Transport Inequalities, Gradient Estimates, Entropy and Ricci Curvature

Karl-Theodor Sturm & **Max-Konstantin von Renesse**

University of Bonn[1]

Abstract. We present various characterizations of uniform lower bounds for the Ricci curvature of a smooth Riemannian manifold M in terms of convexity properties of the entropy (considered as a function on the space of probability measures on M) as well as in terms of transportation inequalities for volume measures, heat kernels and Brownian motions and in terms of gradient estimates for the heat semigroup.

Keywords. transport inequalities, coupling, Wasserstein distance, displacement convexity, gradient estimate, entropy, Ricci curvature, Bakry-Emery estimate.

1 Introduction and statement of the main results

For metric measure spaces there is neither a notion of Ricci curvature nor a common notion of bounds for the Ricci curvature (comparable for instance to Alexandrov's notion of bounds for the sectional curvature for metric spaces).

The goal of this paper is to present various characterizations of uniform lower bounds for the Ricci curvature of a smooth Riemannian manifold M in terms of convexity properties of the entropy (considered as a function on the space of probability measures on M) as well as in terms of transportation inequalities for volume measures, heat kernels and Brownian motions and in terms of gradient estimates for the heat semigroup.

In the sequel, (M, g) always is assumed to be a smooth connected Riemannian manifold with dimension n, Riemannian distance $d(x, y)$ and Riemannian volume $m(dx) = \mathrm{vol}(dx)$. For $r \in [1, \infty[$ the L^r-Wasserstein distance of two measures μ_1, μ_2 on M is defined as

$$d_r^W(\mu_1, \mu_2) := \inf \left\{ \int_{M\times M} d(x_1, x_2)^r \, \mu(dx_1 dx_2) : \ \mu \in \mathcal{C}(\mu_1, \mu_2) \right\}^{1/r}$$

where $\mathcal{C}(\mu_1, \mu_2)$ denotes the set of all couplings of μ_1 and μ_2, that is, the set of all measures μ on $M \times M$ with $\mu(A \times M) = \mu_1(A)$ and $\mu(M \times A) = \mu_2(A)$ for all measurable $A \subset M$.

Here and in the sequel "measure on M" always means: measure on M equipped with its Borel σ-field. The set of probability measures μ on M with $\int_M d(x, y)^r \, \mu(dy) < \infty$ for some

[1]Supported by the Collaborative Research Center SFB 611 of the German Research Council (DFG).

(hence all) $x \in M$ will be denoted by $\mathcal{P}^r(M)$. Equipped with the metric d_r^W, the space $\mathcal{P}^r(M)$ is a geodesic space.

The *entropy* is defined as a function on $\mathcal{P}^r(M)$ by

$$\mathrm{Ent}(\nu) := \int_M \frac{d\nu}{dx} \log \frac{d\nu}{dx} \,\mathrm{vol}(dx)$$

if ν is absolutely continuous w.r.t. vol with $\int_M \frac{d\nu}{dx} \left[\log \frac{d\nu}{dx}\right]_+ \mathrm{vol}(dx) < \infty$ and $\mathrm{Ent}(\nu) := +\infty$ otherwise.

Given an arbitrary geodesic space (X, ρ), a number $K \in \mathbb{R}$ and a function $U : X \to]-\infty, +\infty]$ we say that U is *K-convex* iff for each (constant speed, as usual) geodesic $\gamma : [0,1] \to X$ and for each $t \in [0,1]$:

$$U(\gamma_t) \le (1-t)\, U(\gamma_0) + t\, U(\gamma_1) - \frac{K}{2} t(1-t)\, \rho^2(\gamma_0, \gamma_1).$$

K-convex functions on $\mathcal{P}^2(M)$ are also called *displacement K-convex* (to make sure that $t \mapsto \gamma_t$ is really the geodesic w.r.t. d_2^W and not the linear interpolation $t \mapsto (1-t)\gamma_0 + t\gamma_1$ in the space $\mathcal{P}^2(M)$).

Theorem 1. *For any smooth connected Riemannian manifold M and any $K \in \mathbb{R}$ the following properties are equivalent:*

(i) $\mathrm{Ric}(M) \ge K$,
which always should be read as: $\mathrm{Ric}_x(v,v) \ge K|v|^2$ *for all* $x \in M, v \in T_xM$.

(ii) *The entropy* $\mathrm{Ent}(.)$ *is displacement K-convex on* $\mathcal{P}^2(M)$.

One reason for the importance of Theorem 1 is that it characterizes lower Ricci bounds referring neither to the differential structure of M nor to the dimension of M. Property (ii) may be formulated in any metric measure space.

F. Otto & C. Villani [OV00] gave a very nice heuristic argument for the implication "(i)$\Rightarrow$(ii)". In the case $K = 0$ this implication was proven in [CMS01].

Remark 1. The argument in [OV00] is based on a kind of formal Riemannian calculus on the space $(\mathcal{P}^2(M), d_2^W)$. This calculus allows to interpret the heat flow $\Phi : (t, \mu) \mapsto \mu p_t := \int \mu(dx) p_t(x, .)$ as the gradient flow w.r.t. the entropy $\mathrm{Ent}(.)$ on the space $(\mathcal{P}^2(M), d_2^W)$. If the latter was a finite dimensional smooth Riemannian manifold then one easily could conclude that (ii) is equivalent to:

(iii) *The gradient flow* $\Phi : \mathbb{R}_+ \times \mathcal{P}^2(M) \to \mathcal{P}^2(M)$ *w.r.t.* $\mathrm{Ent}(.)$ *satisfies*

$$d_2^W(\Phi(t,\mu), \Phi(t,\nu)) \le e^{-Kt} \cdot d_2^W(\mu, \nu) \qquad (\forall \mu, \nu \in \mathcal{P}^2(M), \forall t \ge 0).$$

Indeed, both statements are equivalent. However, we will derive the equivalence of (ii) and (iii) only through their equivalence with (i). The equivalence of (i) and (iii) will be obtained as part of the more general Corollary 1 below.

Here and henceforth, $p_t(x,y)$ always denotes the heat kernel on M, i.e. the minimal positive fundamental solution to the heat equation $(\Delta - \frac{\partial}{\partial t}) p_t(x,y) = 0$. It is smooth in (t,x,y), symmetric in (x,y) and satisfies $\int_M p_t(x,y)\mathrm{vol}(dy) \le 1$. Hence, it defines a subprobability measure $p_t(x,dy) := p_t(x,y)\mathrm{vol}(dy)$ as well as operators $p_t : \mathcal{C}_c^\infty(M) \to \mathcal{C}^\infty(M)$ and $p_t : L^2(M) \to L^2(M)$

which are all denoted by the same symbol. Given $\mu \in \mathcal{P}^r(M)$ and $t > 0$ we define a new measure $\mu p_t \in \mathcal{P}^r(M)$ by $\mu p_t(A) = \int_A \int_M p_t(x,y)\mu(dx)\mathrm{vol}(dy)$.

