



# A note on Fisher information hypocoercive decay for the linear Boltzmann equation

Pierre Monmarché<sup>1</sup>

Received: 18 June 2020 / Revised: 27 October 2020 / Accepted: 4 November 2020 /  
Published online: 17 November 2020  
© Springer Nature Switzerland AG 2020

## Abstract

This note deals with the linear Boltzmann equation in the non-compact setting with a confining potential which is close to quadratic. We prove that in this situation, starting from a smooth initial datum, the Fisher Information (and hence, the relative entropy) with respect to the stationary state converges exponentially fast to zero.

**Keywords** Hypocoercivity · Linear Boltzmann equation · Fisher information · Randomized HMC

**Mathematics Subject Classification** 35K99 · 60J25

## 1 Introduction

We are interested in the long-time convergence to equilibrium of the solution  $f$  of the so-called linear Boltzmann (or BGK) equation

$$\partial_t f_t(x, y) = -v \cdot \nabla_x f_t(x, y) + \nabla U(x) \cdot \nabla_y f_t(x, y) + \lambda Q f_t(x, y) \quad (1)$$

where  $(x, y) \in \mathbb{R}^{2d}$ ,  $d \in \mathbb{N}_*$ ,  $\lambda > 0$  is constant,  $U \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$  and  $Q$  is either  $Q_1$  or  $Q_2$  with

---

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

---

Pierre Monmarché acknowledges support from the French ANR-17-CE40-0030 - EFI - Entropy, flows, inequalities.

---

✉ Pierre Monmarché  
pierre.monmarche@sorbonne-universite.fr

<sup>1</sup> Sorbonne Université, LJLL and LCT, 4 place Jussieu, 75 005 Paris, France

$$Q_1 f(x, y) = \gamma_{d,\sigma}(y) \int_{\mathbb{R}^d} f(x, v) dv - f(x, y)$$

$$Q_2 f(x, y) = \sum_{k=1}^d \left( \gamma_{1,\sigma}(y_k) \int_{\mathbb{R}} f(x, y_1, \dots, y_{k-1}, w, y_{k+1}, \dots, y_d) dw - f(x, y) \right)$$

with, for some  $\sigma > 0$ ,

$$\gamma_{p,\sigma}(y) = \frac{e^{-\frac{1}{2\sigma^2}|y|^2}}{(2\pi\sigma^2)^{p/2}}$$

a Gaussian measure on  $\mathbb{R}^p$ . We assume that  $f_0$  is a probability density so that, mass and positivity being conserved through time,  $f_t$  is a probability density for all  $t \geq 0$ . Denoting  $H(x, y) = U(x) + |y|^2/2$ , we suppose that  $\exp(-H/\sigma^2)$  is integrable and we denote by  $\mu$  the probability law with density proportional to it (we also write  $\mu$  this density). Then  $\mu$  is a fixed point of (1). Our goal is to give a quantitative estimate for the convergence of a solution of (1) toward  $\mu$ . In fact we will rather work with the relative density  $h_t = f_t/\mu$ , which solves

$$\partial_t h_t = Lh_t \tag{2}$$

with

$$Lh(x, y) = -v \cdot \nabla_x h(x, y) + \nabla U(x) \cdot \nabla_y h(x, y) + \eta(P_h - h),$$

where  $(P, \eta)$  is either  $(P_1, \eta_1)$  or  $(P_2, \eta_2)$  with  $\eta_1 = \lambda, \eta_2 = \lambda d$  and

$$P_1 h(x, y) = \int_{\mathbb{R}^d} h(x, v) \gamma_{d,\sigma}(v) dv$$

$$P_2 h(x, y) = \frac{1}{d} \sum_{k=1}^d \int_{\mathbb{R}} h(x, y_1, \dots, y_{k-1}, w, y_{k+1}, \dots, y_d) \gamma_{1,\sigma}(w) dw.$$

Remark that  $P_1$  and  $P_2$  are Markov operators.

