

## Concentration of measure on product spaces with applications to Markov processes

by

GORDON BLOWER (Lancaster) and FRANÇOIS BOLLEY (Lyon)

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

$$\int_X \exp(tF(x)) \mu(dx) \leq \exp\left(t \int_X F(x) \mu(dx) + \kappa t^2/2\right) \quad (t \in \mathbb{R})$$

for all 1-Lipschitz functions  $F : (X, d) \rightarrow \mathbb{R}$  (see [4]). Recall that a function  $g : (\Omega_1, d_1) \rightarrow (\Omega_2, d_2)$  between metric spaces is *L-Lipschitz* if  $d_2(g(x), g(y)) \leq Ld_1(x, y)$  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 some hypotheses on the initial distribution  $P^{(1)}$  of  $\xi_1$  and the conditional distributions  $p_k(\cdot | x^{(k-1)})$  of  $\xi_k$  given  $\xi^{(k-1)}$ ;

---

2000 *Mathematics Subject Classification*: Primary 60E15; Secondary 60E05, 39B62.  
*Key words and phrases*: logarithmic Sobolev inequality, optimal transportation.

we recall that  $P^{(n)}$  is given by

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

If the  $(\xi_j)_{j=1}^n$  are mutually independent, and the distribution of each  $\xi_j$  satisfies  $\text{GC}(\kappa)$ , then  $P^{(n)}$  on  $(X^n, d^{(1)})$  is the product of the marginal distributions, and inherits  $\text{GC}(n\kappa)$  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  $1 \leq s < \infty$ ,  $\text{Prob}_s(X)$  denotes the subspace of  $\text{Prob}(X)$  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  $\text{Prob}_s(X)$  by

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

where  $\pi \in \text{Prob}_s(X \times X)$  has marginals  $\pi_1 = \mu$  and  $\pi_2 = \nu$ . Then  $W_s$  defines a metric on  $\text{Prob}_s(X)$ , 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 a homogeneous Markov process with state space  $X$ , initial distribution  $P^{(1)}$  and transition measure  $p(\cdot | x)$ . Suppose that there exist constants  $\kappa_1$  and  $L$  such that:*

- (i)  $P^{(1)}$  and  $p(\cdot | x)$  ( $x \in X$ ) satisfy  $\text{GC}(\kappa_1)$  on  $(X, d)$ ;
- (ii)  $x \mapsto p(\cdot | x)$  is  $L$ -Lipschitz  $(X, d) \rightarrow (\text{Prob}_1(X), W_1)$ .

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

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

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  $\text{Prob}(X)$ , 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

$$\text{Ent}(\nu | \mu) = \int_X \log \frac{d\nu}{d\mu} d\nu;$$

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

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

$$W_s(\nu, \mu) \leq \left( \frac{2}{\alpha} \text{Ent}(\nu | \mu) \right)^{1/2}$$

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

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 [4] that  $\text{GC}(\kappa)$  is equivalent to  $T_1(1/\kappa)$ ; their proof used the Kantorovich–Rubinstein duality result, that

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

where  $\mu, \nu \in \text{Prob}_1(X)$  and  $f : X \rightarrow \mathbb{R}$  runs over the set of 1-Lipschitz functions. A  $\nu \in \text{Prob}(X)$  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, 8] for detailed statements. Moreover, since  $T_s(\alpha)$  implies  $T_r(\alpha)$  for  $1 \leq r \leq s$  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 [8] developed Marton's coupling method [14, 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 a homogeneous Markov process with state space  $X$ , initial distribution  $P^{(1)}$  and transition measure  $p(\cdot | x)$ . Suppose that there exist constants  $1 \leq s \leq 2$ ,  $\alpha > 0$  and  $L \geq 0$  such that:*

- (i)  $P^{(1)}$  and  $p(\cdot | x)$  ( $x \in X$ ) satisfy  $T_s(\alpha)$ ;
- (ii)  $x \mapsto p(\cdot | x)$  is  $L$ -Lipschitz  $(X, d) \rightarrow (\text{Prob}_s(X), W_s)$ .

Then the joint distribution  $P^{(n)}$  of  $(\xi_1, \dots, \xi_n)$  satisfies  $T_s(\alpha_n)$ , where

$$\alpha_n = \begin{cases} n^{1-2/s}(1-L^{1/s})^2\alpha, & L < 1, \\ e^{2/s-2}n^{-2/s-1}\alpha, & L = 1, \\ \left(\frac{L-1}{e^{s-1}L^n}\right)^{2/s} \frac{\alpha}{n+1}, & L > 1; \end{cases}$$

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 a hypothesis related to Dobrushin–Shlosman’s mixing condition [9, 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

$$\int_{\mathbb{R}^m} f^2 \log\left(f^2 / \int_{\mathbb{R}^m} f^2 d\mu\right) d\mu \leq \frac{2}{\alpha} \int_{\mathbb{R}^m} \|\nabla f\|^2 d\mu$$

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

The connections between the various inequalities are summarized by

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

see [4; 18; 24, p. 293]. Conversely, Otto and Villani showed that if  $\mu(dx) = e^{-V(x)}dx$  satisfies  $T_2(\alpha)$  where  $V : \mathbb{R}^m \rightarrow \mathbb{R}$  is convex, then  $\mu$  also satisfies LSI( $\alpha/4$ ) (see [3; 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 Émery [1] showed that if  $V$  is twice continuously differentiable, with  $\text{Hess } V \geq \alpha I_m$  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 [4] characterized in terms of their cumulative distribution functions those  $\mu \in \text{Prob}(\mathbb{R})$  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 a homogeneous Markov process with state space  $\mathbb{R}^m$ , initial distribution  $P^{(1)}$  and transition measure  $p(dy|x) = e^{-u(x,y)}dy$ . Suppose that there exist constants  $\alpha > 0$  and  $L \geq 0$  such that:*