Brownian motion on M is by definition the Markov process with generator $\frac{1}{2}\Delta$. Thus its transition (sub-)probabilities are given by $p_{t/2}$.

If the Ricci curvature of the underlying manifold M is bounded from below then all the $p_t(x,.)$ are probability measures. If the latter holds true we say that the heat kernel and the associated Brownian motion are conservative. It means that the Brownian motion has infinite lifetime.

The heat kernel on a Riemannian manifold is a fundamental object for analysis, geometry and stochastics. Many properties and precise estimates are known. In most of these results, lower bounds on the Ricci curvature of the underlying manifold play a crucial role. Often, the results for the heat kernel on Riemannian manifolds turned out to be a source of inspiration for results in entirely different frameworks or much greater generality e.g. (hypo-)elliptic operators, Hodge-Laplacians, Feller generators or generators of Dirichlet forms instead of the Laplace-Beltrami operator; path spaces, spaces of measures or fractals instead of finite dimensional manifolds.

Our second main result deals with robust versions of gradient estimates.

Theorem 2. *For any smooth connected Riemannian manifold M and any $K \in \mathbb{R}$ the following properties are equivalent:*

(i) $\mathrm{Ric}(M) \geq K$.

(iv) *For all $f \in C_c^\infty(M)$, all $x \in M$ and all $t > 0$*

$$|\nabla p_t f|(x) \leq e^{-Kt}\, p_t|\nabla f|(x).$$

(v) *For all $f \in \mathcal{C}_c^\infty(M)$ and all $t > 0$*

$$\|\nabla p_t f\|_\infty \leq e^{-Kt}\, \|\nabla f\|_\infty .$$

(vi) *For all bounded $f \in \mathcal{C}^{\mathrm{Lip}}(M)$ and all $t > 0$*

$$\mathrm{Lip}(p_t f) \leq e^{-Kt}\, \mathrm{Lip}(f).$$

The equivalence of (i) and (iv), perhaps, is one of the most famous general results which relate heat kernels with Ricci curvature. It is due to D. Bakry & M. Emery [BE84], see also [ABC$^+$00] and references therein. Property (iv) is successfully used in various applications as a replacement (or definition) of lower Ricci curvature bounds for symmetric Markov semigroups on general state spaces. Our result states that (iv) can be weakened in two respects:

– one can replace the pointwise estimate by an estimate between L^∞-norms;
– one can drop the p_t on the RHS.

Besides being formally weaker than (iv) one other advantage of (v) is that it is an explicit statement on the smoothing effect of p_t whereas (iv) is implicit (since p_t appears on both sides).

As an easy corollary to the equivalence of the statements (iv) and (v) one may deduce the well known fact that (iv) is equivalent to the assertion that for all f, x and t as above

$$|\nabla p_t f|(x) \leq e^{-Kt}\, \left[p_t(|\nabla f|^2)(x)\right]^{1/2}.$$

Property (vi) may be considered as a replacement (or as one possible definition) for lower Ricci curvature bounds for Markov semigroups on metric spaces. For several non-classical examples (including nonlocal generators as well as infinite dimensional or singular finite dimensional state spaces) we refer to [Stu03], [DR02] and [vRen03]. This property turned out to be the key ingredient to prove Lipschitz continuity for harmonic maps between metric spaces in [Stu03].

According to the Kantorovich-Rubinstein duality, property (vi) is equivalent to a contraction property for the heat kernels in terms of the L^1-Wasserstein distance d_1^W. Actually, however, much more can be proven:

– one obtains contraction in d_r^W for each $r \in [1, \infty]$ and for any initial data;
– one obtains pathwise contraction for Brownian trajectories.

Corollary 1. *For any smooth connected Riemannian manifold M and any $K \in \mathbb{R}$ the following properties are equivalent:*

(i) $\mathrm{Ric}(M) \geq K$.

(vii) *For all $x, y \in M$ and all $t > 0$ there exists $r \in [1, \infty]$ with*

$$d_r^W(p_t(x,.), p_t(y,.)) \leq e^{-Kt} \cdot d(x,y).$$

(viii) *For all $r \in [1, \infty]$, all $\mu, \nu \in \mathcal{P}^r(M)$ and all $t > 0$:*

$$d_r^W(\mu p_t, \nu p_t) \leq e^{-Kt} \cdot d_r^W(\mu, \nu).$$

(ix) *For all x_1, $x_2 \in M$ there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and two conservative Brownian motions $(X_1(t))_{t \geq 0}$ and $(X_2(t))_{t \geq 0}$ defined on it, with values in M and starting in x_1 and x_2, respectively, such that for all $t > 0$*

$$\mathbb{E}\left[d(X_1(t), X_2(t))\right] \leq e^{-Kt/2} \cdot d(x_1, x_2).$$

(x) *There exists a conservative Markov process $(\Omega, \mathcal{A}, \mathbb{P}^x, X(t))_{x \in M \times M, t \geq 0}$ with values in $M \times M$ such that the coordinate processes $(X_1(t))_{t \geq 0}$ and $(X_2(t))_{t \geq 0}$ are Brownian motions on M and such that for all $x = (x_1, x_2) \in M \times M$ and all $t > 0$*

$$d(X_1(t), X_2(t)) \leq e^{-Kt/2} \cdot d(x_1, x_2) \qquad \mathbb{P}^x\text{-a.s.}$$

Note that each of the statements (vii) and (viii) implicitly includes the conservativity of the heat kernel. Indeed, the finiteness of the Wasserstein distance implies that the measures under consideration must have the same total mass. Thus $p_t(x, M)$ is constant in x, hence also constant in t and therefore equal to 1.

Our interpretation of these results is as follows: if we put mass distributions μ and ν on M and if they spread out according to the heat equation then the lower bound for the Ricci curvature of M controls how fast the distances between these distributions may expand (or have to decay) in time.

Our third main result deals with transportation inequalities for uniform distributions on spheres and analogous inequalities for uniform distributions on balls. Here the lower Ricci bound is characterized as a control for the increase of the distances if we replace Dirac masses δ_x and δ_y by uniform distributions $\sigma_{r,x}$ and $\sigma_{r,y}$ on spheres around x and y, resp. or if we replace them by uniform distributions $m_{r,x}$ and $m_{r,y}$ on balls around x and y, resp.

Theorem 3. *For any smooth connected compact Riemannian manifold M and any $K \in \mathbb{R}$ the following properties are equivalent:*

(i) $\mathrm{Ric}(M) \geq K$.

(xi) *The normalized Riemannian uniform distribution on spheres*

$$\sigma_{r,x}(A) := \frac{\mathcal{H}^{n-1}(A \cap \partial B_r(x))}{\mathcal{H}^{n-1}(\partial B_r(x))}, \qquad A \in \mathcal{B}(M)$$

satisfies the asymptotic estimate

$$d_1^W(\sigma_{r,x}, \sigma_{r,y}) \leq \Big(1 - \frac{K}{2n} r^2 + o(r^2)\Big) \cdot d(x,y) \tag{1}$$

where the error term is uniform w.r.t. $x, y \in M$.