Equation (1) is a classical model in statistical physics, modelling the motion of a particle influenced by an external potential  $U$  and by random collisions with other particles with Gaussian velocities. We refer the interested reader to [12] and references within for details. Moreover, it intervenes in Markov Chain Monte Carlo methods. More precisely, denote  $L^*$  the dual of  $L$  in  $L^2(\mu)$ . Integrating by parts, we see that

$$L^* \varphi(x, y) = v \cdot \nabla_x \varphi(x, y) - \nabla U(x) \cdot \nabla_y \varphi(x, y) + \eta(P\varphi - \varphi). \tag{3}$$

This is the generator of a Markov process  $(X, Y)$  whose law solves (1). When  $Q = Q_1$ , the dynamics of the process is the following: the particle follows the Hamiltonian flow  $\dot{x} = y, \dot{y} = -\nabla_x U(x)$  and, at random times with exponential law of intensity  $\lambda$ ,

the velocity  $y$  is refreshed to a new Gaussian value. The motion is similar when  $Q = Q_2$ , except that each coordinate of the velocity has its own exponential clock, and is refreshed to a new Gaussian value independently from the other components. This process is sometimes called the randomized Hamiltonian Monte Carlo process [6]. Since its law converges to  $\mu$ , ergodic averages of the process can be used as estimators for the expectations of some observables with respect to  $\mu$ . A non-asymptotic, quantitative long-time convergence estimate for (2) then classically provides bounds on the bias and variance of such estimators.

The question of the long-time convergence of (1) [or equivalently (2)] has been studied in much general forms in a number of works (see e.g. [12] and references within). The exponential convergence in the  $L^2$  sense, i.e. the existence of constants  $C, \theta > 0$  such that

$$\|h_t - 1\|_{L^2(\mu)}^2 \leq C e^{-\theta t} \|h_0 - 1\|_{L^2(\mu)}^2,$$

has been established under different assumptions by several authors [1,8,12,13]. This long-time convergence is said to be hypocoercive [8,18], in the sense that  $C$  is necessarily greater than 1 or, in other words,  $h_t$  converges exponentially fast to 0 but not at a constant rate (note that both the  $L^2$  norm and the relative entropy studied below are non-increasing with time).

When one studies a system of  $N$  particles with chain or mean-field interactions (so that  $d = Nd'$ , where  $d'$  is the dimension of the ambient space), the  $L^2$ -norm is not well-adapted, since it scales badly in  $N$ . In these contexts, a more natural way to quantify the distance to equilibrium is the relative entropy  $\int h \ln h d\mu$ , as in [10,15,16]. Nevertheless, entropic hypocoercivity results (see e.g. [17–19]) are usually restricted to diffusion processes (i.e. differential operators). Indeed, since non-local operators such as  $\eta(P - I)$  do not satisfy the chain rule, it is less easy to handle derivatives of non-quadratic quantities of  $h$  and  $\nabla h$ . This is a general and important problem, related to the question of giving good definitions of non-local Fisher Information.

Nevertheless, for the linear Boltzmann equation, this has been achieved by Evans [9] in a recent paper in the case of the periodic torus (namely  $x \in \mathbb{T}^d$ ,  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ ) with no potential ( $U = 0$ ). The purpose of the present note is to show that the computations of Evans, together with the recent results on generalized Ornstein-Uhlenbeck processes ([2,3,17]), allows in fact to deal with the case where  $x \in \mathbb{R}^d$  and  $U$  is close to being quadratic.

**Assumption 1** There exist  $K, \kappa > 0$  such that  $\kappa \leq \nabla^2 U(x) \leq K$  for all  $x \in \mathbb{R}^d$ .

In fact, we won't deal with the entropy itself, but with the classical Fisher Information

$$\mathcal{I}(h) = \int \frac{|\nabla h|^2}{h} d\mu.$$

Under Assumption 1, the Hamiltonian  $H$  is strictly convex, so that, by classical arguments (see e.g. [4]),  $\mu$  satisfies a log-Sobolev inequality

$$\int h \ln h d\mu \leq c\mathcal{I}(h),$$

where the constant  $c$  only depends on  $\kappa$  and  $\sigma$ . For elliptic or hypoelliptic diffusions, such as the kinetic Langevin (or Fokker-Planck) equation, a short-time regularization occurs, so that the Fisher Information is finite for all positive times given that the initial entropy is finite (see for instance [17, Theorem 9]). However, this is not the case for equation (2), and thus we will only consider smooth initial datum with  $\mathcal{I}(h_0) < \infty$ . More precisely, for the sake of simplicity, we will assume that, for some  $\varepsilon \geq 0$ ,

$$h_0 \in \mathcal{A}_\varepsilon := \left\{ g \in C^\infty(\mathbb{R}^{2d}), \int g d\mu = 1, g \geq \varepsilon, \partial^\alpha g \text{ bounded } \forall \alpha \text{ multi-index} \right\}.$$