- (i)  $P^{(1)}$  and  $p(\cdot|x)$  ( $x \in \mathbb{R}^m$ ) satisfy  $\text{LSI}(\alpha)$ ;
- (ii)  $u$  has bounded second-order partial derivatives and the off-diagonal blocks of its Hessian matrix satisfy

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

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

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

$$\alpha_n = \begin{cases} \frac{(\alpha - L)^2}{\alpha}, & 0 \leq L < \alpha, \\ \frac{\alpha}{n(n+1)(e-1)}, & 0 < L = \alpha, \\ \left(\frac{\alpha}{L}\right)^{2n} \frac{(L^2 - \alpha^2)}{\alpha e(n+1)}, & 0 < \alpha < L; \end{cases}$$

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  $\text{LSI}(\alpha)$  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  $n \rightarrow \infty$ , 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  $p_k(\cdot|x^{(k-1)})$  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  $M \geq \varrho_\ell \geq 0$  ( $\ell = 1, \dots, n$ ) such that:*

- (i)  $P^{(1)}$  and  $p_k(\cdot|x^{(k-1)})$  ( $k = 2, \dots, n$ ;  $x^{(k-1)} \in X^{k-1}$ ) satisfy  $T_s(\alpha)$  on  $(X, d)$ ;
- (ii)  $x^{(k-1)} \mapsto p_k(\cdot|x^{(k-1)})$  is Lipschitz as a map  $(X^{k-1}, d^{(s)}) \rightarrow (\text{Prob}_s(X), W_s)$  for  $k = 2, \dots, n$ , in the sense that

$$W_s(p_k(\cdot | x^{(k-1)}), p_k(\cdot | y^{(k-1)}))^s \leq \sum_{j=1}^{k-1} \varrho_{k-j} d(x_j, y_j)^s$$

for all  $x^{(k-1)}, y^{(k-1)} \in X^{k-1}$ . Then  $P^{(n)}$  satisfies the transportation inequality  $T_s(\alpha_n)$  where

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

Suppose further that

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

Then the joint distribution  $P^{(n)}$  on the space  $(X^n, d^{(s)})$  satisfies  $T_s(\alpha_n)$  where

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

In hypothesis (iii), the sequence  $(\varrho_k)_{k=1}^{n-1}$  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  $(\varrho_k)_{k=1}^{n-1}$  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 [8, 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 measures considered.

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

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

for all bounded and continuous  $f : X \rightarrow \mathbb{R}$ . Brenier and McCann showed that if  $\mu$  and  $\nu$  belong to  $\text{Prob}_2(\mathbb{R}^m)$ , and if moreover  $\mu$  is absolutely continuous with respect to Lebesgue measure, then there exists a convex function  $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$  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

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

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 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  $p_j(dx_j | x^{(j-1)})$  ( $j = 2, \dots, n$ ) are all absolutely continuous with respect to Lebesgue measure. Then let  $Q^{(n)} \in \text{Prob}_s(\mathbb{R}^{nm})$  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

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

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

$$(2.1) \quad \text{Ent}(Q^{(n)} | P^{(n)}) = \text{Ent}(Q^{(1)} | P^{(1)}) + \sum_{j=2}^n \int_{\mathbb{R}^{(j-1)m}} \text{Ent}(q_j(\cdot | x^{(j-1)}) | p_j(\cdot | x^{(j-1)}))Q^{(j-1)}(dx^{(j-1)}).$$

When 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  $p_j(dx_j | x^{(j-1)})$  satisfy GC( $\kappa$ ) for  $\kappa = 1/\alpha$ , and then one can check that there exists  $\varepsilon > 0$  such that

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

and likewise for  $p_j$ ; compare with Herbst’s theorem [24, p. 280], and [4, 8]. Hence  $Q^{(1)}$  and  $q_j(dx_j | x^{(j-1)})$  for  $Q^{(j-1)}$ -almost every  $x^{(j-1)}$  have finite second moments, since Young’s inequality  $ab \leq a \log a - a + e^b$  on  $\mathbb{R}$  ensures that

$$\begin{aligned} & \int_{\mathbb{R}^m} \varepsilon\|x^{(1)}\|^2 Q^{(1)}(dx^{(1)}) \\ & \leq \text{Ent}(Q^{(1)} | P^{(1)}) + \log \int_{\mathbb{R}^m} \exp(\varepsilon\|x^{(1)}\|^2) P^{(1)}(dx^{(1)}) < \infty, \end{aligned}$$

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

Let  $\theta_1 : \mathbb{R}^m \rightarrow \mathbb{R}^m$  be an optimal transportation map that induces  $P^{(1)}(dx_1)$  from  $Q^{(1)}(dx_1)$ ; then for  $Q^{(1)}$ -almost every  $x_1$ , let  $x_2 \mapsto \theta_2(x_1, x_2)$  induce  $p_2(dx_2 | \theta_1(x_1))$  from  $q_2(dx_2 | x_1)$  optimally; hence  $\Theta^{(2)} : \mathbb{R}^{2m} \rightarrow \mathbb{R}^{2m}$ , 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  $\Theta^{(j)} : \mathbb{R}^{jm} \rightarrow \mathbb{R}^{jm}$ , we let  $x_{j+1} \mapsto \theta_{j+1}(x^{(j)}, x_{j+1})$  be an optimal transportation map that induces  $p_{j+1}(dx_{j+1} | \Theta^{(j)}(x^{(j)}))$  from  $q_{j+1}(dx_{j+1} | x^{(j)})$  for all  $x^{(j)}$  in a certain set of full  $Q^{(j)}$ -measure; then we let  $\Theta^{(j+1)} : \mathbb{R}^{(j+1)m} \rightarrow \mathbb{R}^{jm} \times \mathbb{R}^m$  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