(xii) *The normalized Riemannian uniform distribution on balls*

$$m_{r,x}(A) := \frac{m(A \cap B_r(x))}{m(B_r(x))}, \qquad A \in \mathcal{B}(M)$$

satisfies the asymptotic estimate

$$d_1^W(m_{r,x}, m_{r,y}) \leq \Big(1 - \frac{K}{2(n+2)} r^2 + o(r^2)\Big) \cdot d(x,y) \tag{2}$$

where the error term is uniform w.r.t. $x, y \in M$.

The advantage of this characterization of Ricci curvature is that it depends only on the basic, robust data: measure and metric. It does not require any heat kernel, any Laplacian or any Brownian motion. It might be used as a guideline in much more general situations.

For instance, let (M, d) be an arbitrary separable metric space equipped with a measure m on its Borel σ-field and assume that (2) holds true (with some numbers $K \in \mathbb{R}$ and $n > 0$). Define an operator m_r acting on bounded measurable functions by $m_r f(x) = \int_M f(y)\, m_{r,x}(dy)$. Then by the Arzela-Ascoli theorem there exists a sequence $(l_j)_j \subset \mathbb{N}$ such that

$$p_t f := \lim_{j\to\infty} \Big(m_{\sqrt{2(n+2)t/l_j}}\Big)^{l_j} f$$

exists (as a uniform limit) for all bounded $f \in \mathcal{C}^{\mathrm{Lip}}(M)$ and it defines a Markov semigroup on M satisfying

$$\mathrm{Lip}(p_t f) \leq e^{-Kt} \mathrm{Lip}(f)$$

(cf. proof of the implication "(xi)$\Rightarrow$(i)" of the above Theorem and proof of Theorem 4.3 in [Stu03]).

2 Ricci curvature and entropy

Proof of Theorem 1.

(ii)$\Rightarrow$(i): Assume $\neg$(i). Then $\mathrm{Ric}_0(e_1, e_1) \leq K - \epsilon$ for some $o \in M$, some unit vector $e_1 \in T_o M$ and some $\epsilon > 0$. Let $e_1, e_2, \ldots, e_n$ be a ONB of $T_o M$ such that

$$R(e_1, e_i)e_1 = k_i e_i$$

for suitable numbers k_i, $i = 1, \ldots, n$ (denoting the sectional curvature of the plane spanned by e_1 and e_i if $i \neq 1$). Then $\sum_{i=1}^n k_i = \mathrm{Ric}_0(e_1, e_1) \leq K - \epsilon$.

For $\delta, r > 0$ let $A_1 := B_\delta(\exp_o(re_1))$ and $A_0 := B_\delta(\exp_o(-re_1))$ be geodesic balls and let

$$A_{1/2} := \exp_o\left(\left\{y \in T_oM : \sum_{i=1}^n (y_i/\delta_i)^2 \leq 1\right\}\right)$$

with $\delta_i := \delta \cdot (1 + r^2(k_i + \frac{\epsilon}{2n})/2)$. Choosing $\delta \ll r \ll 1$ we can achieve that $\gamma_{1/2} \in A_{1/2}$ for each minimizing geodesic $\gamma : [0,1] \to M$ with $\gamma_0 \in A_0$, $\gamma_1 \in A_1$.

Now let μ_0 and μ_1 be the normalized uniform distribution in A_0 and A_1, resp. and let ν be the normalized uniform distribution in $A_{1/2}$. Then

$$\mathrm{Ent}(\mu_0) = -\log \mathrm{vol} A_0 = -\log c_n - n \log \delta + O(\delta^2)$$

$$\mathrm{Ent}(\mu_1) = -\log \mathrm{vol} A_0 = -\log c_n - n \log \delta + O(\delta^2)$$

with $c_n := Vol(B_1)$ in $\mathbb{R}^n$ whereas

$$\begin{aligned}
\mathrm{Ent}(\nu) &= -\log \mathrm{vol} A_{1/2} = -\log c_n - \sum_{i=1}^n \log \delta_i + O(\delta^2) \\
&= -\log c_n - n \log \delta - r^2(\epsilon/2 + \sum_{i=1}^n k_i)/2 + O(r^4) + O(\delta^2) \\
&\geq -\log c_n - n \log \delta - r^2(K - \epsilon/2)/2 + O(r^4) + O(\delta^2)
\end{aligned}$$

Since the optimal mass transport from μ_0 to μ_1 (w.r.t. d_2^W) is along geodesics of M, the support of $\mu_{1/2}$ must be contained in the set $A_{1/2}$. Hence,

$$\mathrm{Ent}(\mu_{1/2}) \geq \mathrm{Ent}(\nu)$$

and thus

$$\begin{aligned}
\mathrm{Ent}(\mu_{1/2}) - \frac{1}{2}\mathrm{Ent}(\mu_0) - \frac{1}{2}\mathrm{Ent}(\mu_1) &\geq -\frac{K}{2}r^2 + \frac{\epsilon}{4}r^2 + O(r^4) + O(\delta^2) \\
&> -\frac{K}{8} d_2^W(\mu_0, \mu_1)^2
\end{aligned}$$

for $\delta \ll r \ll 1$.

(i)⇒(ii): Here we follow closely the argumentation of [CMS01] and use their notation. Assume that $\mathrm{Ric}(M) \geq K$. We have to prove that

$$\mathrm{Ent}(\mu_t) \leq (1-t)\mathrm{Ent}(\mu_0) + t\mathrm{Ent}(\mu_1) - \frac{K}{2} t(1-t) d_2^W(\mu_0, \mu_1)^2$$

for each geodesic $t \mapsto \mu_t$ in $(\mathcal{P}^2(M), d_2^W)$ and each $t \in [0,1]$. Without restriction, we may assume that μ_0 and μ_1 are absolutely continuous (otherwise the RHS is infinite). Hence, there exists a unique geodesic connecting them. It is given as $\mu_t = (F_t)_*\mu_0$ where $F_t(x) = \exp_x(-t\nabla\varphi(x))$ with a suitable function φ. Moreover, with $J_t(x) := \det dF_t(x)$ and $\mathsf{s}(r) := \sin\left(\sqrt{\frac{K}{n-1}} \cdot r\right) / \left(\sqrt{\frac{K}{n-1}} \cdot r\right)$ (which should be read as $\mathsf{s}(r) := \sinh\left(\sqrt{\frac{-K}{n-1}} \cdot r\right) / \left(\sqrt{\frac{-K}{n-1}} \cdot r\right)$ if

