# CONCENTRATION OF MEASURE ON PRODUCT SPACES WITH APPLICATIONS TO MARKOV PROCESSES

GORDON BLOWER AND FRANÇOIS BOLLEY

Abstract. For a stochastic process with state space some Polish space, this paper gives sufficient conditions on the initial and conditional distributions for the joint law to satisfy Gaussian concentration inequalities and transportation inequalities. In the case of the Euclidean space $\mathbb{R}^m$, there are sufficient conditions for the joint law to satisfy a logarithmic Sobolev inequality. In several cases, the obtained constants are of optimal order of growth with respect to the number of random variables, or are independent of this number. These results extend results known for mutually independent random variables to weakly dependent random variables under Dobrushin–Shlosman type conditions. The paper also contains applications to Markov processes including the ARMA process.

## 1. Introduction

Given a complete and separable metric space $(X, d)$, $\mathrm{Prob}(X)$ denotes the space of Radon probability measures on $X$, equipped with the (narrow) weak topology. We say that $\mu \in \mathrm{Prob}\,(X)$ satisfies a *Gaussian concentration inequality* $GC(\kappa)$ with constant $\kappa$ on $(X, d)$ if

$$\int_X \exp\big(tF(x)\big)\mu(dx) \leq \exp\Big(t\int_X F(x)\,\mu(dx) + \kappa\, t^2/2\Big) \qquad (t \in \mathbb{R})$$

holds for all 1-Lipschitz functions $F : (X, d) \to \mathbb{R}$ (see [3]). Recall that a function $g : (\Omega_1, d_1) \to (\Omega_2, d_2)$ between metric spaces is *L-Lipschitz* if $d_2(g(x), g(y)) \leq L\, d_1(x, y)$ holds for all $x, y \in \Omega_1$, and we call the infimum of such $L$ the *Lipschitz seminorm* of $g$.

For $k \geq 1$ and $x_1, \dots, x_k$ in $X$, we let $x^{(k)} = (x_1, \dots, x_k) \in X^k$ and, given $1 \leq s < \infty$, we equip the product space $X^k$ with the metric $d^{(s)}$ defined by $d^{(s)}(x^{(k)}, y^{(k)}) = (\sum_{j=1}^k d(x_j, y_j)^s)^{1/s}$ for $x^{(k)}$ and $y^{(k)}$ in $X^k$.

Now let $(\xi_j)_{j=1}^n$ be a stochastic process with state space $X$. The first aim of this paper is to obtain concentration inequalities for the joint distribution $P^{(n)}$ of $\xi^{(n)} = (\xi_1, \dots, \xi_n)$, under hypotheses on the initial distribution $P^{(1)}$ of $\xi_1$ and the conditional distributions $p_k(. \mid x^{(k-1)})$ of $\xi_k$ given $\xi^{(k-1)}$; we recall that $P^{(n)}$ is given by

$$P^{(n)}(dx^{(n)}) = p_n(dx_n \mid x^{(n-1)}) \dots p_2(dx_2 \mid x_1)\, P^{(1)}(dx_1).$$

*Key words and phrases.* Logarithmic Sobolev inequality, optimal transportation.
2000 Mathematics Subject Classification : 60E15 (60E05 39B62).
16th May 2005.

If the $(\xi_j)_{j=1}^n$ are mutually independent, and the distribution of each $\xi_j$ satisfies $GC(\kappa)$, then $P^{(n)}$ on $(X^n, d^{(1)})$ is the product of the marginal distributions, and inherits $GC(n\kappa)$ from its marginal distributions by a simple 'tensorization' argument. A similar result also applies to product measures for the transportation and logarithmic Sobolev inequalities which we consider later; see [12, 23]. To obtain concentration inequalities for $P^{(n)}$ when $(\xi_j)$ are weakly dependent, we impose additional restrictions on the coupling between the variables, expressed in terms of Wasserstein distances which are defined as follows.

Given $1 \leq s < \infty$, $\mathrm{Prob}_s(X)$ denotes the subspace of $\mathrm{Prob}(X)$ consisting of $\nu$ such that $\int_X d(x_0, y)^s \, \nu(dy)$ is finite for some or equivalently all $x_0 \in X$. Then we define the *Wasserstein distance* of order $s$ between $\mu$ and $\nu$ in $\mathrm{Prob}_s(X)$ by

$$W_s(\mu, \nu) = \inf_\pi \Big( \iint_{X \times X} d(x, y)^s \, \pi(dx\, dy) \Big)^{1/s} \tag{1.1}$$

where $\pi \in \mathrm{Prob}_s\,(X \times X)$ has marginals $\pi_1 = \mu$ and $\pi_2 = \nu$. Then $W_s$ defines a metric on $\mathrm{Prob}_s(X)$, which in turn becomes a complete and separable metric space (see [20, 24]).

In section 3 we obtain the following result for time-homogeneous Markov chains.

**Theorem 1.1.** *Let $(\xi_j)_{j=1}^n$ be an homogeneous Markov process with state space $X$, initial distribution $P^{(1)}$ and transition measure $p(.\mid x)$. Suppose that there exist constants $\kappa_1$ and $L$ such that:*

*(i) $P^{(1)}$ and $p(.\mid x)$ $(x \in X)$ satisfy $GC(\kappa_1)$ on $(X, d)$;*

*(ii) $x \mapsto p(.\mid x)$ is $L$-Lipschitz $(X, d) \to (Prob_1\,(X), W_1)$.*

*Then the joint law $P^{(n)}$ of $(\xi_1, \ldots, \xi_n)$ satisfies $GC(\kappa_n)$ on $(X^n, d^{(1)})$, where*

$$\kappa_n = \kappa_1 \sum_{m=1}^n \Big( \sum_{k=0}^{m-1} L^k \Big)^2.$$

In Example 6.3 we demonstrate sharpness of these constants by providing for each value of $L$ a process such that $\kappa_n$ has optimal growth in $n$.

Concentration inequalities are an instance of the wider class of transportation inequalities, which bound the transportation cost by the relative entropy. We recall the definitions.

Let $\nu$ and $\mu$ be in $\mathrm{Prob}(X)$, where $\nu$ is absolutely continuous with respect to $\mu$, and let $d\nu/d\mu$ be the Radon–Nikodym derivative. Then we define the *relative entropy* of $\nu$ with respect to $\mu$ by

$$\mathrm{Ent}(\nu \mid \mu) = \int_X \log \frac{d\nu}{d\mu} \, d\nu\,;$$

note that $0 \leq \mathrm{Ent}(\nu \mid \mu) \leq \infty$ by Jensen's inequality. By convention we let $\mathrm{Ent}(\nu \mid \mu) = \infty$ if $\nu$ is not absolutely continuous with respect to $\mu$.

Given $1 \leq s < \infty$, we say that $\mu \in \mathrm{Prob}_s(X)$ satisfies a *transportation inequality* $T_s(\alpha)$ for cost function $d(x, y)^s$, with constant $\alpha$, if

$$W_s(\nu, \mu) \leq \Big( \frac{2}{\alpha} \mathrm{Ent}(\nu \mid \mu) \Big)^{1/2}$$

for all $\nu \in \mathrm{Prob}_s(X)$.

Marton [13] introduced $T_2$ as 'distance-divergence' inequalities in the context of information theory; subsequently Talagrand [23] showed that the standard Gaussian distribution on $\mathbb{R}^m$ satisfies $T_2(1)$. Bobkov and Götze showed in [3] that $GC(\kappa)$ is equivalent to $T_1(1/\kappa)$; their proof used the Kantorovich–Rubinstein duality result, that

$$W_1(\mu,\nu) = \sup_f \Big\{ \int_X f(x)\mu(dx) - \int_X f(y)\,\nu(dy) \Big\}$$

where $\mu,\nu \in \mathrm{Prob}_1(X)$ and $f$ runs over the set of 1-Lipschitz functions $f : X \to \mathbb{R}$. A $\nu \in \mathrm{Prob}(X)$ satisfies a $T_1$ inequality if and only if $\nu$ admits a square-exponential moment; that is, $\int_X \exp(\beta d(x,y)^2)\,\nu(dx)$ is finite for some $\beta > 0$ and some, and thus all, $y \in X$; see [5, 9] for detailed statements. Moreover, since $T_s(\alpha)$ implies $T_r(\alpha)$ for $1 \le r \le s$ by Hölder's inequality, transportation inequalities are a tool for proving and strengthening concentration inequalities; they are also related to the Gaussian isoperimetric inequality as in [2]. For applications to empirical distributions in statistics, see [16].

Returning to weakly dependent $(\xi_j)_{j=1}^n$ with state space $X$, we obtain transportation inequalities for the joint distribution $P^{(n)}$, under hypotheses on $P^{(1)}$ and the conditional distributions. Djellout, Guillin and Wu [9] developed Marton's coupling method [13, 15] to prove $T_s(\alpha)$ for $P^{(n)}$ under various mixing or contractivity conditions; see also [22], or [5] where the conditions are expressed solely in terms of exponential moments. We extend these results in sections 2 and 3 below, thus obtaining a strengthened dual form of Theorem 1.1.

**Theorem 1.2.** *Let $(\xi_j)_{j=1}^n$ be an homogeneous Markov process with state space $X$, initial distribution $P^{(1)}$ and transition measure $p(.\mid x)$. Suppose that there exist constants $1 \le s \le 2$, $\alpha > 0$ and $L \ge 0$ such that:*
*(i) $P^{(1)}$ and $p(.\mid x)$ $(x \in X)$ satisfy $T_s(\alpha)$;*
*(ii) $x \mapsto p(.\mid x)$ is L-Lipschitz $(X,d) \to (Prob_s(X), W_s)$.*
*Then the joint distribution $P^{(n)}$ of $(\xi_1,\dots,\xi_n)$ satisfies $T_s(\alpha_n)$, where*

$$\alpha_n = \begin{cases} n^{1-(2/s)}(1-L^{1/s})^2\alpha & \text{if } L<1, \\ e^{(2/s)-2}(n^{-(2/s)-1}\alpha & \text{if } L=1, \\ \Big(\dfrac{L-1}{e^{s-1}L^n}\Big)^{2/s}\dfrac{\alpha}{n+1} & \text{if } L>1; \end{cases}$$

*in particular $\alpha_n$ is independent of $n$ for $s=2$ when $L<1$.*

Our general transportation Theorem 2.1 will involve processes that are not necessarily Markovian, but satisfy some a hypothesis related to Dobrushin–Shlosman's mixing condition [8, p. 352; 15, Definition 2]. When $X = \mathbb{R}^m$, we shall also present some more computable version of hypothesis (ii) in Proposition 2.2, and later consider a stronger functional inequality.