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

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

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

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

$$d_1 = W_s(P^{(1)}, Q^{(1)})$$

and, for  $k = 2, \dots, n$ ,

$$(2.3) \quad d_k = \int_{\mathbb{R}^{(k-1)m}} W_s(p_k(\cdot | \Theta^{(k-1)}(x^{(k-1)})), q_k(\cdot | x^{(k-1)}))^s Q^{(k-1)}(dx^{(k-1)}).$$

Given  $a, b > 0$ ,  $1 \leq s \leq 2$  and  $\gamma > 1$ , we have

$$(a + b)^s \leq \left( \frac{\gamma}{\gamma - 1} \right)^{s-1} a^s + \gamma^{s-1} b^s.$$

Hence by the triangle inequality, the expression (2.3) is bounded by

$$(2.4) \quad \left( \frac{\gamma}{\gamma - 1} \right)^{s-1} \int_{\mathbb{R}^{m(k-1)}} W_s(p_k(\cdot | x^{(k-1)}), q_k(\cdot | x^{(k-1)}))^s Q^{(k-1)}(dx^{(k-1)}) \\ + \gamma^{s-1} \int_{\mathbb{R}^{m(k-1)}} W_s(p_k(\cdot | \Theta^{(k-1)}(x^{(k-1)})), p_k(\cdot | x^{(k-1)}))^s Q^{(k-1)}(dx^{(k-1)}).$$

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

in (2.4) by

$$h_k = \left( \frac{\gamma}{\gamma-1} \right)^{s-1} \left( \frac{2}{\alpha} \right)^{s/2} \left( \int_{\mathbb{R}^{(k-1)m}} \text{Ent}(q_k | p_k) dQ^{(k-1)} \right)^{s/2}.$$

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

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

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

$$(2.5) \quad d_k \leq h_k + \gamma^{s-1} \sum_{j=1}^{k-1} \varrho_{k-j} d_j$$

for  $k = 2, \dots, n$ . For  $k = 1$  we simply have  $d_1 \leq h_1$  by hypothesis (i), where

$$h_1 = \left( \frac{\gamma}{\gamma-1} \right)^{s-1} \text{Ent}(Q^{(1)} | P^{(1)}).$$

In the case when the  $\varrho_\ell$  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

$$\begin{aligned} \sum_{k=1}^n d_k &\leq \sum_{j=1}^n h_j (1 + \gamma^{s-1} M)^{n-j} \\ &\leq \left( \sum_{j=1}^n h_j^{2/s} \right)^{s/2} \left( \sum_{j=1}^n (1 + \gamma^{s-1} M)^{2(n-j)/(2-s)} \right)^{(2-s)/2} \end{aligned}$$

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

$$\left( \sum_{j=1}^n h_j^{2/s} \right)^{s/2} = \left( \frac{\gamma}{\gamma-1} \right)^{s-1} \left( \frac{2}{\alpha} \right)^{s/2} \text{Ent}(Q^{(n)} | P^{(n)})^{s/2}$$

by (2.1), whereas the second term is bounded by

$$\frac{(1 + M\gamma^{s-1})^n}{((1 + M\gamma^{s-1})^{2/(2-s)} - 1)^{(2-s)/2}} \leq \frac{\gamma^{n(s-1)}(1 + M)^n}{M\gamma^{s-1}} \leq \frac{(1 + M)^n}{M} e^{s-1}$$

with  $\gamma = 1 + 1/n$ . Collecting both terms we finally obtain by (2.2) the stated result

$$W_s(Q^{(n)}, P^{(n)})^s \leq \left( \frac{2}{\alpha} \right)^{s/2} \frac{(1 + M)^n}{M} e^{s-1} \text{Ent}(Q^{(n)} | P^{(n)})^{s/2}.$$

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

$$T_{m+1} \leq \sum_{j=1}^{m+1} h_j + \gamma^{s-1} R T_m;$$

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

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

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

$$\begin{aligned} (2.6) \quad W_s(Q^{(n)}, P^{(n)})^s &\leq \left( \frac{\gamma}{\gamma-1} \right)^{s-1} \left( \sum_{m=1}^n \left( \sum_{\ell=0}^{m-1} (\gamma^{s-1} R)^\ell \right)^{2/(2-s)} \right)^{(2-s)/2} \\ &\quad \times \left( \frac{2}{\alpha} \text{Ent}(Q^{(n)} | P^{(n)}) \right)^{s/2} \end{aligned}$$

$$(2.7) \quad \leq \left( \frac{\gamma}{\gamma-1} \right)^{s-1} n^{1-s/2} \sum_{\ell=0}^{n-1} (\gamma^{s-1} R)^\ell \left( \frac{2}{\alpha} \text{Ent}(Q^{(n)} | P^{(n)}) \right)^{s/2};$$

this also holds for  $s = 2$ . Finally, we select  $\gamma$  according to the value of  $R$  so as 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

$$W_s(Q^{(n)}, P^{(n)})^s \leq \left( \frac{2}{\alpha} \right)^{s/2} \frac{n^{1-s/2}}{(1 - R^{1/s})^s} \text{Ent}(Q^{(n)} | P^{(n)})^{s/2}.$$

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

$$\begin{aligned} W_s(Q^{(n)}, P^{(n)})^s &\leq \left( \frac{2}{\alpha} \right)^{s/2} (n+1)^{s-1} n^{1-s/2} \left( \frac{(1 + 1/n)^{n(s-1)} R^n - 1}{(1 + 1/n)^{s-1} R - 1} \right) \\ &\quad \times \text{Ent}(Q^{(n)} | P^{(n)})^{s/2}, \end{aligned}$$