$K<0$ and as $\mathsf{s}(r)=1$ if $K=0$) and with $v_t(x,y)$ being the volume distortion coefficient of [CMS01] we deduce

$$\mathrm{Ent}(\mu_t)=\mathrm{Ent}(\mu_0)-\int_M \log J_t(x)\mu_0(dx)$$

and thus

$$\begin{aligned}
&-\mathrm{Ent}(\mu_t)+(1-t)\mathrm{Ent}(\mu_0)+t\mathrm{Ent}(\mu_1)\\
&\quad= \int_M \log J_t(x)\mu_0(dx)-t\int_M \log J_1(x)\mu_0(dx)\\
&\quad\geq n\int \log\left[(1-t)v_{1-t}(F_1(x),x)^{1/n}+tv_t(x,F_1(x))^{1/n}J_1(x)^{1/n}\right]\mu_0(dx)-t\int_M \log J_1(x)\mu_0(dx)\\
&\quad\geq n\int \log\left[(1-t)\left[\frac{\mathsf{s}((1-t)d(F_1(x),x))}{\mathsf{s}(d(F_1(x),x))}\right]^{1-1/n}+t\left[\frac{\mathsf{s}(td(x,F_1(x)))}{\mathsf{s}(d(F_1(x),x))}\right]^{1-1/n}J_1(x)^{1/n}\right]\mu_0(dx)\\
&\qquad\qquad -t\int_M \log J_1(x)\mu_0(dx)\\
&\quad\geq (n-1)\int\left[(1-t)\log\mathsf{s}((1-t)d(F_1(x),x))+t\log\mathsf{s}(td(F_1(x),x))-\log\mathsf{s}(d(F_1(x),x))\right]\mu_0(dx)\\
&\quad\geq \frac{K}{2}t(1-t)\int d^2(F_1(x),x)\mu_0(dx)\\
&\quad= \frac{K}{2}t(1-t)d_2^W(\mu_0,\mu_1).
\end{aligned}$$

Here the first and second inequality follow from [CMS01], equations (73) and (24). The third inequality follows from the concavity of the logarithm and the last one from the fact that

$$\begin{aligned}
(1-t)\log\mathsf{s}\left((1-t)r\right)+t\log\mathsf{s}\left(tr\right)-\log\mathsf{s}(r) &\;-\; \frac{t(1-t)}{2}\frac{K}{n-1}r^2\\
=\;(1-t)\lambda((1-t)r)+t\lambda(tr)-\lambda(r) &\;\geq\; 0
\end{aligned}$$

for all $r\geq 0$ under consideration and $t\in[0,1]$ since $\lambda'(r)\leq 0$ where $\lambda(r):=\log\mathsf{s}(r)+\frac{1}{6}\frac{K}{n-1}r^2$. Note that according to the Bonnet-Myers theorem we may restrict ourselves to $r\geq 0$ with $\frac{K}{n-1}r^2\leq\pi^2$.

In order to verify that $\lambda'(r)\leq 0$ it suffices to consider the cases $K=\pm(n-1)$. If $K=-(n-1)$ then $\lambda(r)=\log\sinh r-\log r-\frac{1}{6}r^2$ and $\lambda'(r)=\frac{\cosh r}{\sinh r}-\frac{1}{r}-\frac{1}{3}r$. The latter is nonpositive for all $r>0$ if and only if

$$r\cosh r-\sinh r-\frac{1}{3}r^2\sinh r\leq 0$$

for all $r>0$. Differentiating and dividing by $r/3$ we see that this is equivalent to $-r\cosh r+\sinh r\leq 0$ which (again by differentiation) will follow from $-r\sinh r\leq 0$ which is obviously true.

Analogously, if $K=n-1$ the condition $\lambda'(r)\leq 0$ is equivalent to

$$r\cos r-\sin r+\frac{1}{3}r^2\sin r\leq 0$$

which (by the same arguments as before) is equivalent to $-r\sin r\leq 0$. Here of course, we have to restrict ourselves to $r\in[0,\pi]$. □

3 Ricci curvature and transport inequalities for heat kernels and Brownian motions

Proof of Theorem 2 and Corollary 1.

(i)⇒(iv): This is due to D. Bakry & M. Emery [BE84] and can be obtained using their Γ_2 calculus (cf. [ABC+00], prop. 5.4.5.).

(iv)⇒(v): Take $\|.\|_\infty$ on both sides and use (on the RHS) the fact that p_t is a contraction on $L^\infty(M)$.

(v)⇒(i): We prove it by contradiction, assuming $\neg$ (i) $\wedge$ (v). If (i) is not true then there exists a point $0 \in M$ and $v \in S^{n-1} \subset T_0 M$ such that $\mathrm{Ric}_0(v,v) \le K - \epsilon$ for some $\epsilon > 0$. Let $\mathcal{F} = \{x \in M \setminus \mathrm{Cut}(0) \,|\, \log_0 x \perp v\} \subset M$ be the orthogonal hypersurface to v in M and define the signed distance function $d_{\mathcal{F}}^{\pm}$ from $\mathcal{F}$ by

$$d_{\mathcal{F}}^{\pm} : M \setminus \mathrm{Cut}(0) \to \mathbb{R}, \quad d_{\mathcal{F}}^{\pm}(x) := \mathrm{dist}(x, \mathcal{F})\mathrm{sign}\langle v, \log_0 x\rangle. \tag{3}$$

It is shown in Lemma 1 below that $d_{\mathcal{F}}^{\pm} \in C^\infty(\mathcal{U})$ for some neighborhood $\mathcal{U} \ni 0$ and that furthermore

$$\nabla d_{\mathcal{F}}^{\pm}(0) = v \tag{4}$$
$$|\nabla d_{\mathcal{F}}^{\pm}|(x) = 1 \quad \forall\, x \in \mathcal{U} \tag{5}$$
$$\mathrm{Hess}_0(d_{\mathcal{F}}^{\pm}) = 0. \tag{6}$$

Hence for any neighborhood $\mathcal{U}' \ni 0$ with $\overline{\mathcal{U}'} \subset \mathcal{U}$ and any smooth cut-off function $\psi \in C_c^\infty(\mathcal{U})$ with $\psi_{|\mathcal{U}'} \equiv 1$ the function $f = d_{\mathcal{F}}^{\pm} \cdot \psi$ inherits the properties (4), (5) and (6) with $\mathcal{U}$ replaced by $\mathcal{U}'$ in (5). In particular, by continuity of the function $\Gamma_2(f,f) : \mathcal{U} \to \mathbb{R}$

$$\Gamma_2(f,f)(x) := \|\mathrm{Hess}_x(f)\|^2 + \mathrm{Ric}_x(\nabla f, \nabla f)$$

we find

$$\Gamma_2(f,f) \le K - \frac{1}{2}\epsilon$$

on some neighborhood of 0 which contains w.l.o.g. $\mathcal{U}'$. If, moreover, (v) is true then for any nonnegative test function $\varphi \in C_c^\infty$ and $t \ll 1$ we obtain

$$\begin{aligned}\int_M |\nabla p_t f|^2 \varphi \,\mathrm{dvol} &\le \|\nabla p_t f\|_\infty^2 \int_M \varphi \,\mathrm{dvol} \\ &\le \exp(-2Kt) \int_M \varphi \,\mathrm{dvol} \\ &= (1 - 2Kt + o(t)) \int_M \varphi \,\mathrm{dvol}.\end{aligned} \tag{7}$$