Note that, for any  $\varepsilon \geq 0$ , the set  $\mathcal{A}_\varepsilon$  is fixed by Equation (2), as proved in [7, Appendix] (here we don't need uniform in time estimates for the bounds of the derivatives). The result can then be extended by a density argument to all positive  $h_0$  with  $\mathcal{I}(h_0) < +\infty$ .

**Theorem 1** *Under Assumption 1 let  $(h_t)_{t \geq 0}$  be a solution of (2) with  $h_0 \in \mathcal{A}_0$  such that  $\mathcal{I}(h_0) < +\infty$ . Let*

$$\theta = \frac{2}{3} \left( \frac{2\lambda K}{4K + \lambda^2} - \frac{(K - \kappa)^2}{\sqrt{K}} \right), \quad C = 3 \max \left( \frac{4K^2}{4K + \lambda^2}, \frac{4K + \lambda^2}{4K^2} \right).$$

Suppose that  $\theta \geq 0$ . Then, for all  $t \geq 0$ ,

$$\mathcal{I}(h_t) \leq C e^{-\theta t} \mathcal{I}(h_0).$$

**Remarks**

- The result is the same for  $Q = Q_1$  or  $Q_2$ , and  $C$  and  $\theta$  do not depend on  $\sigma$ .
- The log-Sobolev inequality satisfied by  $\mu$  and Theorem 1 imply that, for some  $C' > 0$ ,

$$\int h_t \ln h_t d\mu \leq C' e^{-\theta t} \mathcal{I}(h_0).$$

By the Pinsker's inequality, considering the Markov process  $(X, Y)$  with generator  $L^*$  given by (3) and initial law  $f_0$ , we get that for all measurable set  $A \subset \mathbb{R}^{2d}$ ,

$$|\mathbb{P}((X_t, Y_t) \in A) - \mu(A)| \leq \frac{1}{2} \|f_t - \mu\|_1 \leq \sqrt{\frac{1}{2} \int h_t \ln h_t d\mu}.$$

This gives a bound on the bias of the Monte Carlo estimator of  $\mu(A)$  based on the process  $(X, Y)$ , with constants  $C'$  and  $\theta$  that depends (explicitly) only on  $\kappa, K, \gamma, \sigma$ .

- The assumption that the potential is convex is usual in the studies of the long-time behaviour of Markov processes. The fact that its Hessian is also bounded above, and more precisely that the Hessian is not too far from a constant matrix, is a much more rigid assumption, which already appeared in similar works [2,5]. Besides, there are

examples of kinetic processes with a convex potential with an unbounded Hessian, which does not converge exponentially fast to their equilibrium [11]. Essentially, we are able to deal with the Gaussian case because of some nice algebra, and have some room for a small perturbation. More precisely, in the Gaussian case, the Jacobian of the drift is a constant matrix, so that the question of the contraction of a suitably modified Fisher Information boils down to a linear algebra problem (see Proposition 5 below). Then a Lipschitz perturbation from this linear case can be absorbed by the positive contraction of the linear case (see the end of the proof of Theorem 1).