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 [8]. Given  $j = 1, \dots, n-1$ , assume that

$\pi^{(j)} \in \text{Prob}(X^{2j})$  has marginals  $Q^{(j)}(dx^{(j)})$  and  $P^{(j)}(dy^{(j)})$  and satisfies

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

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

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

Now we let

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

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

$$\begin{aligned} W_s(Q^{(n)}, P^{(n)})^s \\ \leq \sum_{j=1}^n \int_{X^{2(j-1)}} W_s(q_j(\cdot | x^{(j-1)}), p_j(\cdot | y^{(j-1)}))^s \pi^{(j-1)}(dx^{(j-1)} dy^{(j-1)}) \end{aligned}$$

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  $\varrho_1 = L$  and  $\varrho_j = 0$  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  $\varrho_\ell$  when  $(X, d) = (\mathbb{R}^m, \ell^s)$ .

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

- (i)  $p_j(dx_j | x^{(j-1)}) = \exp(-u_j(x^{(j)})) dx_j$  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

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

where  $1/s' + 1/s = 1$  and  $\partial/\partial x_k$  is the gradient with respect to  $x_k$ .

Then the mapping  $x^{(j-1)} \mapsto p_j(\cdot | x^{(j-1)})$  is  $\sqrt{M_s/\alpha}$ -Lipschitz  $(\mathbb{R}^{m(j-1)}, \ell^s) \rightarrow (\text{Prob}_s(\mathbb{R}^m), W_s)$ .

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

$$f(t) = W_s(p_j(\cdot | x^{(j-1)}(t)), p_j(\cdot | \bar{x}^{(j-1)}));$$

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

By the triangle inequality and (i), we have

$$\begin{aligned} (2.8) \quad & \left( \frac{f(t+\delta) - f(t)}{\delta} \right)^2 \\ & \leq \frac{1}{\delta^2} W_s(p_j(\cdot | x^{(j-1)}(t+\delta)), p_j(\cdot | x^{(j-1)}(t)))^2 \\ & \leq \frac{1}{\alpha\delta^2} \{ \text{Ent}(p_j(\cdot | x^{(j-1)}(t+\delta)) | p_j(\cdot | x^{(j-1)}(t))) \\ & \quad + \text{Ent}(p_j(\cdot | x^{(j-1)}(t)) | p_j(\cdot | x^{(j-1)}(t+\delta))) \} \\ & = \frac{1}{\alpha\delta^2} \int_{\mathbb{R}^m} (u_j(x^{(j-1)}(t+\delta), x_j) - u_j(x^{(j-1)}(t), x_j)) \\ & \quad \times \{ \exp(-u_j(x^{(j-1)}(t), x_j)) - \exp(-u_j(x^{(j-1)}(t+\delta), x_j)) \} dx_j. \end{aligned}$$

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

$$\begin{aligned} & u_j(x^{(j-1)}(t+\delta), x_j) - u_j(x^{(j-1)}(t), x_j) \\ & = \delta \sum_{k=1}^{j-1} \left\langle \frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j), x_k - \bar{x}_k \right\rangle \\ & \quad + \frac{\delta^2}{2} \langle (\text{Hess } u_j)(x^{(j-1)} - \bar{x}^{(j-1)}), (x^{(j-1)} - \bar{x}^{(j-1)}) \rangle, \end{aligned}$$

where  $\text{Hess } u_j$  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 in (2.8), we obtain

$$\begin{aligned} & \limsup_{\delta \rightarrow 0^+} \left( \frac{f(t+\delta) - f(t)}{\delta} \right)^2 \\ & \leq \frac{1}{\alpha} \int_{\mathbb{R}^m} \left( \sum_{k=1}^{j-1} \left\langle \frac{\partial u_j}{\partial x_k}(x^{(j-1)}(t), x_j), x_k - \bar{x}_k \right\rangle \right)^2 p_j(dx_j | x^{(j-1)}(t)). \end{aligned}$$

Hence by Hölder's inequality we have

$$\begin{aligned}
 & \limsup_{\delta \rightarrow 0^+} \frac{|f(t + \delta) - f(t)|}{\delta} \\
 & \leq \frac{1}{\sqrt{\alpha}} \left( \int_{\mathbb{R}^n} \left( \sum_{k=1}^{j-1} \left| \frac{\partial u_j}{\partial x_k} (x^{(j-1)}(t), x_j) \right|^{s'} \right)^{2/s'} p_j(dx_j | x^{(j-1)}(t)) \right)^{1/2} \\
 & \quad \times \|x^{(j-1)} - \bar{x}^{(j-1)}\|_{\ell^s}
 \end{aligned}$$

for  $1 < s \leq 2$ , and likewise with obvious changes for  $s = 1$ . By assumption (ii) and Vitali's theorem, the function  $f$  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  $M \geq \varrho_j \geq 0$  ( $j = 1, \dots, n$ ) such that:*

- (i)  $P^{(1)}$  and  $p_k(\cdot | x^{(k-1)})$  ( $k = 2, \dots, n$ ;  $x^{(k-1)} \in X^{k-1}$ ) satisfy  $\text{GC}(\kappa_1)$  on  $(X, d)$ ;
- (ii)  $x^{(k-1)} \mapsto p_k(\cdot | x^{(k-1)})$  is Lipschitz as a map  $(X^{k-1}, d^{(1)}) \rightarrow (\text{Prob}_1(X), W_1)$  for  $k = 2, \dots, n$ , in the sense that

$$W_1(p_k(\cdot | x^{(k-1)}), p_k(\cdot | y^{(k-1)})) \leq \sum_{j=1}^{k-1} \varrho_{k-j} d(x_j, y_j)$$

for all  $x^{(k-1)}, y^{(k-1)} \in X^{k-1}$ .