We choose a test function $0 \le \varphi \in C_c^\infty(\mathcal{U}')$ and define for $t > 0$

$$\Phi(t) := \int_M |\nabla p_t f|^2 \varphi \,\mathrm{dvol}.$$

Hence, since $|\nabla p_t f|^2 \to |\nabla f|^2 \equiv 1$ on $\mathrm{supp}(\varphi) \subset \mathcal{U}'$, the function Φ extends continuously on the entire nonnegative half line by $\Phi(0) := \int_M \varphi$. From Bochner's formula we deduce

$$\begin{aligned}
\Phi'(t) &= \int_M 2\langle \nabla p_t \Delta f, \nabla p_t f\rangle \varphi \,\mathrm{dvol} \\
&= \int_M (2\Delta |\nabla p_t f|^2 - 2\Gamma_2(p_t f, p_t f))\varphi \,\mathrm{dvol} \\
&= \int_M 2|\nabla p_t f|^2 \Delta\varphi - 2\Gamma_2(p_t f, p_t f)\varphi \,\mathrm{dvol} \\
&\xrightarrow{t\to 0} \int_M (2|\nabla f|^2 \Delta\varphi - 2\Gamma_2(f,f)\varphi) \,\mathrm{dvol} \\
&= -2\int_M \Gamma_2(f,f)\varphi \,\mathrm{dvol} \geq (\epsilon - 2K)\int_M \varphi = (\epsilon - 2K)\Phi(0).
\end{aligned}$$

Thus $t \to \Phi(t)$ is differentiable in $t = 0+$ with $\Phi'(0+) > (\epsilon - 2K)\Phi(0)$. Consequently we find for small t that $\Phi(t) \geq \Phi(0) + (\epsilon - 2K)\Phi(0)t + o(t) = \Phi(0)(1 - 2Kt + o(t))$, i.e.

$$\int_M |\nabla p_t f|^2 \varphi \,\mathrm{dvol} \geq (1 + (\epsilon - 2K)t + o(t)) \int_M \varphi \,\mathrm{dvol}$$

which is a contradiction to (7).

(vi)⇒(v): Trivial

(vii)⇒(vi): By Hölder's inequality, property (vii) for some $r \geq 1$ implies property (vii) for $r = 1$ which in turn implies (vi) according to the Kantorovich-Rubinstein duality.

Or explicitly: for each coupling λ of $p_t(x,.)$ and $p_t(y,.)$

$$\begin{aligned}
|p_t f(x) - p_t f(y)| &= \left| \int [f(z) - f(w)]\lambda(dzdw) \right| \\
&\leq \mathrm{Lip}(f) \cdot \int d(z,w)\lambda(dzdw) \\
&\leq \mathrm{Lip}(f) \cdot \left[\int d(z,w)^r \lambda(dzdw) \right]^{1/r}.
\end{aligned}$$

Hence,

$$|p_t f(x) - p_t f(y)| \leq \mathrm{Lip}(f) \cdot d_r^W(p_t(x,.), p_t(y,.)) \leq \mathrm{Lip}(f) \cdot d(x,y) \cdot e^{-Kt}.$$

(viii)⇒(vii): Choose $\mu = \delta_x, \nu = \delta_y$.

(ix)⇒(vii): The distribution $\lambda(.) := \mathbb{P}((X_1(2t), X_2(2t)) \in .)$ of the pair $(X_1(2t), X_2(2t))$ defines a coupling of $p_t(x_1,.)$ and $p_t(x_2,.)$. Hence,

$$d_1^W(p_t(x_1,.), p_t(x_2,.)) \leq \int d(z_1,z_2)\,\lambda(dz_1dz_2) = \mathbb{E}\left[d(X_1(2t), X_2(2t))\right] \leq e^{-Kt} \cdot d(x_1,x_2).$$

(x)⇒(viii): Let λ be an optimal coupling of μ and ν w.r.t. d_r^W and let Π_t be the transition semigroup of the Markov process from (x). Then $\lambda_t := \lambda\Pi_{2t}$ is a coupling of μp_t and νp_t. Hence,

$$
\begin{aligned}
d_r^W(\mu p_t, \nu p_t)^r &\leq \int d(w_1, w_2)^r \lambda_t(dw_1 dw_2) \\
&= \int\int d(w_1, w_2)^r \Pi_{2t}((x_1, x_2), dw_1 dw_2)\lambda(dx_1 dx_2) \\
&= \int \mathbb{E}^{(x_1,x_2)} \left[d(X_1(2t), X_2(2t))^r\right] \lambda(dx_1 dx_2) \\
&\leq e^{-Ktr} \cdot \int d(x_1, x_2)^r \lambda(dx_1, dx_2) \\
&= e^{-Ktr} \cdot d_r^W(\mu, \nu)^r.
\end{aligned}
$$

(x)⇒(ix): Take expectations.

(i)⇒(x) is well known and can be shown using either SDE theory on Riemannian manifolds in order to construct the coupling by parallel transport process on $M \times M$ for two Brownian motions (cf. [Wan97] and references therein) or by a central limit theorem for coupled geodesic random walks and an estimate of the type (2) (cf. [vRen03] for a similar argument). □

Remark 2. There are two more very natural candidates which one could consider as a test function in the proof of the (v)⇒ (i) part above, namely $f_1(z) := \langle \log_0(z), v\rangle_{T_0M}$ and $f_2(z) = d(\exp_0(Lv), z)$ with L being large compared to $\mathcal{U}$. However, even if $\nabla f_1(0) = \nabla f_2(0) = v$ none of the two will help because $|\nabla f_1| \not\equiv 1$ in a neighborhood of 0 and $\mathrm{Hess}_0\, f_2 \neq 0$, which were essential properties of $d_{\mathcal{F}}^{\pm}$ in our proof.