- The rate of convergence is of order  $\lambda$  when  $\lambda$  goes to zero and of order  $\lambda^{-1}$  when  $\lambda$  goes to infinity, which is similar to the kinetic Langevin case ([14]), and expected. Indeed, when  $\lambda$  is small, the typical time for the velocity to be refreshed (and thus, to mix) is  $\lambda^{-1}$ . On the other hand, when  $\lambda$  is large, in a time of order 1, there are many jumps, and by the law of large number, the effective velocity is close to zero, and the position moves (and thus, mixes) slowly. If time is accelerated by a factor  $\lambda$ , the position then converges to an overdamped Langevin process.
- In this particular close-to-quadratic case, Theorem 1 answers the question raised in [9, Section 1.5].
- Consider the case where  $Q = Q_2$  and  $U(x) = a|x|^2 + \frac{1}{d} \sum_{i,j=1}^d W(x_i - x_j)$  with an even potential  $W$  with bounded Hessian and  $a > 0$ . This corresponds to a mean-field interaction between  $N = d$  particles. Provided  $\|W''\|_\infty$  is sufficiently small with respect to  $a$  and  $\lambda$ , Theorem 1 yields a speed of convergence to equilibrium which is independent from the number of particles. Then, the arguments from [10, 16] may be adapted (the parallel coupling with Wiener processes being replaced by a parallel coupling with Poisson processes) to obtain uniform in time propagation of chaos, and long-time convergence for the non-linear PDE obtained at the limit (note that the latter is not the Boltzmann equation, for which the interaction lies at the level of the collisions rather than of the Hamiltonian).

## 2 Proof

We write  $h_t = e^{tL}h_0$  the solution of (2) with initial condition  $h_0$ . In the rest of the paper, we will always consider  $h \in \mathcal{A}_\varepsilon$  with  $\varepsilon > 0$ . Indeed, suppose that Theorem 1 has been proved for  $h \in \mathcal{A}_\varepsilon$  with any arbitrary  $\varepsilon > 0$ , and consider  $h_0 \in \mathcal{A}_0$ . Set  $h_0^{(\varepsilon)} = (1-\varepsilon)h_0 + \varepsilon$ . Then  $h_t^{(\varepsilon)} := e^{tL}h_0^{(\varepsilon)} = (1-\varepsilon)h_t + \varepsilon$  so that, applying Theorem 1 to  $h_t^{(\varepsilon)}$  and letting  $\varepsilon$  go to 0, the monotone convergence theorem yields the result for  $h_t$ . The restriction to the cases  $\varepsilon > 0$  will ensure that all the forthcoming derivations under the integral sign are correct. In particular,  $\mathcal{I}(h) < +\infty$  for all  $h \in \mathcal{A}_\varepsilon$  for  $\varepsilon > 0$ .

We start with a general computation. Denoting by  $A^T$  the transpose of a matrix (and seeing vectors as column matrices, so that the scalar product between two vectors  $u$  and  $v$  can be denoted by  $u^T v$ ), for a symmetric matrix  $M$ , we write

$$\mathcal{I}_M(h) = \int \frac{(\nabla h)^T M \nabla h}{h} d\mu.$$

Our aim is to construct  $M$  such that  $\partial_t (\mathcal{I}_M (e^{tL}h)) \leq -\theta \mathcal{I}_M (e^{tL}h)$  with  $\theta > 0$ . In the following, in a  $2d \times 2d$  matrix, a  $d \times d$  block equal to  $\alpha I_d$  for some  $\alpha \in \mathbb{R}$  will sometimes be denoted only by  $\alpha$ , and  $N \geq M$  stands for the usual order for symmetric matrices  $N, M$ .

For an operator  $A$ , we write  $(\partial_t)_{|A}$  the derivative at  $t = 0$  along the semi-group  $e^{tA}$ .

**Lemma 2** *Let  $P$  be a Markov operator which fixes  $\mathcal{A}_\varepsilon$ ,  $h \in \mathcal{A}_\varepsilon$  and  $M = R^T R$  be a positive symmetric matrix. Then*

$$(\partial_t)_{|P-I} \mathcal{I}_M (h) \leq -\mathcal{I}_M (h) + \mathcal{I}_M (Ph) .$$