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

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

Suppose moreover that

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

Then  $P^{(n)}$  satisfies  $\text{GC}(\kappa_n(R))$  on  $(X^n, d^{(1)})$ , where

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

*Proof.* This follows from the Bobkov–Götze theorem [4] 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  $\varrho_1 = L$  and  $\varrho_j = 0$  for  $j = 2, \dots, n$ , which satisfy (iii). ■

**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  $\nu \in \text{Prob}(\mathbb{R}^m)$  satisfy  $\text{LSI}(\alpha)$ , 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  $\text{LSI}(\alpha/L^2)$ . 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  $\text{LSI}(\alpha)$ . Then for any  $L$ -Lipschitz map  $\Theta : \mathbb{R}^m \rightarrow \mathbb{R}^m$ , the relation*

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

*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  $\text{LSI}(\alpha_n)$  where*

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

*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

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

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

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

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

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

By summing over  $k$ , one notes that  $\varphi_n$  defines a Lipschitz function from

$(\mathbb{R}^{nm}, \ell^2)$  into itself, with Lipschitz seminorm

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

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

$$L_{\varphi_n} \leq \begin{cases} [n(n+1)(e-1)]^{1/2} & \text{for } L = 1, \\ [e(n+1)L^{2n}(L-1)^{-1}]^{1/2} & \text{for } L > 1. \end{cases}$$

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( $\alpha_n$ ), where  $\alpha_n = L_{\varphi_n}^{-2}\alpha$ . ■

The linear case gives the following result for ARMA processes. Here  $\|M\|$  denotes the operator norm between Euclidean spaces of a matrix  $M$ .

**PROPOSITION 4.2.** *Let  $A$  and  $B$  be  $m \times m$  matrices such that the spectral radius  $\varrho$  of  $A$  satisfies  $\varrho < 1$ . 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*

$$(4.4) \quad Z_{j+1} = AZ_j + BY_{j+1} \quad (j = 0, 1, \dots),$$

satisfies LSI( $\alpha$ ) where

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

*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  $A = \sqrt{\varrho} S^{-1}CS$ ; one can choose the similarity so that the operator norms satisfy

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

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

$$(4.5) \quad SZ_{j+1} = \sqrt{\varrho} CSZ_j + SBY_{j+1} \quad (j = 0, 1, \dots)$$

which involves the  $\sqrt{\varrho}$ -Lipschitz linear map  $\Theta : \mathbb{R}^m \rightarrow \mathbb{R}^m$ ,  $\Theta(w) = \sqrt{\varrho}Cw$ . Given  $n \geq 1$ , the linear map  $\Phi_n : \mathbb{R}^{nm} \rightarrow \mathbb{R}^{nm}$ , defined to solve (4.5) by

$$\begin{aligned} (z_0, y_1, \dots, y_n) &\mapsto (Sz_0, SBY_1, \dots, SBY_{n-1}) \\ &\mapsto (Sz_0, Sz_1, \dots, Sz_{n-1}) \mapsto (z_0, z_1, \dots, z_{n-1}), \end{aligned}$$

has operator norm

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

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  $\varrho < 1$ .

(ii) The joint distribution of the ARMA process is discussed by Djellout, Guillin and Wu [8, Section 3]. We have improved upon [8] by obtaining LSI( $\alpha$ ), hence  $T_2(\alpha)$ , under the spectral radius condition  $\varrho < 1$ , where  $\alpha$  is independent of the size  $n$  of the sample considered 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

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

The coupling between variables is measured by the integral

$$A_{j,k}(s) = \sup_{x^{(j-1)}} \int_{\mathbb{R}} \exp\left(\left\langle s, \frac{\partial u_j}{\partial x_k}(x^{(j)}) \right\rangle\right) p_j(dx_j | x^{(j-1)})$$

for all  $s \in \mathbb{R}^m$  and  $1 \leq k < j \leq n$ , 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  $p_k(\cdot | x^{(k-1)})$  ( $k = 2, \dots, n$ ;  $x^{(k-1)} \in \mathbb{R}^{m(k-1)}$ ) satisfy LSI( $\alpha$ );
- (ii)  $A_{j,k}(s) \leq \exp(\kappa_{j,k} \|s\|^2/2)$  for all  $s \in \mathbb{R}^m$ .

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

$$(5.1) \quad \alpha_n = \frac{\alpha}{1 + \varepsilon} \prod_{j=1}^{n-1} (1 + K_j)^{-1}$$

for all  $\varepsilon > 0$ , where

$$K_j = (1 + \varepsilon^{-1}) \sum_{l=0}^{j-1} \kappa_{n-l, n-j} / \alpha \quad \text{for } j = 1, \dots, n-1.$$

Suppose further that there exist  $R \geq 0$  and  $\varrho_l \geq 0$  for  $l = 1, \dots, n-1$  such that

$$(iii) \quad \kappa_{j,k} \leq \varrho_{j-k} \text{ for } 1 \leq k < j \leq n, \text{ and } \sum_{l=1}^{n-1} \sqrt{\varrho_l} \leq \sqrt{R}.$$

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

$$\alpha_n = \begin{cases} (\sqrt{\alpha} - \sqrt{R})^2, & 0 \leq R < \alpha, \\ \frac{\alpha}{n(n+1)(e-1)}, & 0 < R = \alpha, \\ \left(\frac{\alpha}{R}\right)^n \frac{R - \alpha}{e(n+1)}, & 0 < \alpha < R. \end{cases}$$

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 [4].