A probability measure $\mu$ on $\mathbb{R}^m$ satisfies the *logarithmic Sobolev inequality* $LSI(\alpha)$ with constant $\alpha > 0$ if

$$\int_{\mathbb{R}^m} f^2 \log\Big(f^2 / \int_{\mathbb{R}^m} f^2 \, d\mu\Big) d\mu \leq \frac{2}{\alpha} \int_{\mathbb{R}^m} \|\nabla f\|^2_{\ell^2} \, d\mu$$

holds for all $f \in L^2(d\mu)$ that have distributional gradient $\nabla f \in L^2(d\mu; \mathbb{R}^m)$. Given $(a_k) \in \mathbb{R}^m$, let $\|(a_k)\|_{\ell^s} = (\sum_{k=1}^m |a_k|^s)^{1/s}$ for $1 \leq s < \infty$, and $\|(a_k)\|_{\ell^\infty} = \sup_{1\leq k \leq m} |a_k|$.

The connection between the various inequalities is summarized by

$$LSI(\alpha) \Rightarrow T_2(\alpha) \Rightarrow T_1(\alpha) \Leftrightarrow GC(1/\alpha); \tag{1.2}$$

see [3; 18; 24, p. 293]. Conversely, Otto and Villani showed that if $\mu(dx) = e^{-V(x)}\, dx$ satisfies $T_2(\alpha)$ where $V : \mathbb{R}^m \to \mathbb{R}$ is convex, then $\mu$ also satisfies $LSI(\alpha/4)$ (see [4; 18; 24, p. 298]); but this converse is not generally true, as a counter-example in [6] shows.

Gross [11] proved that the standard Gaussian probability measure on $\mathbb{R}^m$ satisfies $LSI(1)$. More generally, Bakry and Emery [1] showed that if $V$ is twice continuously differentiable, with $\operatorname{Hess} V \geq \alpha I_m$ on $\mathbb{R}^m$ for some $\alpha > 0$, then $\mu(dx) = e^{-V(x)}\, dx$ satisfies $LSI(\alpha)$; see for instance [25] for extensions to this result. Whereas Bobkov and Götze [3] characterized in terms of their cumulative distribution functions those $\mu \in \operatorname{Prob}(\mathbb{R})$ that satisfy $LSI(\alpha)$ for some $\alpha$, there is no known geometrical characterization of such probability measures on $\mathbb{R}^m$ when $m > 1$.

Our main Theorem 5.1 gives a sufficient condition for the joint law of a weakly dependent process with state space $\mathbb{R}^m$ to satisfy $LSI$. In section 6 we deduce the following for distributions of time-homogeneous Markov processes. Let $\partial/\partial x$ denote the gradient with respect to $x \in \mathbb{R}^m$.

**Theorem 1.3.** *Let $(\xi_j)_{j=1}^n$ be an homogeneous Markov process with state space $\mathbb{R}^m$, initial distribution $P^{(1)}$ and transition measure $p(dy \mid x) = e^{-u(x,y)}dy$. Suppose that there exist constants $\alpha > 0$ and $L \geq 0$ such that:*

*(i) $P^{(1)}$ and $p(. \mid x)$ $(x \in \mathbb{R}^m)$ satisfy $LSI(\alpha)$;*

*(ii) $u$ is twice continuously differentiable and the off-diagonal blocks of its Hessian matrix satisfy*

$$\Big\| \frac{\partial^2 u}{\partial x \partial y} \Big\| \leq L$$

*as operators $(\mathbb{R}^m, \ell^2) \to (\mathbb{R}^m, \ell^2)$.*

*Then the joint law $P^{(n)}$ of the first $n$ variables $(\xi_1, \ldots, \xi_n)$ satisfies $LSI(\alpha_n)$, where*

$$\alpha_n = \begin{cases} \dfrac{(\alpha - L)^2}{\alpha} & \text{if } L < \alpha, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if } L = \alpha, \\ \Big(\dfrac{\alpha}{L}\Big)^{2n} \dfrac{L^2 - \alpha^2}{\alpha e(n+1)} & \text{if } L > \alpha; \end{cases}$$

*in particular $\alpha_n$ is independent of $n$ when $L < \alpha$.*

The plan of the paper is as follows. In section 2 we state and prove our results on transportation inequalities, which imply Theorem 1.2, and in section 3 we deduce Theorem 1.1. In section 4 we prove $LSI(\alpha)$ for the joint distribution of ARMA processes, with $\alpha$ independent of the size of the sample. In section 5 we obtain a more general $LSI$, which we express in a simplified form for Markov processes in section 6. Explicit examples in section 6 show that several of our results have optimal growth of the constants with respect to $n$ as $n \to \infty$, and that the hypotheses are computable and realistic.

## 2. Transportation inequalities

Let $(\xi_k)_{k=1}^n$ be a stochastic process with state space $X$, let $p_k(. \mid x^{(k-1)})$ denote the transition measure between the states at times $k-1$ and $k$, and let $P^{(n)}$ be the joint distribution of $\xi^{(n)}$. Our main result of this section is a transportation inequality.

**Theorem 2.1.** *Let $1 \leq s \leq 2$, and suppose that there exist $\alpha_1 > 0$ and $M \geq \rho_\ell \geq 0$ $(\ell = 1, \dots, n)$ such that:*

*(i) $P^{(1)}$ and $p_k(. \mid x^{(k-1)})$ $(k = 2, \dots, n;\ x^{(k-1)} \in X^{k-1})$ satisfy $T_s(\alpha)$ on $(X, d)$;*

*(ii) $x^{(k-1)} \mapsto p_k(. \mid x^{(k-1)})$ is Lipschitz as a map $(X^{k-1}, d^{(s)}) \to (Prob_s\,(X), W_s)$ for $k = 2, \dots, n$, in the sense that*

$$W_s\big(p_k(. \mid x^{(k-1)}), p_k(. \mid y^{(k-1)})\big)^s \leq \sum_{j=1}^{k-1} \rho_{k-j}\, d(x_j, y_j)^s \qquad (x^{(k-1)}, y^{(k-1)} \in X^{k-1}).$$

*Then $P^{(n)}$ satisfies the transportation inequality $T_s(\alpha_n)$ where*

$$\alpha_n = \alpha \left( \frac{(n\,e)^{1-s} M}{(1+M)^n} \right)^{2/s}.$$

*Suppose further that*

*(iii) $\sum_{j=1}^n \rho_j \leq R$.*

*Then the joint distribution $P^{(n)}$ satisfies $T_s(\alpha_n)$ where*

$$\alpha_n = \begin{cases} n^{1-(2/s)}(1-R^{1/s})^2 \alpha & \text{if } \ R < 1, \\ e^{(2/s)-2}(n+1)^{-(2/s)-1} & \text{if } \ R = 1, \\ \left( \dfrac{R-1}{e^{s-1} R^n} \right)^{2/s} \dfrac{\alpha}{n+1} & \text{if } \ R > 1. \end{cases}$$

In hypothesis (iii), the sequence $(\rho_k)_{k=1}^{n-1}$ measures the extent to which the distribution of $\xi_n$ depends upon the previous $\xi_{n-1}, \xi_{n-2}, \dots$; so in most examples $(\rho_k)_{k=1}^{n-1}$ is decreasing.

A version of Theorem 2.1 was obtained by Djellout, Guillin and Wu, but with an explicit constant only when $R < 1$; see [9, Theorem 2.5 and Remark 2.9]. Theorem 2.1 also improves upon section 4 of [5], where the assumptions were written in terms of moments of the considered measures.

The Monge–Kantorovich transportation problem involves finding, for given $\mu, \nu \in \mathrm{Prob}(X)$, an optimal transportation strategy in (1.1), namely a $\pi$ that minimises the transportation cost; a compactness and semi-continuity argument ensures that, for suitable cost functions, there always exists such a $\pi$. We recall that, given $\mu \in \mathrm{Prob}(X)$, another Polish space $Y$ and a continuous function $\varphi : X \to Y$, the measure *induced* from $\mu$ by $\varphi$ is the unique $\nu \in \mathrm{Prob}(Y)$ such that

$$\int_Y f(y)\nu(dy) = \int_X f(\varphi(x))\mu(x)$$

for all bounded and continuous $f : X \to \mathbb{R}$. Brenier and McCann showed that if $\mu$ and $\nu$ belong to $\mathrm{Prob}_2(\mathbb{R}^m)$, and if moreover $\mu$ is absolutely continuous with respect to Lebesgue measure, then there exists a convex function $\Phi : \mathbb{R}^m \to \mathbb{R}$ such that the gradient $\varphi = \nabla\Phi$ induces $\mu$ from $\nu$ and gives the unique solution to the Monge–Kantorovich transportation problem for $s = 2$, in the sense that

$$\int_{\mathbb{R}^m} \|\nabla\Phi(x) - x\|^2_{\ell^2}\, \mu(dx) = W_2(\mu, \nu)^2.$$

Further extensions of this result were obtained by Gangbo and McCann for $1 < s \leq 2$, by Ambrosio and Pratelli for $s = 1$, and by McCann [17] in the context of compact and connected $C^3$-smooth Riemannian manifolds that are without boundary (see also [7, 24]).