**Proof** The computation is similar to [9, Lemma 3]. Indeed,

$$\begin{aligned} (\partial_t)_{|P-I} \mathcal{I}_M (h) &= \int \frac{2(\nabla h)^T R^T R \nabla (Ph - h)}{h} - \frac{|R \nabla h|^2 (Ph - h)}{h^2} d\mu \\ &= \int -\frac{|R \nabla h|^2}{h} \left(1 + \frac{Ph}{h}\right) + 2 \frac{(\nabla h)^T R^T R \nabla Ph}{h} d\mu \\ &= -\mathcal{I}_M (h) + \lambda \int -\left| \frac{R \nabla h}{h} - \frac{R \nabla Ph}{Ph} \right|^2 Qh + \frac{|R \nabla Ph|^2}{Ph} d\mu \\ &\leq -\mathcal{I}_M (h) + \mathcal{I}_M (Ph) \end{aligned}$$

where we used the positivity of the density  $h$ . □

For  $k \in \llbracket 1, d \rrbracket$ , let  $E_k$  be the  $2d \times 2d$  diagonal matrix with all its coefficients being zero except the  $(d + k, d + k)^{th}$  being equal to 1, and

$$E = \sum_{k=1}^d E_k = \begin{pmatrix} 0 & 0 \\ 0 & I_d \end{pmatrix}, \quad E' = I_{2d} - E = \begin{pmatrix} I_d & 0 \\ 0 & 0 \end{pmatrix} .$$

In the particular case of (2), Lemma 2 yields the following.

**Lemma 3** *Let  $\lambda > 0$ ,  $h \in \mathcal{A}_\varepsilon$  and  $M = R^T R$  be a positive symmetric matrix. Then,*

$$(\partial_t)_{|\eta(P-I)} \mathcal{I}_M (h) \leq -\lambda (\mathcal{I}_{EM+ME-EME}(h)) \tag{4}$$

for  $(P, \eta) = (P_1, \eta_1)$ . Moreover, this is also true for  $(P, \eta) = (P_2, \eta_2)$  if the right down  $d \times d$  corner of  $M$  is an homothety.

**Proof** We recall the following argument from [9, Lemma 1]. From  $\nabla_y P_1 = 0$  and  $\nabla_x P_1 = P_1 \nabla_x$ ,  $\mathcal{I}_M (P_1 h) = \int \phi (P_1 (\nabla h, h)) d\mu$ , where  $\phi(u, v) = (u^T E' M E' u)/v$ . Applying Jensen's Inequality to the convex function  $\phi$  and the Markov operator  $P_1$ , we get  $\phi (P_1 (h, \nabla h)) \leq P_1 \phi (h, \nabla h)$ . Integrated with respect to  $\mu$  (which is fixed by  $P_1$ ), this reads

$$\mathcal{I}_M (P_1 h) \leq \mathcal{I}_{E' M E'} (h) .$$

Applying Lemma 2 yields (4) (since  $\eta_1 = \lambda$ ).

Similarly, denoting

$$P_{2,k}f(x, y) = \int f(x, y_1, \dots, y_{k-1}, w, y_{k+1}, \dots, y_d) \frac{e^{-\frac{1}{2\sigma^2}w^2}}{\sigma\sqrt{2\pi}} dw,$$

for  $k \in \llbracket 1, d \rrbracket$ , we get with the previous argument

$$\mathcal{I}_M(P_{2,k}h) \leq \mathcal{I}_{(I_{2d}-E_k)M(I_{2d}-E_k)}(h),$$

so that

$$\begin{aligned} (\partial_t)_{|\eta_2(P_2-I)} \mathcal{I}_M(h) &= \sum_{k=1}^d (\partial_t)_{|\lambda(P_{2,k}-I)} \mathcal{I}_M(h) \\ &\leq -\lambda \left( \mathcal{I}_{\sum_{k=1}^d (E_k M + M E_k - E_k M E_k)}(h) \right). \end{aligned}$$

Now, suppose that the right down  $d \times d$  corner of  $M$  is an homothety, i.e. that

$$M = \begin{pmatrix} M_1 & M_2 \\ M_2^T & \alpha I_d \end{pmatrix}$$

for some matrices  $M_i$  and some  $\alpha > 0$ . In that case,

$$EME = \alpha \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \alpha \sum_{k=1}^d E_k E_k = \sum_{k=1}^d E_k M E_k,$$

which means that we have obtained the same bound (4) on  $(\partial_t)_{|\eta_i(P_i-I)} \mathcal{I}_M(h)$  for both  $i = 1, 2$ . □

On the other hand, the derivative of  $\mathcal{I}_M$  along the transport semi-group  $e^{tA}$  where  $A = y \cdot \nabla_x - \nabla_x U(x) \cdot \nabla_y$  is a classical computation (see e.g. [17, Example 8]), which we recall for the sake of completeness:

**Lemma 4** For  $h \in \mathcal{A}_\varepsilon$ ,

$$(\partial_t)_{|A} \mathcal{I}_M(h) = \int \frac{(\nabla h)^T (MJ + J^T M) \nabla h}{h} d\mu \tag{5}$$

with  $J = \begin{pmatrix} 0 & -\nabla^2 U \\ 1 & 0 \end{pmatrix}$ .