PROPOSITION 5.2. *In the above notation, let  $1 \leq k < j$  and suppose that there exist  $\alpha > 0$  and  $L_{j,k} \geq 0$  such that:*

- (i)  $p_j(\cdot | x^{(j-1)})$  satisfies  $\text{GC}(1/\alpha)$  for all  $x^{(j-1)} \in \mathbb{R}^{m(j-1)}$ ;
- (ii)  $u_j$  has bounded second order partial derivatives and the off-diagonal block of its Hessian matrix satisfies

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

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

Then

$$A_{j,k}(s) \leq \exp(L_{j,k}^2 \|s\|^2 / (2\alpha)) \quad (s \in \mathbb{R}^m).$$

*Proof.* Letting  $e$  be the unit vector in the  $k$ th coordinate direction, 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. Then by the mean-value theorem and (ii) we have,

for  $t > 0$ ,

$$\begin{aligned} 0 &= \int_{\mathbb{R}^m} \frac{1}{t} (p_j(dx_j | x^{(j-1)} + te) - p_j(dx_j | x^{(j-1)})) \\ &= \int_{\mathbb{R}^m} \frac{1}{t} (\exp(u_j(x_j, x^{(j-1)} + te) - u_j(x_j, x^{(j-1)})) - 1) p_j(dx_j | x^{(j-1)}) \\ &= \int_{\mathbb{R}^m} \frac{1}{t} \left( \exp\left(-t \left\langle e, \frac{\partial u_j}{\partial x_k}(x_j, x^{(j-1)}) \right\rangle + O(t^2)\right) - 1 \right) p_j(dx_j | x^{(j-1)}); \end{aligned}$$

here the implied constants in  $O(t^2)$  are independent of  $x^{(j)}$ . Now by GC( $1/\alpha$ ) for  $p_j(dx_j | x^{(j-1)})$ , we have

$$0 \leq \frac{1}{t} \left( \exp\left\{-t \int_{\mathbb{R}^m} \left\langle e, \frac{\partial u_j}{\partial x_k}(x_j, x^{(j-1)}) \right\rangle p_j(dx_j | x^{(j-1)}) + O(t^2)\right\} - 1 \right);$$

so letting  $t \rightarrow 0+$ , we deduce that

$$0 \geq \int_{\mathbb{R}^m} \left\langle \frac{\partial u_j}{\partial x_k}(x_j, x^{(j-1)}), e \right\rangle p_j(dx_j | x^{(j-1)}),$$

and likewise with  $-e$  instead of  $e$ . Hence by symmetry we have

$$0 = \int_{\mathbb{R}^m} \left\langle \frac{\partial u_j}{\partial x_k}(x_j, x^{(j-1)}), e \right\rangle p_j(dx_j | x^{(j-1)}).$$

Then, by (i),

$$\int_{\mathbb{R}^m} \exp\left(t \left\langle e, \frac{\partial u_j}{\partial x_k} \right\rangle\right) p_j(dx_j | x^{(j-1)}) \leq \exp(L_{j,k}^2 t^2 / (2\alpha))$$

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

*Proof of Theorem 5.1.* For notational convenience,  $X$  denotes the state space  $\mathbb{R}^m$ . Then let  $f : X^n \rightarrow \mathbb{R}$  be a smooth and compactly supported function, and let  $g_j : X^{n-j} \rightarrow \mathbb{R}$  be defined by  $g_0 = f$  and by

$$(5.2) \quad g_j(x^{(n-j)}) = \left( \int_X g_{j-1}(x^{(n-j+1)})^2 p_{n-j+1}(dx_{n-j+1} | x^{(n-j)}) \right)^{1/2}$$

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

$$(5.3) \quad \int_{X^n} f^2 \log\left(f^2 / \int_{X^n} f^2 dP^{(n)}\right) dP^{(n)} = \sum_{j=0}^{n-1} \int_{X^{n-j}} g_j^2 \log(g_j^2 / g_{j+1}^2) dP^{(n-j)},$$

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  $dp_{n-j} = p_{n-j}(dx_{n-j} | x^{(n-j-1)})$  satisfies LSI( $\alpha$ ), whence

$$(5.4) \quad \int_X g_j^2 \log(g_j^2/g_{j+1}^2) dp_{n-j} \leq \frac{2}{\alpha} \int_X \left( \frac{\partial g_j}{\partial x_{n-j}} \right)^2 dp_{n-j} \quad (j = 0, \dots, n-1),$$

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

$$(5.5) \quad g_j \frac{\partial g_j}{\partial x_{n-j}} = \int_{X^j} f \frac{\partial f}{\partial x_{n-j}} dp_n \dots dp_{n-j+1} \\ - \frac{1}{2} \sum_{l=0}^{j-1} \int_{X^{j-l}} g_l^2 \frac{\partial u_{n-l}}{\partial x_{n-j}} dp_{n-l} \dots dp_{n-j+1},$$

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

$$(5.6) \quad \left\| \int_X g_l^2 \frac{\partial u_{n-l}}{\partial x_{n-j}} dp_{n-l} \right\| \leq g_{l+1} \left( 2\kappa_{n-l, n-j} \int_X g_l^2 \log(g_l^2/g_{l+1}^2) dp_{n-l} \right)^{1/2}.$$

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

$$\int_X \exp \left( \left\langle s, \frac{\partial u_{n-l}}{\partial x_{n-j}} \right\rangle - \log \Lambda_{n-l, n-j}(s) \right) dp_{n-l} \leq 1 \quad (s \in X),$$

and hence by the dual formula for relative entropy, as in [3, p. 693],

$$\int_X \left( \left\langle s, \frac{\partial u_{n-l}}{\partial x_{n-j}} \right\rangle - \log \Lambda_{n-l, n-j}(s) \right) g_l^2 dp_{n-l} \leq \int_X g_l^2 \log(g_l^2/g_{l+1}^2) dp_{n-l}.$$