*Proof of Theorem 2.1.* In order to give an explicit solution in a case of importance, we first suppose that $X = \mathbb{R}^m$ and that $P^{(1)}$ and $p_j(dx_j \mid x^{(j-1)})$ $(j = 2, \dots, n)$ are all absolutely continuous with respect to Lebesgue measure. Then let $Q^{(n)} \in \mathrm{Prob}_s(\mathbb{R}^{nm})$ be of finite relative entropy with respect to $P^{(n)}$. Let $Q^{(j)}(dx^{(j)})$ be the marginal distribution of $x^{(j)} \in \mathbb{R}^{jm}$ with respect to $Q^{(n)}(dx^{(n)})$, and disintegrate $Q^{(n)}$ in terms of conditional probabilities, according to

$$Q^{(j)}(dx^{(j)}) = q_j(dx_j \mid x^{(j-1)})Q^{(j-1)}(dx^{(j-1)}).$$

In particular $q_j(. \mid x^{(j-1)})$ is absolutely continuous with respect to $p_j(. \mid x^{(j-1)})$ and hence with respect to Lebesgue measure, for $Q^{(j-1)}$ almost every $x^{(j-1)}$. A standard computation ensures that

$$\begin{aligned} \mathrm{Ent}\,(Q^{(n)} \mid P^{(n)}) \;&=\; \mathrm{Ent}\,(Q^{(1)} \mid P^{(1)}) \\ &\quad + \sum_{j=2}^{n} \int_{\mathbb{R}^{(j-1)m}} \mathrm{Ent}\,\big(q_j(\,.\, \mid x^{(j-1)}) \mid p_j(\,.\, \mid x^{(j-1)})\big)\, Q^{(j-1)}(dx^{(j-1)}). \end{aligned} \tag{2.1}$$

When the hypothesis (i) of Theorem 2.1 holds for some $1 < s \leq 2$, it also holds for $s = 1$. Consequently, by the Bobkov–Götze theorem, $P^{(1)}$ and $p_j(dx_j \mid x^{(j-1)})$ satisfy $GC(\kappa)$ for $\kappa = 1/\alpha$, and then one can check that there exists $\varepsilon > 0$ such that

$$\int_{\mathbb{R}^m} \exp(\varepsilon\|x^{(1)}\|^2_{\ell^2})P^{(1)}(dx^{(1)}) < \infty$$

and likewise for $p_j$; compare with Herbst's theorem [24, p. 280], and [3, 9]. Hence $Q^{(1)}$ and $q_j(dx_j \mid x^{(j-1)})$ for $Q^{(j-1)}$ almost every $x^{(j-1)}$ have finite second moments, since by Young's inequality

$$\int_{\mathbb{R}^m} \varepsilon \|x^{(n)}\|_{\ell^2}^2 \, Q^{(1)}(dx^{(1)}) \leq \mathrm{Ent}(Q^{(1)} \mid P^{(1)}) + \log \int_{\mathbb{R}^m} \exp\bigl(\varepsilon \|x^{(1)}\|_{\ell^2}^2\bigr) P^{(1)}(dx^{(1)}) < \infty$$

and likewise with $q_j$ and $p_j$ in place of $Q^{(1)}$ and $P^{(1)}$ respectively.

Let $\theta_1 : \mathbb{R}^m \to \mathbb{R}^m$ be an optimal transportation map that induces $P^{(1)}(dx_1)$ from $Q^{(1)}(dx_1)$; then for $Q^{(1)}$ every each $x_1$, let $x_2 \mapsto \theta_2(x_1, x_2)$ induce $p_2(dx_2 \mid \theta_1(x_1))$ from $q_2(dx_2 \mid x_1)$ optimally; hence $\Theta^{(2)} : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$, defined by $\Theta^{(2)}(x_1, x_2) = (\theta_1(x_1), \theta_2(x_1, x_2))$ on a certain set of full $Q^{(2)}$ measure, induces $P^{(2)}$ from $Q^{(2)}$. Generally, having constructed $\Theta^{(j)} : \mathbb{R}^{jm} \to \mathbb{R}^{jm}$, we let $x_{j+1} \mapsto \theta_{j+1}(x^{(j)}, x_{j+1})$ be an optimal transportation map that induces $p_{j+1}(dx_{j+1} \mid \Theta^{(j)}(x^{(j)}))$ from $q_{j+1}(dx_{j+1} \mid x^{(j)})$, for all $x^{(j)}$ in a certain set of full $Q^{(j)}$ measure; then we let $\Theta^{(j+1)} : \mathbb{R}^{(j+1)m} \to \mathbb{R}^{jm} \times \mathbb{R}^m$ be the map defined by

$$\Theta^{(j+1)}(x^{(j+1)}) = (\Theta^{(j)}(x^{(j)}), \theta_{j+1}(x^{(j+1)}))$$

on a set of full $Q^{(j+1)}$ measure. In particular $\Theta^{(j+1)}$ induces $P^{(j+1)}$ from $Q^{(j+1)}$, in the style of Kneser.

This transportation strategy may not be optimal, nevertheless it gives the bound

$$W_s(Q^{(n)}, P^{(n)})^s \leq \int_{\mathbb{R}^{nm}} \|\Theta^{(n)}(x^{(n)}) - x^{(n)}\|_{\ell^s}^s Q^{(n)}(dx^{(n)}) = \sum_{k=1}^{n} d_k \tag{2.2}$$

by the recursive definition of $\Theta^{(n)}$, where we have let

$$d_k = \int_{\mathbb{R}^{km}} \|\theta_k(x^{(k)}) - x_k\|_{\ell^s}^s Q^{(k)}(dx^{(k)}) \qquad (k = 1, \dots, n).$$

However, the transportation at step $k$ is optimal by construction, so

$$d_k = \int_{\mathbb{R}^{(k-1)m}} W_s\bigl(p_k(\,.\, \mid \Theta^{(k-1)}(x^{(k-1)})), q_k(\,.\, \mid x^{(k-1)})\bigr)^s Q^{(k-1)}(dx^{(k-1)}). \tag{2.3}$$

Given $a, b > 0$, $1 \leq s \leq 2$ and $\gamma > 1$, we have $(a+b)^s \leq (\gamma/(\gamma-1))^{s-1} a^s + \gamma^{s-1} b^s$. Hence by the triangle inequality, the expression (2.3) is bounded by

$$\begin{aligned} &\Bigl(\frac{\gamma}{\gamma-1}\Bigr)^{s-1} \int_{\mathbb{R}^{m(k-1)}} W_s\bigl(p_k(\,.\, \mid x^{(k-1)}), q_k(\,.\, \mid x^{(k-1)})\bigr)^s Q^{(k-1)}(dx^{(k-1)}) \\ &\qquad + \gamma^{s-1} \int_{\mathbb{R}^{m(k-1)}} W_s(p_k(\,.\, \mid \Theta^{(k-1)}(x^{(k-1)})), p_k(\,.\, \mid x^{(k-1)}))^s Q^{(k-1)}(dx^{(k-1)}). \end{aligned} \tag{2.4}$$

By hypothesis (i) and then Hölder's inequality, we bound the first integral in (2.4) by

$$h_k = \Bigl(\frac{\gamma}{\gamma-1}\Bigr)^{s-1} \Bigl(\frac{2}{\alpha}\Bigr)^{s/2} \Bigl(\int_{\mathbb{R}^{(k-1)m}} \mathrm{Ent}(q_k \mid p_k) \, dQ^{(k-1)}\Bigr)^{s/2}.$$

Meanwhile, on account of hypothesis (ii) the second integral in (2.4) is bounded by

$$\gamma^{s-1}\int_{\mathbb{R}^{m(k-1)}}\sum_{j=1}^{k-1}\rho_{k-j}\big\Vert\theta_j(x^{(j)})-x_j\big\Vert^s\,Q^{(k-1)}(dx^{(k-1)})=\gamma^{s-1}\sum_{j=1}^{k-1}\rho_{k-j}d_j,$$

and when we combine these contributions to (2.4) we have

$$d_k\leq h_k+\gamma^{s-1}\sum_{j=1}^{k-1}\rho_{k-j}\,d_j. \tag{2.5}$$

In the case when the $\rho_\ell$ are merely bounded by $M$, one can prove by induction that

$$d_k\leq h_k+\gamma^{s-1}M\sum_{j=1}^{k-1}h_j(1+\gamma^{s-1}M)^{k-1-j},$$

so that

$$\sum_{k=1}^n d_k\leq\sum_{j=1}^n h_j(1+\gamma^{s-1}M)^{n-j}\leq\Big(\sum_{j=1}^n h_j^{2/s}\Big)^{s/2}\Big(\sum_{j=1}^n(1+\gamma^{s-1}M)^{2(n-j)/(2-s)}\Big)^{(2-s)/2}$$

by Hölder's inequality. The first sum on the right-hand side is

$$\Big(\sum_{j=1}^n h_j^{2/s}\Big)^{s/2}=\Big(\frac{\gamma}{\gamma-1}\Big)^{s-1}\Big(\frac{2}{\alpha}\Big)^{s/2}\mathrm{Ent}\,(Q^{(n)}\mid P^{(n)})^{s/2}$$

by (2.1). Finally, setting $\gamma=1+1/n$, we obtain by (2.2) the stated result

$$W_s(Q^{(n)},P^{(n)})^s\leq\Big(\frac{2}{\alpha}\Big)^{s/2}\frac{(1+M)^n}{M}\,(n\,e)^{s-1}\,\mathrm{Ent}(Q^{(n)}\mid P^{(n)})^{s/2}.$$

(iii) Invoking the further hypothesis (iii), we see that $T_m=\sum_{j=1}^m d_j$ satisfies on account of (2.5) the recurrence relation

$$T_{m+1}\leq\sum_{j=1}^{m+1}h_j+\gamma^{s-1}RT_m,$$

which enables us to use Hölder's inequality again and bound $T_n$ by

$$\begin{aligned}\sum_{k=1}^n\Big(\sum_{j=1}^k h_j\Big)(\gamma^{s-1}R)^{n-k}&=\sum_{j=1}^n h_j\sum_{\ell=0}^{n-j}(\gamma^{s-1}R)^\ell\\&\leq\Big(\sum_{j=1}^n h_j^{2/s}\Big)^{s/2}\Big(\sum_{j=1}^n\Big(\sum_{\ell=0}^{n-j}(\gamma^{s-1}R)^\ell\Big)^{2/(2-s)}\Big)^{(2-s)/2}\end{aligned}$$

for $1 \le s < 2$. By (2.2) and the definition of $T_n$ this leads to

$$W_s(Q^{(n)},P^{(n)})^s \le \Big(\frac{\gamma}{\gamma-1}\Big)^{s-1}\Big(\sum_{m=1}^{n}\big(\sum_{\ell=0}^{m-1}(\gamma^{s-1}R)^{\ell}\big)^{2/(2-s)}\Big)^{(2-s)/2}\Big(\frac{2}{\alpha}\,\mathrm{Ent}(Q^{(n)}\mid P^{(n)})\Big)^{s/2} \tag{2.6}$$

$$\le \Big(\frac{\gamma}{\gamma-1}\Big)^{s-1} n^{1-s/2}\sum_{\ell=0}^{n-1}(\gamma^{s-1}R)^{\ell}\,\Big(\frac{2}{\alpha}\,\mathrm{Ent}(Q^{(n)}\mid P^{(n)})\Big)^{s/2}; \tag{2.7}$$

this also holds for $s=2$. Finally we select $\gamma$ according to the value of $R$ to make the bound (2.7) precise. When $R<1$, we let $\gamma = R^{-1/s} > 1$, so that $\gamma^{s-1}R = R^{1/s} < 1$, and we deduce the transportation inequality

$$W_s(Q^{(n)},P^{(n)})^s \le \Big(\frac{2}{\alpha}\Big)^{s/2}\frac{n^{1-s/2}}{(1-R^{1/s})^s}\mathrm{Ent}(Q^{(n)}\mid P^{(n)})^{s/2}.$$

