKER–PLANCK EQUATION IN $\mathbb{R}^n$

In this Section we investigate the rate of decay of the Energy distance of order α along the solution to the linear n -dimensional Fokker–Planck equation, $n > 1$

$$(4.1) \quad \frac{\partial}{\partial t} f(\mathbf{x}, t) = \nabla \cdot [\nabla f(\mathbf{x}, t) + \mathbf{x} f(\mathbf{x}, t)].$$

For the sake of simplicity, we will study separately the decay of the Energy distance on the linear drift equation

$$(4.2) \quad \frac{\partial}{\partial t} f(\mathbf{x}, t) = \nabla \cdot [\mathbf{x} f(\mathbf{x}, t)]$$

and the linear diffusion equation

$$(4.3) \quad \frac{\partial}{\partial t} f(\mathbf{x}, t) = \Delta f(\mathbf{x}, t).$$

Let $0 < \alpha < 2$. Thanks to definition (2.10) we have

$$\begin{aligned}
 (4.4) \quad \frac{d}{dt} E_\alpha(f(t), g(t)) &= -\frac{d}{dt} \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{x}, t) - g(\mathbf{x}, t)][f(\mathbf{y}, t) - g(\mathbf{y}, t)] d\mathbf{x} d\mathbf{y} \\
 &= -\int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{y}, t) - g(\mathbf{y}, t)] \frac{\partial}{\partial t} [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y} \\
 &\quad - \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{x}, t) - g(\mathbf{x}, t)] \frac{\partial}{\partial t} [f(\mathbf{y}, t) - g(\mathbf{y}, t)] d\mathbf{x} d\mathbf{y}.
 \end{aligned}$$

If we apply identity (4.4) to the drift equation (4.2) and integrate by parts, we obtain (4.5)

$$\begin{aligned}
& \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{y}, t) - g(\mathbf{y}, t)] \frac{\partial}{\partial t} [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y} \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{y}, t) - g(\mathbf{y}, t)] \nabla_{\mathbf{x}} \cdot [\mathbf{x}(f(\mathbf{x}, t) - g(\mathbf{x}, t))] d\mathbf{x} d\mathbf{y} \\
&= - \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{x} \cdot \nabla_{\mathbf{x}} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{y}, t) - g(\mathbf{y}, t)] [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y} \\
&= -\alpha \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^{\alpha-2} \sum_{k=1}^n x_k (x_k - y_k) [f(\mathbf{y}, t) - g(\mathbf{y}, t)] [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y}.
\end{aligned}$$

Likewise

$$\begin{aligned}
& \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{x}, t) - g(\mathbf{x}, t)] \frac{\partial}{\partial t} [f(\mathbf{y}, t) - g(\mathbf{y}, t)] d\mathbf{x} d\mathbf{y} \\
&= -\alpha \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^{\alpha-2} \sum_{k=1}^n y_k (y_k - x_k) [f(\mathbf{y}, t) - g(\mathbf{y}, t)] [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y}.
\end{aligned}$$

Therefore, taking the sum of (4.5) and (4.6) we obtain that the drift equation (4.2) satisfies

$$(4.7) \quad \frac{d}{dt} E_\alpha(f(t), g(t)) = -\alpha E_\alpha(f(t), g(t)).$$

Hence, along the drift equation, the Energy distance of order α between two solutions decays exponentially at a rate proportional to its order.

Let us now apply identity (4.4) to the linear diffusion equation (4.3) and integrate twice by parts. We obtain

$$\begin{aligned}
& \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{y}, t) - g(\mathbf{y}, t)] \Delta [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y} \\
(4.8) \quad &= \int_{\mathbb{R}^n \times \mathbb{R}^n} (\Delta_{\mathbf{x}} |\mathbf{x} - \mathbf{y}|^\alpha) [f(\mathbf{y}, t) - g(\mathbf{y}, t)] [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y} \\
&= \alpha(n-2+\alpha) \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|\mathbf{x} - \mathbf{y}|^{2-\alpha}} [f(\mathbf{y}, t) - g(\mathbf{y}, t)] [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y}.
\end{aligned}$$