**Proof** Since  $A$  satisfies the chain rule,

$$\begin{aligned} (\partial_t)_{|A} \mathcal{I}_M(h) &= \int \left( \frac{2(\nabla h)^T M \nabla A h}{h} - \frac{A h}{h^2} (\nabla h)^T M \nabla h \right) d\mu \\ &= \int \left( A \left( \frac{2(\nabla h)^T M \nabla h}{h} \right) + \frac{2(\nabla h)^T M (\nabla A h - A \nabla h)}{h} \right) d\mu \end{aligned}$$

where  $A \nabla h$  should be understood coordinate by coordinate. Conclusion follows from  $\int A g d\mu = 0$  for all  $g$  and  $\nabla A h - A \nabla h = J \nabla h$ .  $\square$

Combining the two previous results, we get:

**Proposition 5** *Under Assumption 1, let  $h_t = e^{tL} h_0$  where  $h_0 \in \mathcal{A}_\varepsilon$ . Suppose that there exist  $a, b, \theta \in \mathbb{R}$  with  $b^2 < a$  and such that, for  $\xi \in \{\kappa, K\}$ ,*

$$N(\xi) := \begin{pmatrix} 2b & a - \xi + \lambda b \\ a - \xi + \lambda b & -2b\xi + \lambda a \end{pmatrix} \geq \theta \begin{pmatrix} 1 & b \\ b & a \end{pmatrix} := \theta M. \tag{6}$$

Then, for all  $t \geq 0$ ,

$$\mathcal{I}_M(h_t) \leq e^{-\theta t} \mathcal{I}_M(h_0).$$

**Proof** Let  $a, b, \theta \in \mathbb{R}$  and  $M$  be as in the proposition. Since  $L = -A + \eta(P - I)$ , Lemmas 3 and 4 read

$$\partial_t \mathcal{I}_M(h_t) = (\partial_t)_{|\eta(P-I)} \mathcal{I}_M(h_t) - (\partial_t)_{|A} \mathcal{I}_M(h) \leq -\mathcal{I}_{N'}(h)$$

with, for all  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} N'(x) &= M \begin{pmatrix} 0 & -\nabla^2 U(x) \\ 1 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -\nabla^2 U(x) & \lambda \end{pmatrix} M - \lambda \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix} \\ &= \begin{pmatrix} 2b & a - \nabla^2 U(x) + \lambda b \\ a - \nabla^2 U(x) + \lambda b & -2b \nabla^2 U(x) + \lambda a \end{pmatrix}. \end{aligned}$$

The proof will be concluded (by the Gronwall's Lemma) if we prove that  $N'(x) \geq \theta M$  for all  $x \in \mathbb{R}^d$ . Fix  $x \in \mathbb{R}^d$ , and let  $\mathcal{O}(x)$  be an orthonormal  $d \times d$  matrix such that  $\mathcal{O}^T(x) \nabla^2 U(x) \mathcal{O}(x)$  is diagonal. Let

$$\mathcal{O}' = \begin{pmatrix} \mathcal{O}(x) & 0 \\ 0 & \mathcal{O}(x) \end{pmatrix}.$$

Notice that  $N'(x) \geq \theta M$  if and only if  $\mathcal{O}'^T N'(x) \mathcal{O}' \geq \theta M$ , where we used that  $\mathcal{O}'^T M \mathcal{O}' = M$ . Now,  $\mathcal{O}'^T N'(x) \mathcal{O}' \geq \theta M$  if and only if  $N(\xi_k) \geq \theta M$  for all eigenvalues  $\xi_k$  of  $\nabla^2 U(x)$ ,  $k \in \llbracket 1, d \rrbracket$ . Writing such an eigenvalue as  $\xi_k = p_k \kappa + (1 - p_k) K$  for some  $p_k \in [0, 1]$ , we get

$$N(\xi_k) = p_k N(\kappa) + (1 - p_k) N(K) \geq \theta M$$



by assumption, which concludes. □

With Proposition 5 in hand, the proof of our main result has boiled down to elementary computations.