When $R\ge 1$, we let $\gamma = 1+1/n$ to obtain the transportation inequality

$$W_s(Q^{(n)},P^{(n)})^s \le \Big(\frac{2}{\alpha}\Big)^{s/2}(n+1)^{s-1}\,n^{1-s/2}\,\Big(\frac{(1+1/n)^{n(s-1)}R^n-1}{(1+1/n)^{s-1}R-1}\Big)\,\mathrm{Ent}(Q^{(n)}\mid P^{(n)})^{s/2},$$

which leads to the stated result by simple analysis, and completes the proof when $X=\mathbb{R}^m$.

For typical Polish spaces $(X,d)$, we cannot rely on the existence of optimal maps, but we can use a less explicit inductive approach to construct the transportation strategy, as in [9]. Given $j=1,\dots,n-1$, assume that $\pi^{(j)}\in \mathrm{Prob}(X^{2j})$ has marginals $Q^{(j)}(dx^{(j)})$ and $P^{(j)}(dy^{(j)})$ and satisfies

$$W_s(Q^{(j)},P^{(j)})^s \le \int_{X^{2j}}\sum_{k=1}^{j} d(x_k,y_k)^s\pi^{(j)}(dx^{(j)}dy^{(j)}).$$

Then, for each $(x^{(j)},y^{(j)})\in X^{2j}$, let $\sigma_{j+1}(\,.\mid x^{(j)},y^{(j)})\in\mathrm{Prob}(X^2)$ be an optimal transportation strategy that has marginals $q_{j+1}(dx_{j+1}\mid x^{(j)})$ and $p_{j+1}(dy_{j+1}\mid y^{(j)})$ and that satisfies

$$W_s(q_{j+1}(\,.\mid x^{(j)}),p_{j+1}(\,.\mid y^{(j)}))^s = \int_{X^2} d(x_{j+1},y_{j+1})^s\,\sigma_{j+1}(dx_{j+1}\,dy_{j+1}\mid x^{(j)},y^{(j)}).$$

Now we let

$$\pi^{(j+1)}(dx^{(j+1)}dy^{(j+1)}) = \sigma_{j+1}(dx_{j+1}\,dy_{j+1}\mid x^{(j)},y^{(j)})\,\pi^{(j)}(dx^{(j)}dy^{(j)}),$$

which defines a probability on $X^{2(j+1)}$ with marginals $Q^{(j+1)}(dx^{(j+1)})$ and $P^{(j+1)}(dy^{(j+1)})$. This may not give an optimal transportation strategy; nevertheless, the recursive definition shows that

$$W_s(Q^{(n)},P^{(n)})^s \le \sum_{j=1}^{n}\int_{X^{2(j-1)}} W_s\big(q_j(\,.\mid x^{(j-1)}),p_j(\,.\mid y^{(j-1)})\big)^s\pi^{(j-1)}(dx^{(j-1)}dy^{(j-1)})$$

and one can follow the preceding proof from (2.2) onwards. □

*Proof of Theorem 1.2.* Under the hypotheses of Theorem 1.2, we can take $\rho_1 = L$ and $\rho_j = 0$ for $j = 2, \dots, n$, which satisfy Theorem 2.1 with $R = L$ in assumption (iii). □

The definition of $W_s$ not being well suited to direct calculation, we now give a computable sufficient condition for hypothesis (ii) of Theorem 2.1 to hold with some constant coefficients $\rho_\ell$ when $(X, d) = (\mathbb{R}^m, \ell^s)$.

**Proposition 2.2.** *Let $u_j : \mathbb{R}^{jm} \to \mathbb{R}$ be a twice continuously differentiable function that has bounded second-order partial derivatives. Let $1 \leq s \leq 2$ and suppose further that:*

*(i) $p_j(dx_j \mid x^{(j-1)}) = \exp(-u_j(x^{(j)}))\, dx_j$ satisfies $T_s(\alpha)$ for some $\alpha > 0$ and all $x^{(j-1)} \in \mathbb{R}^{m(j-1)}$;*

*(ii) there exists some real number $M_s$ such that*

$$\sup_{x^{(j-1)}} \int_{\mathbb{R}^m} \Big\| \Big(\frac{\partial u_j}{\partial x_k}\Big)_{k=1}^{j-1} \Big\|^2_{\ell^{s'}} p_j(dx_j \mid x^{(j-1)}) = M_s,$$

*where $1/s' + 1/s = 1$ and $\partial/\partial x_k$ denotes the gradient with respect to $x_k$.*
*Then $x^{(j-1)} \mapsto p_j(. \mid x^{(j-1)})$ is $\sqrt{(M_s/\alpha)}$-Lipschitz $(\mathbb{R}^{m(j-1)}, \ell^s) \to (Prob_s(\mathbb{R}^m), W_s)$.*

*Proof.* Given $x^{(j-1)}, \bar{x}^{(j-1)} \in \mathbb{R}^{m(j-1)}$, we let $x^{(j-1)}(t) = (1-t)\bar{x}^{(j-1)} + t x^{(j-1)}$ $(0 \leq t \leq 1)$ be the straight-line segment that joins them, and we consider

$$f(t) = W_s\big(p_j(. \mid x^{(j-1)}(t)), p_j(. \mid \bar{x}^{(j-1)})\big);$$

then it suffices to show that $f : [0,1] \to \mathbb{R}$ is Lipschitz and to bound its Lipschitz seminorm.

By the triangle inequality and (i), we have

$$\Big(\frac{f(t+\delta) - f(t)}{\delta}\Big)^2$$

$$\leq \frac{1}{\delta^2} W_s\big(p_j(. \mid x^{(j-1)}(t+\delta)), p_j(. \mid x^{(j-1)}(t))\big)^2$$

$$\leq \frac{1}{\alpha\delta^2}\Big\{\mathrm{Ent}\big(p_j(. \mid x^{(j-1)}(t+\delta)) \mid p_j(. \mid x^{(j-1)}(t))\big) + \mathrm{Ent}\big(p_j(. \mid x^{(j-1)}(t)) \mid p_j(. \mid x^{(j-1)}(t+\delta))\big)\Big\}$$

$$= \frac{1}{\alpha\delta^2} \int_{\mathbb{R}^m} \big(u_j(x^{(j-1)}(t+\delta), x_j) - u_j(x^{(j-1)}(t), x_j)\big)$$
$$\big\{\exp(-u_j(x^{(j-1)}(t), x_j)) - \exp(-u_j(x^{(j-1)}(t+\delta), x_j))\big\}\, dx_j. \quad (2.8)$$

However, by the assumptions on $u_j$ and the mean-value theorem, we have

$$u_j(x^{(j-1)}(t+\delta), x_j) - u_j(x^{(j-1)}(t), x_j)$$
$$= \delta \sum_{k=1}^{j-1} \Big\langle \frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j), x_k - \bar{x}_k \Big\rangle + \frac{\delta^2}{2} \big\langle \mathrm{Hess}\, u_j\, (x^{(j-1)} - \bar{x}^{(j-1)}), (x^{(j-1)} - \bar{x}^{(j-1)}) \big\rangle,$$

where $\operatorname{Hess} u_j$ is computed at some point between $(x^{(j-1)}, x_j)$ and $(\bar{x}^{(j-1)}, x_j)$ and is uniformly bounded. Proceeding in the same way for the other term (2.8), we obtain

$$\limsup_{\delta\to 0+}\Big(\frac{f(t+\delta)-f(t)}{\delta}\Big)^2 \le \frac{1}{\alpha}\int_{\mathbb{R}^m}\Big(\sum_{k=1}^{j-1}\Big\langle\frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j), x_k-\bar{x}_k\Big\rangle\Big)^2 p_j(dx_j \mid x^{(j-1)}(t)).$$

Hence by Hölder's inequality we have

$$\limsup_{\delta\to 0+}\frac{|f(t+\delta)-f(t)|}{\delta}$$
$$\le \frac{1}{\sqrt{\alpha}}\Big(\int_{\mathbb{R}^m}\Big(\sum_{k=1}^{j-1}\Big|\frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j)\Big|^{s'}\Big)^{2/s'} p_j(dx_j \mid x^{(j-1)}(t))\Big)^{1/2}\|x^{(j-1)}-\bar{x}^{(j-1)}\|_{\ell^s}$$

for $1 < s \le 2$, and likewise with obvious changes for $s = 1$. By assumption (ii) and Vitali's theorem, $f$ is Lipschitz with constant $\sqrt{(M_s/\alpha)}\|x^{(j-1)}-\bar{x}^{(j-1)}\|_{\ell^s}$, as required. □

## 3. Concentration inequalities for weakly dependent sequences

In terms of concentration inequalities, the dual version of Theorem 2.1 reads as follows.