Likewise

$$\begin{aligned}
& \int_{\mathbb{R}^n \times \mathbb{R}^n} |\mathbf{x} - \mathbf{y}|^\alpha [f(\mathbf{x}, t) - g(\mathbf{x}, t)] \Delta [f(\mathbf{y}, t) - g(\mathbf{y}, t)] d\mathbf{x} d\mathbf{y} \\
(4.9) \quad &= \alpha(n-2+\alpha) \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|\mathbf{x} - \mathbf{y}|^{2-\alpha}} [f(\mathbf{y}, t) - g(\mathbf{y}, t)] [f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y}.
\end{aligned}$$

For $0 < \alpha < 2$, let us denote

$$(4.10) \quad W_\alpha(\mathbf{x}) = |\mathbf{x}|^{-(2-\alpha)}.$$

As shown in Theorem 4.1 from [24], since for $n \geq 2$ one has $2 - \alpha < n$, the Fourier transform of W_α equals

$$(4.11) \quad \widehat{W}_\alpha(\boldsymbol{\xi}) = \int_{\mathbb{R}^n} |\mathbf{x}|^{-(2-\alpha)} e^{-i\boldsymbol{\xi} \cdot \mathbf{x}} d\mathbf{x} = \pi^{n/2} 2^{n-2+\alpha} \frac{\Gamma\left(\frac{n-2+\alpha}{2}\right)}{\Gamma\left(\frac{2-\alpha}{2}\right)} \frac{1}{|\boldsymbol{\xi}|^{n-2+\alpha}}$$

where $\Gamma(\cdot)$ is the gamma function. This implies [1]

$$(4.12) \quad \begin{aligned} 0 \leq E_{-(2-\alpha)}(f(t), g(t)) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|\mathbf{x} - \mathbf{y}|^{2-\alpha}} [f(\mathbf{y}, t) - g(\mathbf{y}, t)][f(\mathbf{x}, t) - g(\mathbf{x}, t)] d\mathbf{x} d\mathbf{y} \\ &= d_{n,\alpha} \int_{\mathbb{R}^n} \frac{|\widehat{f}(\boldsymbol{\xi}, t) - \widehat{g}(\boldsymbol{\xi}, t)|^2}{|\boldsymbol{\xi}|^{n-2+\alpha}} d\boldsymbol{\xi}, \end{aligned}$$

where

$$(4.13) \quad d_{n,\alpha} = \frac{2^\alpha}{2^{2n/2}} \frac{\Gamma\left(\frac{n-2+\alpha}{2}\right)}{\Gamma\left(\frac{2-\alpha}{2}\right)}$$

The Energy distance of negative order $2 - \alpha$, with $0 < \alpha < 2$, defined in (4.12) has been introduced and studied in [1]. It is interesting to remark that, at difference with the Energy distance of positive order, the mixed integral is now dominated by the sum of the other two. Indeed, expression (4.12) can be equivalently expressed as

$$(4.14) \quad \begin{aligned} E_{-(2-\alpha)}(f, g) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|\mathbf{x} - \mathbf{y}|^{2-\alpha}} f(\mathbf{x}) f(\mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &\quad + \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|\mathbf{x} - \mathbf{y}|^{2-\alpha}} g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} - 2 \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|\mathbf{x} - \mathbf{y}|^{2-\alpha}} f(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y}. \end{aligned}$$

Joining the two results, if $f(\mathbf{x}, t)$ solves the Fokker–Planck equation (4.1) and $f_\infty(\mathbf{x})$ denotes the steady solution Gaussian density we proved that

$$(4.15) \quad \frac{d}{dt} E_\alpha(f(t), f_\infty) = -2\alpha(n-2+\alpha) E_{-(2-\alpha)}(f(t), f_\infty) - \alpha E_\alpha(f(t), f_\infty).$$

5. THE ENERGY DISTANCE AND THE LINEAR DIFFUSION EQUATION

It is interesting to remark that, for any given $0 < \alpha < 2$, and $n \geq 2$, given two probability densities $f(\mathbf{x})$ and $g(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, the Energy distance $E_{-(2-\alpha)}(f, g)$, defined as in (4.12), controls the Energy distance $E_\alpha(f, g)$ defined in (2.11). An explicit bound can be found as follows. Let $R > 0$. Then

$$(5.1) \quad \int_{|\boldsymbol{\xi}| \leq R} \frac{|\widehat{f}(\boldsymbol{\xi}) - \widehat{g}(\boldsymbol{\xi})|^2}{|\boldsymbol{\xi}|^{n+\alpha}} d\boldsymbol{\xi} \leq \sup_{\boldsymbol{\xi}} \frac{|\widehat{f}(\boldsymbol{\xi}) - \widehat{g}(\boldsymbol{\xi})|^2}{|\boldsymbol{\xi}|^2} \int_{|\boldsymbol{\xi}| \leq R} \frac{1}{|\boldsymbol{\xi}|^{n+\alpha-2}} d\boldsymbol{\xi}.$$