**Proof of Theorem 1** Let us find  $a$  and  $b$  such that  $N(\xi)$  as given by (6) is definite positive for a given  $\xi$  (to be chosen later on). For simplicity, we want to enforce the following conditions:

$$4b^2 \leq a, \quad a + \lambda b = \xi, \quad \lambda a \geq 4b\xi,$$

which ensures that  $N(\xi)$  is diagonal with positive terms and that the corresponding  $M$  is definite positive. It is clear that such conditions are satisfied for  $b$  small enough with  $a = \xi - \lambda b$ . More precisely, the first condition is implied by the third if  $b \leq \xi/\lambda$ , and the third is implied by the second if  $b \leq \lambda\xi/(4\xi + \lambda^2)$  (notice that  $\lambda\xi/(4\xi + \lambda^2) \leq \xi/\lambda$ ). As a consequence, we chose

$$b = \frac{\lambda\xi}{4\xi + \lambda^2}, \quad a = \xi - \lambda b = \frac{4\xi^2}{4\xi + \lambda^2}.$$

The condition  $4b^2 \leq a$  is such that the corresponding matrix  $M$  satisfies

$$\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \leq M \leq \frac{3}{2} \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}.$$

The choice of  $a$  and  $b$  also ensures that

$$N(\xi) = \begin{pmatrix} 2b & 0 \\ 0 & -2b\xi + \lambda a \end{pmatrix} \geq \begin{pmatrix} 2b & 0 \\ 0 & \frac{\lambda a}{2} \end{pmatrix}.$$

For  $\xi' \in \{\kappa, K\}$ , for all  $\gamma > 0$ ,

$$N(\xi') = N(\xi) + \begin{pmatrix} 0 & \xi - \xi' \\ \xi - \xi' & 2b(\xi - \xi') \end{pmatrix} \geq \begin{pmatrix} 2b - \gamma(\xi - \xi')^2 & 0 \\ 0 & \frac{\lambda a}{2} + 2b(\xi - \xi') - \frac{1}{\gamma}(\xi - \xi')^2 \end{pmatrix}.$$

In other words,

$$N(\xi') \geq \theta_1 \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$$

with

$$\theta_1 = \min \left( 2b \left( 1 - \frac{\gamma}{2b} (\xi - \xi')^2 \right), \frac{\lambda}{2} \left( 1 - \frac{2}{\lambda a \gamma} (\xi - \xi')^2 + \frac{\xi - \xi'}{\xi} \right) \right).$$

Using that  $2b \leq \lambda/2$ , we chose  $\gamma^2 = 4b/(\lambda a) = 1/\xi$  to get that

$$\theta_1 \geq \theta_2 := 2b \left( 1 - \frac{1}{2b\sqrt{\xi}}(\xi - \xi')^2 + \frac{\xi - \xi'}{\xi} \right).$$

Finally, we simply chose  $\xi = K$ , so that  $\xi - \xi' \geq 0$  for  $\xi' \in \{\kappa, K\}$ . Assuming that  $(K - \kappa)^2 \leq 2b\sqrt{\xi}$ , we get that  $\theta_2 \geq 0$  for both  $\xi' \in \{\kappa, K\}$ , and thus

$$N(\xi') \geq \theta_2 \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \geq \frac{2\theta_2}{3} M = \theta M,$$

and we conclude by

$$\mathcal{I}(h_t) \leq \frac{2}{\min(1, a)} \mathcal{I}_M(h_t) \leq \frac{2}{\min(1, a)} e^{-\theta t} \mathcal{I}_M(h_0) \leq \frac{3 \max(1, a)}{\min(1, a)} e^{-\theta t} \mathcal{I}(h_0).$$

□

**Acknowledgements** The author would like to thank Stefano Olla for introducing him to this question, and Max Fathi for fruitful discussions.