**Theorem 3.1.** *Suppose that there exist $\kappa_1 > 0$ and $M \ge \rho_j \ge 0$ $(j = 1, \dots, n)$ such that:*
*(i) $P^{(1)}$ and $p_k(.\mid x^{(k-1)})$ $(k = 2, \dots, n;\ x^{(k-1)} \in X^{k-1})$ satisfy $GC(\kappa_1)$ on $(X, d)$;*
*(ii) $x^{(k-1)} \mapsto p_k(.\mid x^{(k-1)})$ is Lipschitz as a map $(X^{k-1}, d^{(1)}) \to (Prob_1(X), W_1)$ for $k = 2, \dots, n$, in the sense that*

$$W_1\big(p_k(.\mid x^{(k-1)}), p_k(.\mid y^{(k-1)})\big) \le \sum_{j=1}^{k-1}\rho_{k-j}\, d(x_j, y_j) \qquad (x^{(k-1)}, y^{(k-1)} \in X^{k-1}).$$

*Then the joint law $P^{(n)}$ satisfies $GC(\kappa_n)$ on $(X^n, d^{(1)})$, where*

$$\kappa_n = \kappa_1\frac{(1+M)^{2n}}{M^2}.$$

*Suppose moreover that*
*(iii) $\sum_{j=1}^n \rho_j \le R$.*
*Then $P^{(n)}$ satisfies $GC(\kappa_n(R))$ on $(X^n, d^{(1)})$, where*

$$\kappa_n(R) = \kappa_1\sum_{m=1}^{n}\Big(\sum_{k=0}^{m-1}R^k\Big)^2.$$

*Proof of Theorem 3.1.* This follows from the Bobkov–Götze theorem [3] and the bound (2.6) with $s = 1$ in the proof of Theorem 2.1. □

Alternatively, one can prove Theorem 3.1 directly by induction on the dimension, using the definition of $GC$.

*Proof of Theorem 1.1.* Under the hypotheses of Theorem 1.1, we can apply Theorem 3.1 with $\rho_1 = L$ and $\rho_j = 0$ for $j = 2, \dots, n$, which satisfy (iii) $\square$

## 4. Logarithmic Sobolev inequalities for ARMA models

In this section we give logarithmic Sobolev inequalities for the joint law of the first $n$ variables from two auto-regressive moving average processes. In both results we obtain constants that are independent of $n$, though the variables are not mutually independent, and we rely on the following general result which induces logarithmic Sobolev inequalities from one probability measure to another. For $m \geq 1$, let $\nu \in \mathrm{Prob}(\mathbb{R}^m)$ satisfy $LSI(\alpha)$, and let $\varphi$ be a $L$-Lipschitz map from $(\mathbb{R}^m, \ell^2)$ into itself; then, by the chain rule, the probability measure that is induced from $\nu$ by $\varphi$ satisfies $LSI(\alpha/L^2)$. Our first application is the following.

**Proposition 4.1.** *Let $Z_0$ and $Y_j$ $(j = 1, 2, \dots)$ be mutually independent random variables in $\mathbb{R}^m$, and let $\alpha > 0$ be a constant such that the distribution $P^{(0)}$ of $Z_0$ and the distribution of $Y_j$ $(j = 1, 2, \dots)$ satisfy $LSI(\alpha)$.*
*Then for any $L$-Lipschitz map $\Theta : \mathbb{R}^m \to \mathbb{R}^m$, the relation*

$$Z_{j+1} = \Theta(Z_j) + Y_{j+1} \qquad (j = 0, 1, \dots) \tag{4.1}$$

*determines a stochastic process such that, for any $n \geq 1$, the joint distribution $P^{(n-1)}$ of $(Z_j)_{j=0}^{n-1}$ satisfies $LSI(\alpha_n)$ where*

$$\alpha_n = \begin{cases} (1-L)^2\alpha & \text{if} \quad 0 \leq L < 1, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if} \quad L = 1, \\ \dfrac{L-1}{L^n}\dfrac{\alpha}{e(n+1)} & \text{if} \quad L > 1. \end{cases}$$

*Proof.* For $(z_0, y_1, \dots, y_{n-1}) \in \mathbb{R}^{nm}$, let $\varphi_n(z_0, y_1, \dots, y_{n-1})$ be the vector $(z_0, \dots, z_{n-1})$, defined by the recurrence relation

$$z_{k+1} = \Theta(z_k) + y_{k+1} \qquad (k = 0, \dots, n-2). \tag{4.2}$$

Using primes to indicate another solution of (4.2), we deduce the following inequality from the Lipschitz condition on $\Theta$:

$$\|z_{k+1} - z'_{k+1}\|^2 \leq (1+\varepsilon)L^2\|z_k - z'_k\|^2 + (1+\varepsilon^{-1})\|y_{k+1} - y'_{k+1}\|^2 \tag{4.3}$$

for all $\varepsilon > 0$. In particular (4.3) implies the bound

$$\|z_k - z'_k\|^2 \leq \big((1+\varepsilon)L^2\big)^k\|z_0 - z'_0\|^2 + (1+\varepsilon^{-1})\sum_{j=1}^{k}\big((1+\varepsilon)L^2\big)^{k-j}\|y_j - y'_j\|^2.$$

By summing over $k$, one notes that $\varphi_n$ defines a Lipschitz function from $(\mathbb{R}^{nm}, \ell^2)$ into itself, with Lipschitz seminorm

$$L_{\varphi_n} \leq \Big( (1+\varepsilon^{-1}) \sum_{k=0}^{n-1} \big( (1+\varepsilon) L^2 \big)^k \Big)^{1/2}$$

We now select $\varepsilon > 0$ according to the value of $L$: when $L < 1$, we let $\varepsilon = L^{-1} - 1 > 0$, so that $L_{\varphi_n} \leq (1-L)^{-1}$; whereas when $L \geq 1$, we let $\varepsilon = n^{-1}$, and obtain $L_{\varphi_n} \leq [n(n+1)(e-1)]^{(1/2)}$ for $L = 1$, and $L_{\varphi_n} \leq [e(n+1)L^n(L-1)^{-1}]^{1/2}$ for $L > 1$.

Moreover, $\varphi_n$ induces the joint distribution of $(Z_j)_{j=0}^{n-1}$ from the joint distribution of $(Z_0, Y_1, \dots, Y_{n-1})$. By independence, the joint distribution of $(Z_0, Y_1, \dots, Y_{n-1})$ is a product measure on $(\mathbb{R}^{nm}, \ell^2)$ that satisfies $LSI(\alpha)$. Hence the joint distribution of $(Z_j)_{j=0}^{n-1}$ satisfies $LSI(\alpha)$, where $\alpha = L_{\varphi_n}^{-2}\, \alpha$. □

The linear case gives the following result for ARMA processes.

**Proposition 4.2.** *Let $A$ and $B$ be $m \times m$ matrices such that the spectral radius $\rho$ of $A$ satisfies $\rho < 1$. Let also $Z_0$ and $Y_j$ $(j = 1, 2, \dots)$ be mutually independent standard Gaussian $N(0, I_m)$ random variables in $\mathbb{R}^m$. Then, for any $n \geq 1$, the joint distribution of the ARMA process $(Z_j)_{j=0}^{n-1}$, defined by the recurrence relation*

$$Z_{j+1} = AZ_j + BY_{j+1} \qquad (j = 0, 1, \dots),$$

*satisfies $LSI(\alpha)$ where*

$$\alpha = \Big( \frac{(1-\sqrt{\rho})}{\max\{1, \|B\|\}} \Big)^2 \Big( \sum_{j=0}^{\infty} \rho^{-j} \|A^j\|^2 \Big)^{-2}.$$

*Proof.* By Rota's Theorem [19], $A$ is similar to a strict contraction on $(\mathbb{R}^m, \ell^2)$; that is, there exists an invertible $m \times m$ matrix $S$ and a matrix $C$ such that $\|C\| \leq 1$ and $A = \sqrt{\rho} S^{-1} C S$; one can choose the similarity so that the operator norms satisfy

$$\|S\| \|S^{-1}\| \leq \sum_{j=0}^{\infty} \rho^{-j} \|A^j\|^2 < \infty.$$

Hence the ARMA process reduces to the solution of the recurrence relation

$$SZ_{j+1} = \sqrt{\rho}\, CSZ_j + SBY_{j+1} \qquad (j = 0, 1, \dots) \tag{4.4}$$

which involves the $\sqrt{\rho}$-Lipschitz linear map $\Theta : \mathbb{R}^m \to \mathbb{R}^m : \Theta(w) = \sqrt{\rho}\, C\, w$. Given $n \geq 1$, the linear map $\Phi_n : \mathbb{R}^{nm} \to \mathbb{R}^{nm}$, defined to solve (4.4) by

$$(z_0, y_1, \dots, y_n) \mapsto (Sz_0, SBy_1, \dots, SBy_{n-1}) \mapsto (Sz_0, Sz_1, \dots, Sz_{n-1}) \mapsto (z_0, z_1, \dots, z_{n-1}),$$

has operator norm

$$\|\Phi_n\| \leq \|S\| \|S^{-1}\| (1-\sqrt{\rho})^{-1} \max\{1, \|B\|\};$$

moreover, $\Phi_n$ induces the joint distribution of $(Z_0, \dots, Z_{n-1})$ from the joint distribution of $(Z_0, Y_1, \dots, Y_{n-1})$. By Gross's Theorem (see [11]), the latter distribution satisfies $LSI(1)$, and hence the induced distribution satisfies $LSI(\alpha)$, with $\alpha = \|\Phi_n\|^{-2}$. □

**Remarks 4.3.** (i) As compared to Proposition 4.1, the condition imposed in Proposition 4.2 involves the spectral radius of the matrix $A$ and not its operator norm. In particular, for matrices with norm 1, Proposition 4.1 only leads to $LSI$ with constant of order $n^{-2}$; whereas Proposition 4.2 ensures $LSI$ with constant independent of $n$ under the spectral radius assumption $\rho < 1$.

(ii) The joint distribution of the ARMA process is discussed by Djellout, Guillin and Wu [9, Section 3]. We have improved upon [9] by obtaining $LSI(\alpha)$, hence $T_2(\alpha)$, under the spectral radius condition $\rho < 1$, where $\alpha$ is independent of the size $n$ of the considered sample and the size of the matrices.