Thanks to Proposition 2.6 from [6], we know that, for any pair of probability densities $f(\mathbf{x})$ and $g(\mathbf{x})$, the Fourier-based distance

$$d_1(f, g) = \sup_{\boldsymbol{\xi} \in \mathbb{R}^n} \frac{|\widehat{f}(\boldsymbol{\xi}) - \widehat{g}(\boldsymbol{\xi})|}{|\boldsymbol{\xi}|}$$

is bounded. Since $\alpha < 2$, by solving the integral in (5.1) we obtain

$$\int_{|\xi| \leq R} \frac{1}{|\xi|^{n+\alpha-2}} d\xi = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} \int_0^R \rho^{1-\alpha} d\rho = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} \frac{R^{2-\alpha}}{2-\alpha}.$$

Therefore, if $c_{n,\alpha}$ is the constant defined in (2.12),

$$(5.2) \quad c_{n,\alpha} \int_{|\xi| \leq R} \frac{|\widehat{f}(\xi) - \widehat{g}(\xi)|^2}{|\xi|^{n+\alpha}} d\xi \leq c_{n,\alpha} \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} d_1(f, g)^2 \frac{R^{2-\alpha}}{2-\alpha}.$$

Further we have

$$\int_{|\xi| > R} \frac{|\widehat{f}(\xi) - \widehat{g}(\xi)|^2}{|\xi|^{n+\alpha}} d\xi = \int_{|\xi| > R} \frac{1}{|\xi|^2} \frac{|\widehat{f}(\xi) - \widehat{g}(\xi)|^2}{|\xi|^{n+\alpha-2}} d\xi \leq \frac{1}{R^2} \int_{|\xi| > R} \frac{|\widehat{f}(\xi) - \widehat{g}(\xi)|^2}{|\xi|^{n+\alpha-2}} d\xi.$$

Consequently

$$(5.3) \quad c_{n,\alpha} \int_{|\xi| > R} \frac{|\widehat{f}(\xi) - \widehat{g}(\xi)|^2}{|\xi|^{n+\alpha}} d\xi \leq \frac{c_{n,\alpha}}{d_{n,\alpha}} E_{-(2-\alpha)}(f, g) \frac{1}{R^2}.$$

Joining (5.2) and (5.3) we obtain

$$(5.4) \quad E_\alpha(f, g) \leq A_{n,\alpha} d_1(f, g)^2 \frac{R^{2-\alpha}}{2-\alpha} + B_{n,\alpha} E_{-(2-\alpha)}(f, g) \frac{1}{R^2},$$

with obvious meaning of the constants $A_{n,\alpha}$ and $B_{n,\alpha}$. Optimizing over R in (5.4) we finally obtain

$$(5.5) \quad E_\alpha(f, g) \leq D_{n,\alpha} d_1(f, g)^{\frac{4}{4-\alpha}} E_{-(2-\alpha)}(f, g)^{\frac{2-\alpha}{4-\alpha}},$$

where

$$(5.6) \quad D_{n,\alpha} = A_{n,\alpha}^{\frac{2}{4-\alpha}} B_{n,\alpha}^{\frac{2-\alpha}{4-\alpha}} \left[\frac{2^{\frac{2-\alpha}{4-\alpha}}}{2-\alpha} + \left(\frac{1}{2}\right)^{\frac{2}{4-\alpha}} \right].$$