Lemma 1. *Let M be a Riemannian manifold, $0 \in M$, $v \in T_0M$ and $\mathcal{F} = \{\exp_0(u)\,|u \in T_0M, u\perp v\} \subset M$ the $(n-1)$-dimensional hypersurface through 0 orthogonal to v. Then the signed distance function $d_{\mathcal{F}}^{\pm} : M \to \mathbb{R}$ belongs to $C^\infty(\mathcal{U})$ for some neighborhood $\mathcal{U} \ni 0$ and $\mathrm{Hess}_0(d_{\mathcal{F}}^{\pm}) = 0$.*

Proof. The level sets $\mathcal{F}_\epsilon = \{x \in M|d_{\mathcal{F}}^{\pm}(x) = \epsilon\}$ of $d_{\mathcal{F}}^{\pm}$ for $|\epsilon| \ll 1$ define a foliation of (a sufficiently small) neighborhood $\mathcal{U} \ni 0$ by smooth hypersurfaces. The unit normal vector field to $\mathcal{F}_\epsilon$ is given by $\nu = \nabla d_{\mathcal{F}}^{\pm}$ which is well defined and smooth sufficiently close to $\mathcal{F}$ ($d_{\mathcal{F}}^{\pm}$ is a 'distance function' on $\mathcal{U}$ in the sense of [Pet98], sec. 2.3.). Hence the Hessian of $d_{\mathcal{F}}^{\pm}$ in a point $p \in \mathcal{U}$ may be interpreted as the shape operator of $\mathcal{F}_\epsilon$ in $p \in \mathcal{F}_\epsilon$ with $\epsilon = d_{\mathcal{F}}^{\pm}(p)$, i.e.

$$
\mathrm{Hess}_p(X, X) = \Pi_p^{\mathcal{F}_\epsilon}(X, X) = \langle S_p^{\mathcal{F}_\epsilon}(X), X\rangle_{T_pM},
$$

where $\Pi_p^{\mathcal{F}_\epsilon}$ is the second fundamental form of the hypersurface $\mathcal{F}_\epsilon \subset M$ and $S^{\mathcal{F}_\epsilon} : T_pM \to T_p\mathcal{F}_\epsilon$ is the associated shape operator. The claim $\mathrm{Hess}_0(d_{\mathcal{F}}^{\pm}) = 0$ then follows from the construction of $\mathcal{F}$ which implies that $\mathcal{F} = \mathcal{F}_0 \subset M$ is flat in 0, i.e. $S_0^{\mathcal{F}} = 0$. □

4 Ricci curvature and transport inequalities for uniform distributions on spheres and balls

Proof of Theorem 3.

(xi) $\Rightarrow$ (i): Let us define the family of Markov operators $\sigma_r : \mathcal{F}_B \to \mathcal{F}_B$ on the set $\mathcal{F}_B$ of bounded Borel measurable functions on M by $\sigma_r f(x) = \int_M f(y)\sigma_{r,x}(dy)$. Using that for $f \in C^3(M)$

$$\sigma_r f(x) = f(x) + \frac{r^2}{2n}\Delta f(x) + o(r^2) \tag{8}$$

and an appropriate version of the Trotter-Chernov product formula (cf. thm. 1.6.5. in [EK86], applied to $p_t = \exp(t\Delta)$ as Feller semigroup on $(C(M), \|.\|_\infty))$ we find for all $f \in C(M)$

$$\left(\sigma_{\sqrt{2nt/j}}\right)^j f(x) \stackrel{j\to\infty}{\longrightarrow} p_t f(x)$$

uniformly in $x \in M$ and locally uniformly in $t \geq 0$. By the Rubinstein-Kantorovich duality condition (xi) implies

$$|\sigma_r f(x) - \sigma_r f(y)| \leq \left(1 - \frac{K}{2n}r^2 + o(r^2)\right) \cdot d(x,y) \cdot \mathrm{Lip}(f)$$

for all $f \in \mathcal{C}^{\mathrm{Lip}}(M)$ and $x, y \in M$, i.e.

$$\mathrm{Lip}(\sigma_r f) \leq \big(1 - \frac{K}{2n}r^2 + o(r^2)\big) \cdot \mathrm{Lip}(f)$$

and hence by iteration for $j \in \mathbb{N}$, $r = \sqrt{2nt/j}$

$$\mathrm{Lip}\left(\left(\sigma_{\sqrt{2nt/j}}\right)^j f\right) \leq \big(1 - Kt/j + o(t/j)\big)^j \cdot \mathrm{Lip}(f).$$

Passing to the limit for $j \to \infty$ yields

$$\mathrm{Lip}(p_t f) \leq \exp(-Kt) \cdot \mathrm{Lip}(f),$$

which is equivalent to (i) by Theorem 2.

For the proof of the converse we construct an explicit transport from $\sigma_{r,x}$ to $\sigma_{r,y}$ in the following lemma, whose proof is given below.

Lemma 2. *Let M be a smooth connected compact Riemannian manifold and for $x \in M$ let $\sigma_{r,x}(.)$ denote the normalized Riemannian uniform distribution on $S_r(x) := \partial B_r(x)$. Then for r sufficiently small for each $x, y \in M$ there exits a geodesic segment $\gamma = \gamma_{xy}$ and a measurable map $\Psi_r^{x,y} : S_r(x) \to S_r(y)$ such that the push forward measure $\Psi_{r,*}^{x,y}\sigma_{r,x}$ equals $\sigma_{r,y}$ and*

$$\sup_{z\in S_r(x)} \sup_{y\in M} \frac{|\log_x z - /\!\!/_\gamma^{-1} \log_y \Psi_r^{x,y}(z)|}{d(x,y)} = o(r^2) \tag{9}$$

where the error term $o(r^2)$ is uniform in $x \in M$.

(i) ⇒ (xi): We show it for the case $K < 0$, the case $K \geq 0$ is treated analogously. Due to Lemma 2 we can assume that up to a negligible error $\Psi_r^{x,y}$ is indeed nothing but parallel transport because

$$\begin{aligned} d(z, \Psi_r^{x,y}(z)) &\leq d(z, \exp_y(/\!/_{\gamma_{xy}} \log_x(z))) + d(\exp_y(/\!/_{\gamma} \log_x(z)), \Psi_r^{x,y}(z)) \\ &\leq d(z, \exp_y(/\!/_{\gamma} \log_x(z))) + L|/\!/_{\gamma} \log_x(z) - \log_y \Psi_r^{x,y}(z)| \\ &\leq d(z, \exp_y(/\!/_{\gamma} \log_x(z))) + Ld(x,y)o(r^2), \end{aligned}$$

where L is some uniform upper bound for the Lipschitz constant of $\log_{.}(.)$ with respect to the second argument. The asymptotic inequality (1) is now easily verified from (8), since

$$\begin{aligned} d_1^W(\sigma_{r,x}, \sigma_{r,y}) &\leq \frac{1}{\mathcal{H}^{n-1}(\partial B_r(x))} \int\limits_{\partial B_r(x)} d(z, \Psi_r^{x,y}(z)) \mathcal{H}^{n-1}(dz) \\ &= \frac{1}{\mathcal{H}^{n-1}(\partial B_r(x))} \int\limits_{\partial B_r(x)} d(z, \exp_y /\!/_{\gamma} \log_x z) \mathcal{H}^{n-1}(dz) + d(x,y)o(r^2) \\ &= d(x,y) + \frac{r^2}{2n} \Delta D^{x,y}(x) + d(x,y)o(r^2) \end{aligned}$$

with $z \to D^{x,y}(z) = d(z, \exp_y /\!/_{\gamma} \log_x z)$. Since

$$\Delta D^{x,y}(x) = \operatorname{trace} \operatorname{Hess}_x D^{x,y} = \sum_{i=2}^{n} I_M^{\gamma}(J_i, J_i)$$

where $I_M^{\gamma}(J_i, J_i)$ is the Indexform of M along γ_{xy} applied to the Jacobi field induced from parallel geodesic varations of γ in the direction e_i with $\{\dot{\gamma}_{xy}^e, e_2, \cdots, e_n\}$ being an orthonormal basis of $T_x M$. Hence we may conclude by the standard Ricci comparison argument that

$$\Delta D^{x,y}(x) \leq 2(n-1)\sqrt{\frac{-K}{n-1}} \frac{\cosh\left(\sqrt{\frac{-K}{n-1}} d(x,y)\right) - 1}{\sinh\left(\sqrt{\frac{-K}{n-1}} d(x,y)\right)} \leq -K d(x,y)$$

such that we finally arrive at

$$d_1^W(\sigma_{r,x}, \sigma_{r,y}) \leq d(x,y) - \frac{r^2}{2n} K d(x,y) + d(x,y) o(r^2).$$