## 5. Logarithmic Sobolev inequality for weakly dependent processes

In this section we consider a stochastic process $(\xi_j)_{j=1}^n$, with state space $\mathbb{R}^m$ and initial distribution $P^{(1)}$, which is not necessarily Markovian; we also assume that the transition kernels have positive densities with respect to Lebesgue measure, and write

$$dp_j = p_j(dx_j \mid x^{(j-1)}) = e^{-u_j(x^{(j)})} dx_j \qquad (j = 2, \dots, n).$$

The coupling between variables is measured by the following integral

$$\Lambda_{j,k}(s) = \sup_{x^{(j-1)}} \int_{\mathbb{R}} \exp\Big(\big\langle s, \frac{\partial u_j}{\partial x_k}(x^{(j)})\big\rangle\Big)\, p_j(dx_j \mid x^{(j-1)}), \qquad (s \in \mathbb{R}^m,\ 1 \leq k < j \leq n)$$

where as above $\partial/\partial x_k$ denotes the gradient with respect to $x_k \in \mathbb{R}^m$. The main result in this section is the following.

**Theorem 5.1.** *Suppose that there exist constants $\alpha > 0$ and $\kappa_{j,k} \geq 0$ for $1 \leq k < j \leq n$ such that*

*(i) $P^{(1)}$ and $p_k(. \mid x^{(k-1)})$ $(k = 2, \dots, n;\ x^{(k-1)} \in \mathbb{R}^{m(k-1)})$ satisfy $LSI(\alpha)$;*

*(ii) $\Lambda_{j,k}(s) \leq \exp(\kappa_{j,k}\, \|s\|^2/2)$ holds for all $s \in \mathbb{R}^m$.*

*Then the joint distribution $P^{(n)}$ satisfies $LSI(\alpha_n)$ with*

$$\alpha_n = \frac{\alpha}{1+\varepsilon}\Big(1 + \sum_{k=0}^{n-2} \prod_{m=k+1}^{n-1} (1 + K_m)\Big)^{-1} \tag{5.1}$$

*ll $\varepsilon > 0$, where $K_j = (1+\varepsilon^{-1}) \sum_{\ell=0}^{j-1} \kappa_{n-\ell,n-j}/\alpha$ for $j = 1, \dots, n-1$.*

*Suppose further that there exist $R \geq 0$ and $\rho_\ell \geq 0$ for $\ell = 1, \dots, n-1$ such that*

*(iii) $\kappa_{j,k} \leq \rho_{j-k}$ for $1 \leq k < j \leq n$, and $\sum_{\ell=1}^{n-1} \sqrt{\rho_\ell} \leq \sqrt{R}$.*

*Then* $P^{(n)}$ *satisfies* $LSI(\alpha_n)$ *where*

$$\alpha_n = \begin{cases} \frac{\left(\sqrt{\alpha}-\sqrt{R}\right)^2}{\alpha} & \text{if} \quad R<\alpha, \\ \frac{\alpha}{n(n+1)(e-1)} & \text{if} \quad R=\alpha, \\ \left(\frac{\alpha}{R}\right)^n \frac{R-\alpha}{e(n+1)} & \text{if} \quad R>\alpha. \end{cases}$$

Before proving this theorem, we give simple sufficient conditions for hypothesis (ii) to hold. When $m = 1$, hypothesis (i) is equivalent to a condition on the cumulative distribution functions by the criterion for $LSI$ given in [3].

**Proposition 5.2.** *In the above notation, let* $1 \le k < j$ *and suppose that there exist* $\alpha > 0$ *and* $L_{j,k} \ge 0$ *such that*

*(i)* $p_j(. \mid x^{(j-1)})$ *satisfies* $GC(1/\alpha)$ *for all* $x^{(j-1)} \in \mathbb{R}^{m(j-1)}$*;*

*(ii)* $u_j$ *is twice continuously differentiable and the off-diagonal blocks of its Hessian matrix satisfy*

$$\left\| \frac{\partial^2 u_j}{\partial x_j \partial x_k} \right\| \le L_{j,k}$$

*as matrices* $(\mathbb{R}^m, \ell^2) \to (\mathbb{R}^m, \ell^2)$.
*Then*

$$\Lambda_{j,k}(s) \le \exp(L_{j,k}^2 \|s\|^2/(2\alpha)) \qquad (s \in \mathbb{R}^m).$$

*Proof of Proposition 5.2.* Letting $s = \|s\|\, e$ for some unit vector $e$, we note that by (ii) the real function $x_j \mapsto \langle e, \partial u_j/\partial x_k\rangle$ is $L_{j,k}$-Lipschitz in the variable of integration, and that

$$\int_{\mathbb{R}} \Big\langle e, \frac{\partial u_j}{\partial x_k}\Big\rangle\, p_j(dx_j \mid x^{(j-1)}) = -\Big\langle e, \frac{\partial}{\partial x_k} \int_{\mathbb{R}} p_j(dx_j \mid x^{(j-1)})\Big\rangle = 0$$

since $p_j(. \mid x^{(j-1)})$ is a probability measure. Then, by (i),

$$\int_{\mathbb{R}} \exp\Big(\Big\langle s, \frac{\partial u_j}{\partial x_k}\Big\rangle\Big)\, p_j(dx_j \mid x^{(j-1)}) \le \exp(\kappa\, L_{j,k}^2 \|s\|^2/2)$$

holds for all $x^{(j-1)}$ in $\mathbb{R}^{m(j-1)}$. This inequality implies the Proposition. □

*Proof of Proposition 5.2.* For notational convenience, $X$ denotes the state space $\mathbb{R}^m$. Then let $f : X^n \to \mathbb{R}$ be a smooth and compactly supported function, and let $g_j : X^{n-j} \to \mathbb{R}$ be defined by $g_0 = f$ and by

$$g_j(x^{(n-j)}) = \Big(\int_X g_{j-1}(x^{(n-j+1)})^2\, p_{n-j}(dx_{n-j+1} \mid x^{(n-j)})\Big)^{1/2} \tag{5.2}$$

for $j = 1, \dots, n-1$; finally, let $g_n$ be the constant $(\int f^2 dP^{(n)})^{1/2}$.

From the recursive formula (5.2) one can easily verify the identity

$$\int_{X^n} f^2 \log\Big(f^2 / \int_{X^n} f^2 dP^{(n)}\Big) dP^{(n)} = \sum_{j=0}^{n-1} \int_{X^{n-j}} g_j^2 \log\big(g_j^2/g_{j+1}^2\big) dP^{(n-j)} \tag{5.3}$$

which is crucial to the proof; indeed, it allows us to obtain the result from logarithmic Sobolev inequalities on $X$.

By hypothesis (i), the measure $dp_{n-j} = p_{n-j}(dx_{n-j} \mid x^{(n-j-1)})$ satisfies $LSI(\alpha)$, whence

$$\int_X g_j^2 \log\big(g_j^2/g_{j+1}^2\big)\,dp_{n-j} \le \frac{2}{\alpha}\int_X \Big(\frac{\partial g_j}{\partial x_{n-j}}\Big)^2 dp_{n-j} \qquad (j=0,\dots,n-1), \tag{5.4}$$

where for $j=n-1$ we take $dp_1 = P^{(1)}(dx_1)$. The next step is to express these derivatives in terms of the gradient of $f$, using the identity

$$g_j \frac{\partial g_j}{\partial x_{n-j}} = \int_{X^{n-j}} f \frac{\partial f}{\partial x_{n-j}}\, dp_n \dots dp_{n-j+1} - \frac{1}{2}\sum_{\ell=0}^{j-1}\int_{X^{j-\ell}} g_\ell^2 \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\, dp_{n-\ell}\dots dp_{n-j+1} \tag{5.5}$$

which follows from the definition (5.2) of $g_j^2$ and that of $p_{n-j}$. The integrals on the right-hand side of (5.5) will be bounded by the following Lemma.

**Lemma 5.3.** *Let $0 \le \ell < j \le n-1$, and assume that hypothesis (ii) holds. Then*

$$\Big\|\int_X g_\ell^2 \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\, dp_{n-\ell}\Big\| \le g_{\ell+1}\Big(2\,\kappa_{n-\ell,n-j}\int_X g_\ell^2 \log(g_\ell^2/g_{\ell+1}^2)\, dp_{n-\ell}\Big)^{1/2}. \tag{5.6}$$

*Proof of Lemma 5.3.* By definition of $\Lambda_{n-\ell,n-j}$, we have

$$\int_X \exp\Big(\langle s, \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\rangle - \log\Lambda_{n-\ell,n-j}(s)\Big) dp_{n-\ell} \le 1 \qquad (s\in X),$$

and hence by the dual formula for relative entropy, as in [4, p. 693],

$$\int_X \Big(\langle s, \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\rangle - \log\Lambda_{n-\ell,n-j}(s)\Big) g_\ell^2\, dp_{n-\ell} \le \int_X g_\ell^2 \log\big(g_\ell^2/g_{\ell+1}^2\big)\, dp_{n-\ell}.$$

Then hypothesis (ii) of the Theorem ensures that

$$\langle s, \int_X \frac{\partial u_{n-\ell}}{\partial x_{n-j}}\, g_\ell^2\, dp_{n-\ell}\rangle \le \frac{\|s\|^2}{2}\,\kappa_{n-\ell,n-j}\, g_{\ell+1}^2 + \int_X g_\ell^2 \log\big(g_\ell^2/g_{\ell+1}^2\big)\, dp_{n-\ell}$$

and the stated result follows by optimizing this over $s \in \mathbb{R}^m$. □

*Conclusion of the Proof of Theorem 5.1.* When we integrate (5.6) with respect to $dp_{n-\ell-1}\dots dp_{n-j+1}$, we deduce by the Cauchy–Schwarz inequality that

$$\Big|\int_{\mathbb{R}^{j-\ell}} g_\ell^2 \frac{\partial u_{n-\ell}}{\partial x_{n-j}} dp_{n-\ell}\dots dp_{n-j+1}\Big| \le g_j\Big(2\,\kappa_{n-\ell,n-j}\int_{\mathbb{R}^{j-\ell}} g_\ell^2 \log(g_\ell^2/g_{\ell+1}^2) dp_{n-\ell}\dots dp_{n-j+1}\Big)^{1/2}.$$