Then hypothesis (ii) of the theorem ensures that

$$\left\langle s, \int_X \frac{\partial u_{n-l}}{\partial x_{n-j}} g_l^2 dp_{n-l} \right\rangle \leq \frac{\|s\|^2}{2} \kappa_{n-l, n-j} g_{l+1}^2 + \int_X g_l^2 \log(g_l^2/g_{l+1}^2) dp_{n-l}$$

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

$$\left\| \int_{X^{j-l}} g_l^2 \frac{\partial u_{n-l}}{\partial x_{n-j}} dp_{n-l} \dots dp_{n-j+1} \right\| \\ \leq g_j \left( 2\kappa_{n-l, n-j} \int_{X^{j-l}} g_l^2 \log(g_l^2/g_{l+1}^2) dp_{n-l} \dots dp_{n-j+1} \right)^{1/2}.$$

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

$$(5.7) \quad \int_{X^{n-j}} \left\| \frac{\partial g_j}{\partial x_{n-j}} \right\|^2 dP^{(n-j)} \\ \leq (1 + \varepsilon) \int_{X^n} \left\| \frac{\partial f}{\partial x_{n-j}} \right\|^2 dP^{(n)} + \frac{1 + \varepsilon^{-1}}{4} \left\{ \sum_{l=0}^{j-1} (2\kappa_{n-l, n-j} h_l)^{1/2} \right\}^2$$

where  $\varepsilon > 0$  is arbitrary and  $h_l$  is given by

$$h_l = \int_{X^{n-l}} g_l^2 \log(g_l^2 / g_{l+1}^2) dP^{(n-l)}.$$

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

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

where we have let

$$d_j = \frac{2(1 + \varepsilon)}{\alpha} \int_{X^n} \left( \frac{\partial f}{\partial x_{n-j}} \right)^2 dP^{(n)} \quad (j = 0, \dots, n-1), \\ K_j = \frac{1 + \varepsilon^{-1}}{\alpha} \sum_{l=0}^{j-1} \kappa_{n-l, n-j} \quad (j = 1, \dots, n-1).$$

Since  $h_0 \leq d_0$  by (5.4) with  $j = 0$  and since all terms are positive, the partial sums  $H_k = \sum_{j=0}^k h_j$  satisfy the system of inequalities

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

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

$$H_{n-1} \leq d_{n-1} + \sum_{k=0}^{n-2} d_k \prod_{l=k+1}^{n-1} (1 + K_l),$$

which in turn implies the bound

$$H_{n-1} \leq \prod_{l=1}^{n-1} (1 + K_l) \sum_{j=0}^{n-1} d_j.$$

By (5.3) this is equivalent to the inequality

$$\int_{X^n} f^2 \log \left( f^2 / \int_{X^n} f^2 dP^{(n)} \right) dP^{(n)} \leq \frac{2(1 + \varepsilon)}{\alpha} \prod_{l=1}^{n-1} (1 + K_l) \int_{X^n} \|\nabla f\|_{l^2}^2 dP^{(n)}.$$

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

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

$$h_j \leq d_j + \frac{1 + \varepsilon^{-1}}{\alpha} \left( \sum_{l=0}^{j-1} \sqrt{\varrho_{j-l}} \sqrt{h_l} \right)^2$$

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

$$\sum_{j=0}^k h_j \leq \sum_{j=0}^k d_j + \frac{1 + \varepsilon^{-1}}{\alpha} \sum_{j=1}^k \left( \sum_{l=0}^{j-1} \sqrt{\varrho_{j-l}} \sqrt{h_l} \right)^2,$$

which implies by Young's convolution inequality that

$$\sum_{j=0}^k h_j \leq \sum_{j=0}^k d_j + \frac{1 + \varepsilon^{-1}}{\alpha} \left( \sum_{l=1}^k \sqrt{\varrho_l} \right)^2 \sum_{l=0}^{k-1} h_l.$$

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

$$H_k \leq D_k + \sum_{j=0}^{k-1} D_j \prod_{l=j+1}^k \frac{1 + \varepsilon^{-1}}{\alpha} R_l$$

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

$$\begin{aligned} (5.8) \quad H_{n-1} &\leq \left( 1 + \sum_{j=0}^{n-2} \left( \frac{1 + \varepsilon^{-1}}{\alpha} R \right)^{n-j-1} \right) D_{n-1} \\ &= \sum_{l=0}^{n-1} \left( \frac{1 + \varepsilon^{-1}}{\alpha} R \right)^l D_{n-1} \end{aligned}$$

since  $D_j \leq D_{n-1}$  and  $R_j \leq R$  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  $\text{LSI}(\alpha)$  for  $P^{(n)}$  as expected, since here  $P^{(n)}$  is the tensor product of its marginal distributions, which satisfy  $\text{LSI}(\alpha)$ .

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

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

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

$$\int_{X^n} f^2 \log \left( f^2 / \int_{X^n} f^2 dP^{(n)} \right) dP^{(n)} \leq \frac{2}{(\sqrt{\alpha} - \sqrt{R})^2} \int_{X^n} \|\nabla f\|_{\ell^2}^2 dP^{(n)}.$$

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

$$H_{n-1} \leq \frac{2(n+1)}{\alpha} \left( \frac{(1+1/n)^n (R/\alpha)^n - 1}{(1+1/n)(R/\alpha) - 1} \right) \int_{X^n} \|\nabla f\|_{\ell^2}^2 dP^{(n)};$$

as above this leads to the stated result by (5.3). ■

**6. Logarithmic Sobolev inequalities for Markov processes.** The results of the preceding section simplify considerably when we have a 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  $p(dy|x) = e^{-u(x,y)} dy$  where  $u$  is a twice continuously differentiable function such that

$$(6.1) \quad \Lambda(s|x) = \int_{\mathbb{R}^m} \exp\left(\left\langle s, \frac{\partial u}{\partial x}(x,y) \right\rangle\right) p(dy|x) < \infty \quad (s, x \in \mathbb{R}^m).$$

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  $p(\cdot|x)$  ( $x \in \mathbb{R}^m$ ) satisfy LSI( $\alpha$ );
- (ii)  $\Lambda(s|x) \leq \exp(\kappa\|s\|^2/2)$  for all  $s, x \in \mathbb{R}^m$ .