(xii) ⇒ (i): This is shown in the same way as the implication "(xi) ⇒ (i)" with the slight difference that instead of (8) one uses

$$m_r f(x) = f(x) + \frac{r^2}{2(n+2)} \Delta f(x) + o(r^2).$$

(i) ⇒ (xii): We proceed as before for "(i) ⇒ (xi)" where now we have to construct a map $\Phi_r^{x,y}$ which preserves the normalized uniform distributions on balls. However, since similarly to the condition (16) in the proof of Lemma 2 we have

$$m_{r,x}(A) = \frac{1}{m(B_r(x))} \int\limits_0^r \sigma_{u,x}(A) \mathcal{H}^{n-1}(\partial B_u(x)) du$$

such a map can be constructed from a map $\Psi^{x,y}_{r_1,r_2}:\partial B_{r_1}(x)\to\partial B_{r_2}(y)$ with $(\Psi^{x,y}_{r_1,r_2})_*\sigma_{r_1,x}=\sigma_{r_2,y}$ and which is almost induced from parallel transport in the sense of (9). It is clear that Lemma 2 can easily be generalized to yield such a map $\Psi^{x,y}_{r_1,r_2}$ which is all we need.

□

Proof of Lemma 2. We show the lemma for the two dimensional case first and inductively generalize this result to higher dimensions later. Let $n=2$ and choose a parametrization of $S_r(x)$ and $S_r(y)$ (using Riemannian polar coordinates, for example) on $S^1\subset\mathbb{R}^2\simeq T_xM$, i.e. for all $f:S_r(x)\to\mathbb{R}$

$$\int\limits_{S_r(x)} f(z)\mathcal{H}^{n-1}(dz)=\int\limits_{S^1} D_x(r,\vartheta)\tilde f(\vartheta)\mathcal{S}(d\vartheta)=\int\limits_0^{2\pi} D_x(r,s)\tilde f(s)ds$$

with a density $D_x(r,\vartheta)$ given by

$$\begin{aligned}D_x(t,\vartheta)&=\sqrt{\det\left(\langle Y_i(t,\vartheta),Y_j(t,\vartheta)\rangle\right)_{ij}}\\&=t^{n-1}\big(1-\frac{t^2}{6}c_x(\vartheta,\vartheta)+o(t^2)\big),\end{aligned}\tag{10}$$

where c_x is the Gaussian curvature of (M,g) in x and $Y_j(t,\vartheta)$ is the Jacobi field along $t\to\exp_x(t\vartheta)$ with $J_i(0)=0$ and $J_i'(0)=e_i$ for an orthonormal basis $\{e_i\,|\,i=1,2\}$ of T_xM (cf. [GHL90], for instance).

For x,y and γ_{xy} fixed let the parametrization of $S^1_x\subset T_xM$ and $S^1_y\subset T_yM$ on $[0,2\pi]$ be chosen in such a way that

$$S^1_x\ni 0\simeq\dot\gamma_{xy}(0)\text{ und }S^1_y\ni 0\simeq\dot\gamma_{xy}((d(xy)).$$

Next, we choose a function $\tau=\tau^{x,y}_r:[0,2\pi]\to[0,2\pi]$ with $\tau(0)=0$ satisfying

$$\frac{1}{H^{n-1}(S_r(x))}\int\limits_0^u D_x(r,s)ds=\frac{1}{H^{n-1}(S_r(y))}\int\limits_0^{\tau(u)} D_y(r,s)ds\tag{11}$$

for all $u\in[0,2\pi]$. Identifying $\tau:[0,2\pi]\to[0,2\pi]$ with the associated $\tau:S^1\to S^1$ then (11) just means that the induced map $\Psi=\Psi^{x,y}:S_r(x)\to S_r(y)$

$$\Psi(z):=\exp_y(r\tau(\frac{1}{r}\log_x(z)))$$

transports the measure $\mathcal{H}^{n-1}_{r,x}$ into $\mathcal{H}^{n-1}_{r,y}$. By definition of $\Psi^{x,y}$ estimate (9) is equivalent to

$$\sup_{z\in[0,2\pi]}\sup_{y\in M}\frac{|\tau^{x,y}_r(z)-z|}{d(x,y)}=o(r).\tag{12}$$

Denoting

$$E_x(r,s):=\frac{D_x(r,s)}{\mathcal{H}^{n-1}(S_r(x))}=1+O(r^2),\tag{13}$$

with $E_.(.,s)\in C^2(M\times[0,\epsilon])$ for all $s\in[0,2\pi]$ and some $\epsilon>0$, (11) yields

$$(1+O(r^2))|\tau(z)-z|=\left|\int\limits_z^{\tau(z)} E_y(r,s)ds\right|=\left|\int\limits_0^z\left(E_x(r,s)-E_y(r,s)\right)ds\right|.$$

Consequently

$$\sup_{y\in M}\frac{|\tau_r^{x,y}(z)-z|}{d(x,y)}\le (1+O(r^2))\int_0^z |\nabla_x E_\cdot(r,s)|\,ds.$$

Due to $E_\cdot(0,s)\equiv 1=\text{const}$ we find $\lim_{r\to 0}|\nabla_x E_\cdot(r,s)|=0$ and hence

$$\begin{aligned}\lim_{r\to 0}\sup_{y\in M}\frac{|\tau_r^{x,y}(z)-z|}{rd(x,y)} &\le \lim_{r\to 0}\int_0^z \frac{\partial}{\partial r}|\nabla_x E_\cdot(r,s)|\,ds\\ &=\lim_{r\to 0}\int_0^z \langle\frac{\nabla_x E_\cdot(r,s)}{|\nabla_x E_\cdot(r,s)|},\frac{\partial}{\partial r}\nabla_x E_\cdot(r,s)\rangle_{T_xM}ds\\ &=\lim_{r\to 0}\int_0^z \langle\frac{\nabla_x E_\cdot(r,s)}{|\nabla_x E_\cdot(r,s)|},\nabla_x\frac{\partial}{\partial r}E_\cdot(r,s)\rangle_{T_xM}ds. \qquad (14)\end{aligned}$$

The right hand side of (13) yields $\frac{\partial}{\partial r}_{|r=0}E_\cdot(r,s)=0$ from which we see that the integral above vanishes for r tending to 0. This establishes (12) for fixed x. By the smoothness of (M,g) the error term is locally uniform in $x\in M$ and hence it is also globally uniform since M is compact.