Then, by integrating the square of (5.5) with respect to $dP^{(n-j)}$ and making a further application of the Cauchy–Schwarz inequality, we obtain

$$\int_{X^{n-j}} \Big\| \frac{\partial g_j}{\partial x_{n-j}} \Big\|^2 dP^{(n-j)} \le (1+\varepsilon) \int_{X^n} \Big\| \frac{\partial f}{\partial x_{n-j}} \Big\|^2 dP^{(n)} + \frac{1+\varepsilon^{-1}}{4} \Big\{ \sum_{\ell=0}^{j-1} \Big( 2\, \kappa_{n-\ell,n-j}\, h_\ell \Big)^{1/2} \Big\}^2 \tag{5.7}$$

where $\varepsilon > 0$ is arbitrary and $h_\ell$ is given by

$$h_\ell = \int_{X^{n-\ell}} g_\ell^2 \log\big(g_\ell^2 / g_{\ell+1}^2\big)\, dP^{(n-\ell)}.$$

From (5.7), which holds true for $j = 1, \dots, n-1$, we first prove the general result given in (5.1). By (5.4) and the Cauchy–Schwarz inequality again, we obtain from (5.7) the crucial inequality

$$h_j \le d_j + K_j \sum_{m=0}^{j-1} h_m \qquad (j = 1, \dots, n-1)$$

where we have let

$$\begin{aligned} d_j &= \frac{2(1+\varepsilon)}{\alpha} \int_{\mathbb{R}^n} \Big( \frac{\partial f}{\partial x_{n-j}} \Big)^2 dP^{(n)} \qquad (j = 0, \dots, n-1), \\ K_j &= \frac{1+\varepsilon^{-1}}{\alpha} \sum_{\ell=0}^{j-1} \kappa_{n-\ell,n-j} \qquad\qquad (j = 1, \dots, n-1). \end{aligned}$$

Since $h_0 \le d_0$ and all terms are positive, the partial sums $H_k = \sum_{j=0}^{k} h_j$ satisfy the system of inequalities

$$H_k \le d_k + (1 + K_k) H_{k-1} \qquad (k = 1, \dots, n-1),$$

with $H_0 \le d_0$. By induction, one can deduce that

$$H_{n-1} \le d_{n-1} + \sum_{k=0}^{n-2} d_k \prod_{\ell=k+1}^{n-1} (1 + K_\ell),$$

which in turn implies the bound

$$H_{n-1} \le \Big( 1 + \sum_{k=0}^{n-2} \prod_{\ell=k+1}^{n-1} (1 + K_\ell) \Big) \sum_{j=0}^{n-1} d_j.$$

By (5.3) this is equivalent to the inequality

$$\int_{X^n} f^2 \log\Big( f^2 \Big/ \int_{X^n} f^2 dP^{(n)} \Big) dP^{(n)} \le \frac{2(1+\varepsilon)}{\alpha} \Big( 1 + \sum_{k=0}^{n-2} \prod_{\ell=k+1}^{n-1} (1 + K_\ell) \Big) \int_{X^n} \|\nabla f\|^2 dP^{(n)}.$$

Since $f$ is arbitrary, this ensures that $P^{(n)}$ satisfies $LSI(\alpha_n)$ with $\alpha_n$ as in (5.1).

(iii) The extra hypothesis (iii) enables us to strengthen the preceding inequalities, so (5.7) leads to the convolution-type inequality

$$h_j \le d_j + \frac{1+\varepsilon^{-1}}{\alpha}\Big(\sum_{\ell=0}^{j-1}\sqrt{\rho_{j-\ell}}\,\sqrt{h_\ell}\Big)^2$$

for $j = 1, \dots, n-1$, and $h_0 \le d_0$ for $j = 0$. By summing over $j$ we obtain

$$\sum_{j=0}^{k} h_j \le \sum_{j=0}^{k} d_j + \frac{1+\varepsilon^{-1}}{\alpha}\sum_{j=1}^{k}\Big(\sum_{\ell=0}^{j-1}\sqrt{\rho_{j-\ell}}\,\sqrt{h_\ell}\Big)^2,$$

which implies by Young's convolution inequality that

$$\sum_{j=0}^{k} h_j \le \sum_{j=0}^{k} d_j + \frac{1+\varepsilon^{-1}}{\alpha}\Big(\sum_{\ell=1}^{k}\sqrt{\rho_\ell}\Big)^2\sum_{\ell=0}^{k-1} h_\ell.$$

Now let $R_j = (\sum_{\ell=1}^{j}\sqrt{\rho_\ell})^2$ and $D_j = \sum_{\ell=0}^{j} d_\ell$; then by induction one can prove that

$$H_k \le D_k + \sum_{j=0}^{k-1} D_j \prod_{\ell=j+1}^{k}\frac{1+\varepsilon^{-1}}{\alpha}R_\ell$$

for $k = 1, \dots, n-1$, and hence

$$H_{n-1} \le \Big(1 + \sum_{j=0}^{n-2}\Big(\frac{1+\varepsilon^{-1}}{\alpha}R\Big)^{n-j-1}\Big)D_{n-1} = \sum_{\ell=0}^{n-1}\Big(\frac{1+\varepsilon^{-1}}{\alpha}R\Big)^{\ell} D_{n-1} \tag{5.8}$$

since $D_j \le D_{n-1}$ and $R_j \le R$ by hypothesis (iii). We finally select $\varepsilon$ to make the bound (5.8) precise, according to the relative values of $R$ and $\alpha$.

When $R = 0$, we recover $LSI(\alpha)$ for $P^{(n)}$ as expected, since here $P^{(n)}$ is the tensor product of its marginal distributions, which satisfy $LSI(\alpha)$.

When $0 < R < \alpha$, we choose $\varepsilon = (\sqrt{(\alpha/R)}-1)^{-1} > 0$ so that $(1+\varepsilon^{-1})R/\alpha = \sqrt{(R/\alpha)} < 1$ and hence

$$H_{n-1} \le D_{n-1}\sum_{\ell=0}^{\infty}(R/\alpha)^{\ell/2} = \frac{D_{n-1}}{1-\sqrt{(R/\alpha)}},$$

which by (5.3) and the definition of $H_{n-1}$ and $D_{n-1}$ implies the inequality

$$\int_{X^n} f^2\log\Big(f^2\Big/\int_{X^n} f^2 dP^{(n)}\Big)dP^{(n)} \le \frac{2}{(\sqrt{\alpha}-\sqrt{R})^2}\int_{X^n}\|\nabla f\|^2 dP^{(n)}.$$

When $R \ge \alpha$, we choose $\varepsilon = n$ in (5.8), obtaining

$$H_{n-1} \le \frac{2(n+1)}{\alpha}\Big(\frac{(1+1/n)^n(R/\alpha)^n - 1}{(1+1/n)(R/\alpha)-1}\Big)\int_{X^n}\|\nabla f\|^2 dP^{(n)};$$

as above this leads to the stated result by (5.3). This concludes the proof. □

## 6. Logarithmic Sobolev inequalities for Markov processes

The results of the preceding section simplify considerably when we have an homogeneous Markov process $(\xi_j)_{j=1}^n$ with state space $\mathbb{R}^m$, as we shall now show. Suppose that the transition measure is $p(dy \mid x) = e^{-u(x,y)}dy$ where $u$ is a twice continuously differentiable function such that

$$\Lambda(s \mid x) = \int_{\mathbb{R}} \exp\Big(\big\langle s, \frac{\partial u}{\partial x}(x,y)\big\rangle\Big) p(dy \mid x) < \infty \qquad (s, x \in \mathbb{R}^m). \tag{6.1}$$

Then Theorem 5.1 has the following consequence.

**Corollary 6.1.** *Suppose that there exist constants $\kappa \geq 0$ and $\alpha > 0$ such that:*
*(i) $P^{(1)}$ and $p(. \mid x)$ $(x \in \mathbb{R}^m)$ satisfy $LSI(\alpha)$;*
*(ii) $\Lambda(s \mid x) \leq \exp(\kappa \|s\|^2/2)$ holds for all $s, x \in \mathbb{R}^m$.*
*Then the joint law $P^{(n)}$ of the first $n$ variables satisfies $LSI(\alpha_n)$, where*

$$\alpha_n = \begin{cases} \big(\sqrt{\alpha} - \sqrt{\kappa}\big)^2 & \text{if} \quad \kappa < \alpha, \\ \dfrac{\alpha}{n(n+1)(e-1)} & \text{if} \quad \kappa = \alpha, \\ \Big(\dfrac{\alpha}{\kappa}\Big)^n \dfrac{\kappa - \alpha}{e(n+1)} & \text{if} \quad \kappa > \alpha. \end{cases}$$

*Proof.* In the notation of section 5, we have $u_j(x^{(j)}) = u(x_{j-1}, x_j)$, so we can take $\kappa_{j,m} = 0$ for $m = 1, \dots, j-2$, and $\kappa_{j,j-1} = \kappa$ for $j = 2, \dots, n$; hence we can take $\rho_1 = \kappa$ and $\rho_j = 0$ for $j = 2, 3, \dots$. Now we can apply Theorem 5.1 (iii) and obtain the stated result with $R = \kappa$ in the various cases. (In fact (5.7) simplifies considerably for a Markov process, and hence one can obtain an easier direct proof of Corollary 6.1.) □

*Proof of Theorem 1.3.* By the mean-value theorem and hypothesis (ii) of Theorem 1.3, the function $y \mapsto \langle e, \partial u/\partial x\rangle$ is $L$-Lipschitz $(\mathbb{R}^m, \ell^2) \to \mathbb{R}$ for any unit vector $e$ in $\mathbb{R}^m$, and hence $\Lambda(s \mid x) \leq \exp(\|s\|^2 L^2/(2\alpha))$ holds for all $s \in \mathbb{R}^m$ as in Proposition 5.2. Hence we can take $\kappa = L^2/\alpha$ in Corollary 6.1 and deduce Theorem 1.3 with the various values of the constant. □

**Remarks 6.2.** (i) Theorem 5.1 and Corollary 6.1 extend with suitable changes in notation when the state space is a connected $C^1$-smooth Riemannian manifold $X$. The proofs reduce to calculations in local co-ordinate charts. McCann [17] has shown that a locally Lipschitz function on $X$ is differentiable except on a set that has zero Riemannian volume; so a $L$-Lipschitz condition on $f : X \to \mathbb{R}$ is essentially equivalent to $\|\nabla f\| \leq L$.