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

$$\alpha_n = \begin{cases} (\sqrt{\alpha} - \sqrt{\kappa})^2, & 0 \leq \kappa < \alpha, \\ \frac{\alpha}{n(n+1)(e-1)}, & 0 < \kappa = \alpha, \\ \left(\frac{\alpha}{\kappa}\right)^n \frac{\kappa - \alpha}{e(n+1)}, & 0 < \alpha < \kappa. \end{cases}$$

*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  $\varrho_1 = \kappa$  and  $\varrho_j = 0$  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 ( $\mathbb{R}^m, \ell^2$ )  $\rightarrow \mathbb{R}$  for any unit vector  $e$  in  $\mathbb{R}^m$ , and hence  $\Lambda(s|x) \leq \exp(\|s\|^2 L^2 / (2\alpha))$  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 coordinate

charts. McCann [17] has shown that a locally Lipschitz function on  $X$  is differentiable except on a set that has zero Riemannian volume; so the  $L$ -Lipschitz condition on  $f : X \rightarrow \mathbb{R}$  is essentially equivalent to  $\|\nabla f\| \leq L$ .

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

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

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

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

where  $(B_t^{(0)})$  is a real standard Brownian motion starting at 0, and  $\varrho \in \mathbb{R}$ . In financial modelling, OU processes with  $\varrho < 0$  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  $F_n : (\mathbb{R}^n, \ell^1) \rightarrow \mathbb{R}$  defined by  $F_n(x^{(n)}) = \sum_{j=1}^n x_j$ .

The exponential integral satisfies

$$\begin{aligned} (6.2) \quad \int_{\mathbb{R}^n} \exp(sF_n(x^{(n)})) P^{(n)}(dx^{(n)}) &= \mathbb{E} \exp(sF_n(\xi^{(n)})) \\ &= \mathbb{E} \exp\left(s \sum_{j=1}^n Z_{j\tau}^{(x)}\right). \end{aligned}$$

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

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

with  $\theta = e^{-\varrho\tau}$ . 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  $\sigma^2 = (1 - \theta^2)/(2\varrho)$  when  $\varrho \neq 0$ , and  $\sigma^2 = \tau$  when  $\varrho = 0$ . Hence the exponential integral (6.2) equals

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

where

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

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

$$(6.4) \quad \begin{aligned} W_1(p(\cdot | x), p(\cdot | x')) &= W_1(N(\theta x, \sigma^2), N(\theta x', \sigma^2)) \\ &= \theta |x - x'| \quad (x, x' \in \mathbb{R}). \end{aligned}$$

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  $\varrho = 0$ ,  $\varrho < 0$  and  $\varrho > 0$ .

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

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

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

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

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  $f(x^{(n)}) = \exp(\sum_{j=1}^n \theta^j x_j)$ , one can prove that the joint law  $P^{(n)}$  cannot satisfy a logarithmic Sobolev inequality with  $\alpha_n$  greater than some constant multiple of  $n^{-2}$  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. The authors are also grateful to the referee for suggesting improvements to the exposition.

### References

- [1] D. Bakry et M. Émery, *Diffusions hypercontractives*, 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. Bobkov, I. Gentil and M. Ledoux, *Hypercontractivity of Hamilton–Jacobi equations*, J. Math. Pures Appl. (9) 80 (2001), 669–696.
- [4] S. G. Bobkov and F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. 163 (1999), 1–28.
- [5] F. Bolley and C. Villani, *Weighted Csiszár–Kulback–Pinsker inequalities and applications to transportation inequalities*, Ann. Fac. Sci. Toulouse Math. (6) 14 (2005), 331–352.
- [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] 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.
- [9] 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.
- [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*, Amer. J. Math. 97 (1975), 1061–1083.
- [12] M. Ledoux, *The Concentration of Measure Phenomenon*, Amer. Math. Soc., Providence, RI, 2001.
- [13] K. Marton, *Bounding  $\bar{d}$ -distance by informational divergence: a method to prove measure concentration*, Ann. Probab. 24 (1996), 857–866.
- [14] —, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. 6 (1996), 556–571.
- [15] —, *Measure concentration for Euclidean distance in the case of dependent random variables*, Ann. Probab. 32 (2004), 2526–2544.
- [16] P. Massart, *Saint Flour lecture 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 Sci. Tech., Harlow, 1986.
- [20] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, 1991.
- [21] M. Rosenblatt, *Markov Processes: Structure and Asymptotic Behavior*, Springer, 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*, Amer. Math. Soc., Providence, RI, 2003.
- [25] F. Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds*, Probab. Theory Related Fields 108 (1997), 417–424.

Department of Mathematics and Statistics  
Lancaster University  
Lancaster LA1 4YF, UK  
E-mail: g.blower@lancaster.ac.uk

École Normale Supérieure de Lyon  
Umpa, 46 allée d'Italie  
F-69364 Lyon Cedex 07, France

*Current address:*  
Institut de mathématiques – LSP  
Université Paul Sabatier  
F-31062 Toulouse Cedex 9, France  
E-mail: bolley@cict.fr

*Received May 16, 2005*  
*Revised version February 8, 2006*

(5660)