## Compliance with ethical standards

**Conflict of interest** The author declare that there is no conflict of interest.

## References

- Achleitner, F., Arnold, A., Carlen, E.A.: On linear hypocoercive BGK models. In: Gonçalves, P., Soares, A.J. (eds.) *From Particle Systems to Partial Differential Equations III*, pp. 1–37. Springer, Cham (2016)
- Achleitner, F., Arnold, A., Stürzer, D.: Large-time behavior in non-symmetric Fokker–Planck equations. *Riv. Math. Univ. Parma (N.S.)* **6**(1), 1–68 (2015)
- Arnold, A., Erb, J.: Sharp entropy decay for hypocoercive and non-symmetric Fokker–Planck equations with linear drift (2014). [arXiv:1409.5425](https://arxiv.org/abs/1409.5425)
- Bakry, D., Gentil, I., Ledoux, M.: Analysis and geometry of Markov diffusion operators, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 348. Springer, Cham (2014). <https://doi.org/10.1007/978-3-319-00227-9>
- Baudoin, F.: Bakry–Émery meet Villani. *J. Funct. Anal.* **273**(7), 2275–2291 (2017). <https://doi.org/10.1016/j.jfa.2017.06.021>
- Bou-Rabee, N., Sanz-Serna, J.: Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27**(4), 2159–2194 (2017). <https://doi.org/10.1214/16-AAP1255>
- Cáceres, M.J., Carrillo, J.A., Goudon, T.: Equilibration rate for the linear inhomogeneous relaxation-time Boltzmann equation for charged particles. *Commun. Partial Differ. Equ.* **28**(5–6), 969–989 (2003). <https://doi.org/10.1081/PDE-120021182>
- Dolbeault, J., Mouhot, C., Schmeiser, C.: Hypocoercivity for linear kinetic equations conserving mass. *Trans. Am. Math. Soc.* **367**(6), 3807–3828 (2015). <https://doi.org/10.1090/S0002-9947-2015-06012-7>
- Evans, J.: Hypocoercivity in Phi-entropy for the Linear Boltzmann equation on the torus (2017). [ArXiv e-prints](https://arxiv.org/abs/1708.02801)
- Guillin, A., Monmarché, P.: Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes (2020). [arXiv e-prints arXiv:2003.00735](https://arxiv.org/abs/2003.00735)
- Hairer, M., Mattingly, J.: Slow energy dissipation in anharmonic oscillator chains. *Commun. Pure Appl. Math.* **62**(8), 999–1032 (2009). <https://doi.org/10.1002/cpa.20280>

12. Han-Kwan, D., Léautaud, M.: Geometric analysis of the linear Boltzmann equation I. Trend to equilibrium. *Ann. PDE* **1**(1), Art. 3, 84 (2015)
13. Hérau, F.: Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation. *Asymptot. Anal.* **46**(3–4), 349–359 (2006)
14. Iacobucci, A., Olla, S., Stoltz, G.: Convergence rates for nonequilibrium Langevin dynamics. *Ann. Math. Québec* **43**, 73–98 (2019)
15. Letizia, V., Olla, S.: Nonequilibrium isothermal transformations in a temperature gradient from a microscopic dynamics. *Ann. Probab.* **45**(6A), 3987–4018 (2017). <https://doi.org/10.1214/16-AOP1156>
16. Monmarché, P.: Long-time behaviour and propagation of chaos for mean field kinetic particles. *Stochastic Process. Their Appl.* **127**, 1721–1737 (2016)
17. Monmarché, P.: Generalized  $\Gamma$  calculus and application to interacting particles on a graph. *Potential Anal.* **50**(3), 439–466 (2019). <https://doi.org/10.1007/s11118-018-9689-3>
18. Villani, C.: Hypocoercivity. *Mem. Am. Math. Soc.* **202**(950), iv+141 (2009). <https://doi.org/10.1090/S0065-9266-09-00567-5>
19. Wang, F.Y.: Hypercontractivity and applications for stochastic hamiltonian systems. *J. Funct. Anal.* **272**(12), 5360–5383 (2017). <https://doi.org/10.1016/j.jfa.2017.03.015>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.