(ii) Corollary 6.1 is a natural refinement of Theorems 1.1 and 1.2. Indeed $LSI(\alpha)$ implies $T_s(\alpha)$. Then, in the notation of the mentioned results, suppose that $u$ is a twice continuously differentiable function with bounded second-order partial derivatives. Then, by Proposition 2.2, hypotheses (i) and (ii) of Corollary 6.1 together imply that the map $x \mapsto p(. \mid x)$ is $(\kappa/\alpha)^{1/2}$ Lipschitz as a function $\mathbb{R}^m \to (\mathrm{Prob}_2(\mathbb{R}^m), W_2)$, hence $\mathbb{R}^m \to$

$(\mathrm{Prob}_s(\mathbb{R}^m), W_s)$ as in Theorems 1.1 or 1.2. Similarly Proposition 2.2 ensures that Theorem 5.1 is a refinement of Theorem 2.1 with, for $s = 2$,

$$M \leq M_2/\alpha = \frac{1}{\alpha} \sup_{x^{(j-1)}} \sum_{k=1}^{j-1} \int_{\mathbb{R}^m} \Big\| \frac{\partial u_j}{\partial x_k} \Big\|^2 p_j(dx_j \mid x^{(j-1)}) \leq \frac{1}{\alpha} \sum_{k=1}^{j-1} \kappa_{j,k}.$$

Note also the similarity between the constants in Theorem 2.1 (iii) and Theorem 5.1 (iii) when $s = 2$ and one rescales $R$ suitably. In Example 6.3 we show these constants to be optimal.

**Example 6.3.** (Ornstein–Uhlenbeck Process) We now show that the constants $\kappa_n$ of Theorem 1.1 (or Theorem 3.1(iii)) and $\alpha_n$ of Corollary 6.1 have optimal growth in $n$. For this purpose we consider the real Ornstein–Uhlenbeck process conditioned to start at $x \in \mathbb{R}$, namely the solution to the Itô stochastic differential equation

$$dZ_t^{(x)} = -\rho Z_t^{(x)} dt + dB_t^{(0)}, \qquad (t \geq 0)$$

where $(B_t^{(0)})$ is a real standard Brownian motion starting at 0, and $\rho \in \mathbb{R}$. In financial modelling, OU processes with $\rho < 0$ are used to model stock prices in a rising market (see [10, p 26] for instance). More precisely we consider the discrete-time Markov process $(\xi_j)_{j=1}^n$ defined by $\xi_j = Z_{j\tau}^{(x)}$ where $\tau > 0$, and test the Gaussian concentration inequality with the 1-Lipschitz function $F_n : (\mathbb{R}^n, \ell^1) \to \mathbb{R}$ defined by $F_n(x^{(n)}) = \sum_{j=1}^n x_j$.

The exponential integral satisfies

$$\int_{\mathbb{R}^n} \exp\big(sF_n(x^{(n)})\big)\, P^{(n)}(dx^{(n)}) = \mathbb{E}\,\exp\big(sF_n(\xi^{(n)})\big) = \mathbb{E}\,\exp\Big(s \sum_{j=1}^n Z_{j\tau}^{(x)}\Big). \tag{6.2}$$

This sum can be expressed in terms of the increments of the OU process

$$\sum_{j=1}^n Z_{j\tau}^{(x)} = \sum_{i=1}^n \theta^i\, Z_0^{(x)} + \sum_{j=0}^{n-1} \sum_{i=0}^{n-j-1} \theta^i \,\big(Z_{(j+1)\tau}^{(x)} - \theta Z_{j\tau}^{(x)}\big),$$

with $\theta = e^{-\rho\tau}$. Moreover one can integrate the stochastic differential equation and prove that $(Z_{(j+1)\tau}^{(x)} - \theta Z_{j\tau}^{(x)})_{0 \leq k \leq n-1}$ are independent random variables each with $N(0, \sigma^2)$ distribution, where $\sigma^2 = (1-\theta^2)/(2\rho)$ when $\rho \neq 0$, and $\sigma^2 = \tau$ when $\rho = 0$. Hence the exponential integral (6.2) equals

$$\exp\Big(s \sum_{i=1}^n \theta^i x\Big) \prod_{j=0}^{n-1} \mathbb{E} \exp\Big[s\Big(\sum_{i=0}^{n-j-1} \theta^i\Big)\big(Z_{(j+1)\tau}^{(x)} - \theta Z_{j\tau}^{(x)}\big)\Big] = \exp\big(s\, \mathbb{E}\, F_n(\xi^{(n)}) + s^2 \kappa_n/2\big)$$

where

$$\kappa_n = \sigma^2 \sum_{j=0}^{n-1} \Big( \sum_{i=0}^{n-j-1} \theta^i \Big)^2. \tag{6.3}$$

However, hypothesis (i) of Theorem 1.1 holds with $L = \theta$, since $P^{(1)}$ with distribution $N(x, \sigma^2)$ and $p(. \mid x)$ with distribution $N(\theta x, \sigma^2)$ satisfy $GC(\kappa_1)$ where $\kappa_1 = \sigma^2$, while hypothesis (ii) is satisfied with

$$W_1(p(. \mid x), p(. \mid x')) = W_1(N(\theta x, \sigma^2), N(\theta x', \sigma^2)) = \theta|x - x'| \qquad (x, x' \in \mathbb{R}). \tag{6.4}$$

Hence the constant $\kappa_n(L)$ given by Theorem 1.1 is exactly the directly computed constant $\kappa_n$ in (6.3), in each of the cases $L = 1$, $L > 1$ and $L < 1$, corresponding to $\rho = 0$, $\rho < 0$ and $\rho > 0$.

As regards Corollary 6.1, note that the transition probability is given by

$$p(dy \mid x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(y - \theta x)^2}{2\sigma^2}\Big) dy$$

since $Z_\tau^{(x)}$ is distributed as $\theta x + B_{\sigma^2}^{(0)}$. Hence by direct calculation we have

$$\alpha = \frac{1}{\sigma^2}, \quad \kappa = \frac{\theta^2}{\sigma^2}, \quad L = \frac{\theta}{\sigma^2};$$

consequently the dependence parameters $(\kappa/\alpha)^{1/2}$ and $\theta$ given in (6.4) coincide, as in Remark 6.2(ii).

Further, by considering the function $f(x^{(n)}) = \exp\big(\sum_{j=1}^n \theta^j x_j\big)$, one can prove that the joint law $P^{(n)}$ cannot satisfy a logarithmic Sobolev inequality with $\alpha_n$ greater than some constant multiple of $n^{-3}$ for $\theta = 1$, and $(\alpha/\kappa)^n$ for $\theta > 1$. Thus for $\theta \geq 1$, we recover the order of growth in $n$ of the constants given in Corollary 6.1; whereas for $\theta < 1$, the constant given in Corollary 6.1 is independent of $n$.

The OU process does not satisfy the Doeblin condition $D_0$, as Rosenblatt observes; see [21, p. 214].

**Acknowledgements.** This research was supported in part by the European Network PHD, MCRN -511953. The authors thank Professors P.J. Diggle, M. Ledoux, K. Marton and C. Villani for helpful conversations.

## References


[1] D. Bakry and M. Émery. Diffusions hypercontractive, in: *Séminaire de probabilités XIX*, Lecture Notes in Math. 1123, Springer, Berlin 1985, 177–206.

[2] G. Blower. The Gaussian isoperimetric inequality and transportation. *Positivity 7* (2003), 203–224.

[3] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal. 163* (1999), 1–28.

[4] S.G. Bobkov, I. Gentil and M. Ledoux. Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures et Appl. 80*, 7 (2001), 669–696.

[5] F. Bolley and C. Villani. Weighted Csiszár–Kulback–Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math.* To appear.

[6] P. Cattiaux and A. Guillin. Talagrand's like quadratic transportation cost inequalities. Preprint (2005).

[7] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. Invent. Math. 146 (2001), 219–257.

[8] R.L. Dobrushin and S.B. Shlosman. Constructive criterion for the uniqueness of Gibbs field, in: *Statistical Physics and Dynamical Systems (Rigorous Results)*, J. Fritz, A. Jaffe and D. Szász (eds.), Birkhäuser, Boston, 1985, 347–370.

[9] H. Djellout, A. Guillin and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann.Probab. 32* (2004), 2702–2732.

[10] H. Föllmer. Stock price fluctuation as a diffusion in a random environment, in: *Mathematical Models in Finance*, S.D. Howison, F.P. Kelly and P. Wilmott (eds.), Chapman and Hall, London, 1995, 21–33.

[11] L. Gross. Logarithmic Sobolev inequalities. *Amr. J. Math. 97* (1975), 1061–1083.

[12] M. Ledoux. The Concentration of Measure Phenomenon. American Mathematical Society, Providence, RI, 2001.

[13] K. Marton. Bounding $\overline{d}$-distance by information divergence: a method to prove measure concentration. *Ann. Probab. 24* (1996), 857–866.

[14] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal. 6* (1996), 556–571.

[15] K. Marton. Measure concentration for Euclidean distance in the case of dependent random variables. *Ann. Probab. 32* (2004), 2526–2544.

[16] P. Massart. Saint Flour Lectures Notes. `http://www.math.u-psud.fr/~massart`, 2003.

[17] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal. 11* (2001), 589–608.

[18] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal. 173* (2000), 361–400.

[19] V. I. Paulsen. Completely Bounded Maps and Dilations. Longman Science and Technical, Harlow, 1986.

[20] S. T. Rachev. Probability Metrics and the Stability of Stochastic Models. John Wiley and Sons Ltd., Chichester, 1991.

[21] M. Rosenblatt. Markov Processes: Structure and Asymptotic Behavior. Springer–Verlag, Heidelberg, 1971.

[22] P. -M. Samson. Concentration of measure inequalities for Markov chains and $\phi$-mixing processes. *Ann. Probab. 28* (2000), 416–461.

[23] M. Talagrand. Transportation cost for gaussian and other product measures. *Geom. Funct. Anal. 6* (1996), 587–600.

[24] C. Villani. Topics in Optimal Transportation. American Mathematical Society, Providence, RI, 2003.

[25] F. Y. Wang Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. *Prob. Th. Rel. Fields 108* (1997), 417–424.

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK

*E-mail address*: `g.blower@lancaster.ac.uk`

ENS Lyon, Umpa, 46 allée d'Italie, F-69364 Lyon Cedex 07

*E-mail address*: `fbolley@umpa.ens-lyon.fr`