Let $f(t)$ and $g(t)$ be two solutions to equation (4.3) with initial data $f_0(\mathbf{x})$ and $g_0(\mathbf{x})$, both being probability densities on $\mathbb{R}^n$. By examining the Fourier transform formulation of equation (4.3), it is straightforward to verify that the distance $d_1(f(t), g(t))$ is monotone non-increasing in time. Applying inequality (5.5) to the linear diffusion equation (4.3), we then obtain from (4.15) that

$$(5.7) \quad \frac{d}{dt} E_\alpha(f(t), g(t)) = -2\alpha(n-2+\alpha) E_{-(2-\alpha)}(f(t), g(t)) \leq -C E_\alpha(f(t), g(t))^{\frac{4-\alpha}{2-\alpha}}.$$

where

$$(5.8) \quad C = C(\alpha, n, d_1(f_0, g_0)) = 2\alpha(n-2+\alpha) \left[D_{n,\alpha} d_1(f_0, g_0)^{\frac{4}{4-\alpha}} \right]^{-\frac{4-\alpha}{2-\alpha}}.$$

Integrating (5.7) we obtain

$$(5.9) \quad E_\alpha(f(t), g(t)) \leq \left(\frac{1}{E_\alpha(f_0, g_0)^{(2-\alpha)/2}} + C \frac{2}{2-\alpha} t \right)^{-\frac{2-\alpha}{2}}.$$

Hence the Energy distance of order α between two different solutions decays to zero at a rate $t^{-\frac{2-\alpha}{2}}$, a polynomial rate decreasing with respect to α . Clearly, the same result

remains valid if we choose as $g(\mathbf{x}, t)$ the fundamental solution of the heat equation, namely the Gaussian density of zero mean and variance $2t$.

6. CONCLUSIONS

In this paper, we studied the rate of convergence to equilibrium of the solutions to Fokker–Planck type equations with linear drift by means of Cramér and Energy distances. In all cases in which the Fokker–Planck type equations can be treated through these distances, it is shown that the rate of decay is improved with respect to known results which are based on the decay of relative entropy. It is interesting to remark that the study of the decay to equilibrium by means of these distances is not restricted to Fokker–Planck type equations, but it can be usefully applied to systems of Fokker–Planck type equations, as recently done in [19], or to the linear diffusion equation, as briefly shown in Section 5. The present results suggest that both Cramér and Energy distances could be of high interest in the study of both theoretical and numerical problems related to Fokker–Planck type equations.

7. ACKNOWLEDGEMENT

This work has been written within the activities of GNFM (Gruppo Nazionale per la Fisica Matematica) of INdAM (Istituto Nazionale di Alta Matematica), Italy. The results contained in the present paper have been partially presented in WASCOM 2025. The authors state that there is no conflict of interest.

REFERENCES

- [1] Auricchio, G., Brigati, G., Giudici, P., and Toscani, G., From kinetic theory to AI: a rediscovery of high-dimensional divergences and their properties. *Mathematical Models and Methods in Applied Sciences* (in press) (2025)
- [2] Auricchio, G., Codegani, A., Gualandi, S., Toscani, G., and Veneroni, M., On the equivalence between Fourier-based and Wasserstein metrics, *Rendiconti Lincei. Matematica e Applicazioni*, **31** (2020) 627–649.
- [3] Auricchio, G., Codegani, A., Gualandi, S., and Zambon, L., The Fourier discrepancy function, *Communications in Mathematical Sciences*, **21** (2023) 627–639.
- [4] Bellemare, M. G., Danihelka, I., Dabney, W., Mohamed, S., Lakshminarayanan, B., and Hoyer, S., and Munos, R., The Cramer distance as a solution to biased Wasserstein gradients, arXiv preprint arXiv:1705.10743 (2017).
- [5] Bouchaud, J.F., and Mézard, M., Wealth condensation in a simple model of economy *Physica A*, **282** (2000) 536–545.
- [6] Carrillo, J.A. and Toscani, G., Contractive probability metrics and asymptotic behavior of dissipative kinetic equations, *Rivista di Matematica dell’Università di Parma*, **6** (2007) 75–198.
- [7] Carrillo, J.A. and Toscani, G., Renyi entropy and improved equilibration rates to self-similarity for nonlinear diffusion equations, *Nonlinearity*, **27** (2014) 3159–3177.
- [8] Cordier, S., Pareschi, L., and Toscani, G., On a kinetic model for a simple market economy, *J. Statist. Phys.*, **120** (2005) 253–277.
- [9] Cramér, H., On the composition of elementary errors, *Skandinavisk Aktuarietidskrift*, **11**, (1928) 141–180.
- [10] Cuturi, M. and Doucet, A., Fast computation of Wasserstein barycenters, *International conference on machine learning*, (2014) 685–693.
- [11] Dorfman, R., A Formula for the Gini Coefficient, *Review of Economics and Statistics*, **61** (1979) 146–149.