The case $\dim M=3$ will show how we can deal with arbitrary dimensions $n\in\mathbb{N}$. Fix $x,y\in M$ as well as some segment γ_{xy} from x to y. By means of the inverse of the exponential map we lift the measures $\sigma_{r,x}$ and $\sigma_{r,y}$ onto the unit spheres in T_xM and T_yM respectively which we disintegrate along the $\dot\gamma_{xy}$-direction as follows. Choose an o.n. basis $\{e_1,e_2,e_3\}\subset T_xM$ with $e_1=\dot\gamma_{xy}$ and $\{e_1',e_2',e_3'\}=/\!\!/_\gamma\{e_1,e_2,e_3\}\subset T_yM$ and for $u\in[-1,1]$ let $S_{r,x}^\perp(u)=\{\exp_x\big(r(ue_1,se_2,te_3)\big)|u^2+s^2+t^2=1\}$ denote the 'orthogonal' part of $S_r(x)$ w.r.t. γ at ue_1. Define a probability measure $c_{r,x}(u)(du)$ on $[-1,1]\subset e_1\mathbb{R}\subset T_xM$ by

$$\begin{aligned}c_{r,x}(u)(du)&=\frac{1}{\mathcal{H}^{n-1}(S_r(x))}\mathcal{H}^{n-2}\big(S_{r,x}^\perp(u)\big)du\\ &=\frac{1}{\int_{S^2}D_x(r,\vartheta)\mathcal{S}^2(d\vartheta)}\int_{S^1_{\sqrt{1-u^2}}}D_x(r,(u,\eta))\mathcal{S}^1(d\eta)du\end{aligned}$$

with the Riemannian volume density

$$D_x(t,\vartheta)=t^{n-1}\left(1-\frac{t^2}{6}\mathrm{Ric}_x(\vartheta,\vartheta)+o(t^2)\right) \qquad (15)$$

and $c_{r,y}(u)$ analogously. Let $\tau_1:[-1,1]e_1\to[-1,1]e_1'$ be the function defined by

$$\int_0^s c_{r,x}(u)du=\int_0^{\tau_1(s)}c_{r,y}(u)du\quad\forall\, s\in[-1,1]. \qquad (16)$$

For each $s\in[-1,1]$ define a transport

$$\tau_s^\perp:S_{r,x}^\perp(s)\to S_{r,y}^\perp(\tau_1(s)),$$

analogously to (11) which preserves the probability measures $\sigma^{\perp}_{r,x}(s)(.)$ and $\sigma^{\perp}_{r,y}(\tau(s))(.)$ that are obtained from conditioning $\sigma_{r,x}$ and $\sigma_{r,y}$ on $S^{\perp}_{r,x}(s)$ and $S^{\perp}_{r,y}(\tau_1(s))$ respectively. Hence the map $\Psi^{x,y}: S_r(x) \to S_r(y)$, $\Psi^{x,y}(z) := \exp_y(r\tau(\frac{1}{r}\log_x(z)))$ induced from

$$\tau = \tau_r^{x,y} : T_xM \supset S^2(x) \to S^2(y) \subset T_yM$$
$$\tau(u) := \left(\tau_1(u_1), \tau^{\perp}_{u_1}((u_2,u_3))\right)$$

will push forward $\sigma_r(x)$ into $\sigma_r(y)$ and it remains to prove the asymptotic estimate (12). Since the distance $|\tau_r^{x,y}(u)-u|^2$ is Euclidean we may use the estimate (12) from the two dimensional case for the $|\tau_r^{x,y}((u_2,u_3))-(u_2,u_3)|$ part which persists also in this situation. Indeed, it is sufficient to note that the expression (14) and hence the error estimate

$$\sup_{z\in[0,2\pi]}\sup_{y\in M}\frac{|\tau^{\perp}_{x,y}((u_2,u_3))-(u_2,u_3)|}{d(x,y)} = o(r)$$

holds true also for the embedded orthogonal spheres $S^{\perp}_{r,x}(u)$, $S^{\perp}_{r,y}(u)$ since they are parallel translates of one another and that by triangle inequality this also generalizes to the situation $\tau^{\perp}_{x,y}: S^{\perp}_{r_1}(x) \to S^{\perp}_{r_2}(y)$ with $r_1, r_2 \le r$. Thus it remains to prove

$$\sup_{u_1\in[-1,1]}\sup_{y\in M}\frac{|\tau_1 u_1 - u_1|}{d(x,y)} = o(r),$$

which follows from (16) by similar arguments which established (12) in the 2D-case. This completes the proof in three dimensions. For arbitrary $n \in \mathbb{N}$ one proceeds in a similar fashion by inductively reducing the problem to lower dimensions. □

References

[ABC+00] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les inégalités de Sobolev logarithmiques.* Société Mathématique de France, Paris, 2000.

[BE84] Dominique Bakry and Michel Emery. Hypercontractivite de semi-groupes de diffusion. *C. R. Acad. Sci., Paris, Ser. I*, 299:775–778, 1984.

[CMS01] Dario Cordero-Erausquin, Robert McCann, and Michael Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamb and Lieb. *Invent. Math.*, 146: 219-257, 2001.

[DR02] Giuseppe Da Prato and Michael Röckner. Singular dissipative stochastic equations in Hilbert spaces. *Probab. Theory Relat. Fields*, 124: 261-303, 2002.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes – Characterization and convergence.* John Wiley & Sons Inc., New York, 1986.

[GHL90] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry.* Springer-Verlag, Berlin, second edition, 1990.

[OV00] Felix Otto and Cedric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[Pet98] Peter Petersen. *Riemannian geometry.* Springer-Verlag, New York, 1998.

[Stu98] Karl-Theodor Sturm. Diffusion processes and heat kernels on metric spaces. *Ann. Probab.*, 26(1):1–55, 1998.

[Stu03] Karl-Theodor Sturm. A semigroup approach to harmonic maps. Prepint no. 39, SFB 611 Univ. Bonn, 2003.

[Vil03] Cedric Villani. *Topics in Mass Transportation.* Graduate Studies in Mathematics. American Mathematical Society, 2003.

[vRen03] Max-K. von Renesse. Intrinsic Coupling on Riemannian Manifolds and Polyhedra. Prepint no. 64, SFB 611 Univ. Bonn, 2003.

[Wan97] Feng-Yu Wang. On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups. *Probab. Theory Related Fields*, 108(1):87–101, 1997.

Address:
Institut für Angewandte Mathematik, Universität Bonn
Wegelerstrasse 6, 53115 Bonn

Fax: +49 - 228 - 734871

E-mail
KTS: sturm@uni-bonn.de, MKvR: kostja@iam.uni-bonn.de