- [12] Furioli, G., Pulvirenti, A., Terraneo, E., and Toscani, G., Fokker–Planck equations in the modelling of socio-economic phenomena, *Math. Models Methods Appl. Sci.* **27** (1) (2017) 115–158.
- [13] Furioli, G., Pulvirenti, A., Terraneo, E., and Toscani, G., Wright-Fisher-type equations for opinion formation, large time behavior and weighted logarithmic-Sobolev inequalities. *Ann. IHP, Analyse Non Linéaire* **36** (2019) 2065–2082.
- [14] Gabetta, E., Toscani, G., and Wennberg, B., Metrics for probability measures and the trend to equilibrium for solutions of the Boltzmann equation, *Journal of Statistical Physics*, **81** (1995) 901–934.
- [15] Gastwirth, J.L., The Estimation of the Lorenz Curve and Gini Index, *Review of Economics and Statistics*, **54** (1972) 306–316.
- [16] Gini, C., Sulla misura della concentrazione e della variabilità dei caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **73** (1914) 1203–1248. English translation in *Metron*, 3–38, (2005).
- [17] Gini, C., Measurement of inequality of incomes, *The Economic Journal*, **31** (1921) 124–126.
- [18] Kendall, M.G., and Stuart, A., *The Advanced Theory of Statistics, Vol. 1, Distribution Theory*, 4th ed. London: Charles Griffin, 1977
- [19] Martalò, G., Toscani, G. and Zanella, M., Individual-based foundation of SIR-type epidemic models: mean-field limit and large time behaviour. (Preprint) (2025).
- [20] Naldi, G., Pareschi, L., and Toscani, G., *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer Verlag, Heidelberg, 2010.
- [21] Pareschi, L., and Toscani, G., *Interacting multiagent systems. Kinetic equations and Monte Carlo methods*. Oxford University Press, Oxford 2013.
- [22] Pareto, V., *Cours d’Économie Politique*, Lausanne and Paris (1897).
- [23] Scagliotti, A. and Farinelli, S., *Normalizing flows as approximations of optimal transport maps via linear-control neural ODEs*, *Nonlinear Analysis*, Elsevier, **257** (2025), 113811.
- [24] Stein, E.M., and Weiss, G., *Introduction to Fourier analysis on Euclidean spaces*, Princeton University Press, Princeton, NJ (1970)
- [25] Székely, G.J., Potential and Kinetic Energy in Statistics. Lecture Notes. Budapest Institute of Technology (Technical University) (1989).
- [26] Székely, G.J., E-statistics: Energy of Statistical Samples. Bowling Green State University, Department of Mathematics and Statistics Technical Report No. 03-05, 2003 (also technical reports by same title, from 2000–2003 and NSA Grant # MDA 904-02-1-0091 (2000–2002).)
- [27] Székely, G.J. and Rizzo, M.I., Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, **143** (2013) 1249–1272
- [28] Torregrossa, M. and Toscani, G., On a Fokker-Planck equation for wealth distribution, *Kinet. Relat. Models* **11** (2) (2018) 337–355.
- [29] Torregrossa, M. and Toscani, G., Wealth distribution in presence of debts. A Fokker-Planck description, *Communications in Mathematical Sciences*, **16** (2018) 537–560.
- [30] G. Toscani, Kinetic models of opinion formation. *Commun. Math. Sci.* **4** (2006) 481–496.
- [31] Xu, K., How has the literature on Gini’s index evolved in the past 80 years? *Dalhousie University, Economics Working Paper* (2003).
- [32] Villani, C., *Optimal transport: old and new*, Vol. 338, (Springer-Verlag, Basel, 2009).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PADUA, VIA TRIESTE 36, PADOVA, 35100 ITALY

Email address: gennaro.auricchio@unipd.it

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PAVIA, AND IMATI OF THE NATIONAL COUNCIL FOR RESEARCH; VIA FERRATA 1, PAVIA, 27100 ITALY

Email address: giuseppe.toscani@unipv.it