Entropic Measure on Multidimensional Spaces

Karl-Theodor Sturm

Abstract. We construct the entropic measure $\mathbb{P}^\beta$ on compact manifolds of any dimension. It is defined as the push forward of the Dirichlet process (another random probability measure, well-known to exist on spaces of any dimension) under the conjugation map

$$\mathcal{C}: \mathcal{P}(M) \rightarrow \mathcal{P}(M).$$

This conjugation map is a continuous involution. It can be regarded as the canonical extension to higher dimensional spaces of a map between probability measures on 1-dimensional spaces characterized by the fact that the distribution functions of $\mu$ and $\mathcal{C}(\mu)$ are inverse to each other.

We also present an heuristic interpretation of the entropic measure as

$$d\mathbb{P}^\beta(\mu) = \frac{1}{Z} \exp\left(-\beta \cdot \text{Ent}(\mu|m)\right) \cdot d\mathbb{P}^0(\mu).$$

Mathematics Subject Classification (2000). 60G57; 28C20; 49N90; 49Q20; 58J65.

Keywords. Optimal transport, entropic measure, Wasserstein space, entropy, gradient flow, Brenier map, Dirichlet distribution, random probability measure.

1. Introduction

Gradient flows of entropy-like functionals on the Wasserstein space turned out to be a powerful tool in the study of various dissipative PDEs on Euclidean or Riemannian spaces $M$, the prominent example being the heat equation. See e.g. the monographs [Vi03, AGS05] for more examples and further references.

In [RS08], von Renesse and the author presented an approach to stochastic perturbation of the gradient flow of the entropy. It is based on the construction of a Dirichlet form

$$\mathcal{E}(u,u) = \int_{\mathcal{P}(M)} \|
abla u\|^2(\mu) \ d\mathbb{P}^\beta(\mu).$$
where $\|\nabla u\|$ denotes the norm of the gradient in the Wasserstein space $\mathcal{P}(M)$ as introduced by Otto [Ot01]. The fundamental new ingredient was the measure $\mathbb{P}^\beta$ on the Wasserstein space. This so-called *entropic measure* is an interesting and challenging object in its own right. It is formally introduced as

$$
\frac{d\mathbb{P}^\beta}{d\mathbb{P}^0}(\mu) = \frac{1}{Z} \exp \left( -\beta \cdot \text{Ent}(\mu|m) \right) \cdot d\mathbb{P}^0(\mu)
$$

(1.1)

with some (non-existing) ‘uniform distribution’ $\mathbb{P}^0$ on the Wasserstein space $\mathcal{P}(M)$ and the relative entropy as a potential.

A rigorous construction was presented for 1-dimensional spaces. In the case $M = [0,1]$ it is based on the bijections

$$
\mu \xrightarrow{(x)=\mu([0,x])} f \xrightarrow{g=f^{-1}} g \xrightarrow{g(y)=\nu([0,y])} \nu
$$

between probability measures, distribution functions and inverse distribution functions (where $f^{-1}(y) = \inf\{x \geq 0 : f(x) \geq y\}$ more precisely denotes the ‘right inverse’ of $f$). If $\mathcal{C}: \mathcal{P}(M) \to \mathcal{P}(M)$ denotes the map $\mu \mapsto \nu$ then the entropic measure $\mathbb{P}^\beta$ is just the push forward under $\mathcal{C}$ of the Dirichlet-Ferguson process $\mathbb{Q}^\beta$. The latter is a random probability measure which is well-defined on every probability space.

For long time it seemed that the previous construction is definitively limited to dimension 1 since it heavily depends on the use of distribution functions (and inverse distribution functions), – objects which do not exist in higher dimensions. The crucial observation to overcome this restriction is to interpret $g$ as the unique *optimal transport map* which pushes forward $m$ (the normalized uniform distribution on $M$) to $\mu$:

$$
\mu = g_* m.
$$

Due to Brenier [Br87] and McCann [Mc01] such a ‘monotone map’ exists for each probability measure $\mu$ on a Riemannian manifold of arbitrary dimension. Moreover, also in higher dimensions such a monotone map $g$ has a unique generalized inverse $f$, again being a monotone map (with generalized inverse being $g$). This observation allows to define the *conjugation map*

$$
\mathcal{C}: \mathcal{P}(M) \to \mathcal{P}(M), \ \mu \mapsto \nu
$$

for any compact manifold $M$. It is a continuous involution. By means of this map we define the entropic measure as follows:

$$
\mathbb{P}^\beta := \mathcal{C}_* \mathbb{Q}^\beta
$$

where $\mathbb{Q}^\beta$ denotes the Dirichlet-Ferguson process on $M$ with intensity measure $\beta \cdot m$. (Actually, such a random probability measure exists on every probability space.)

In order to justify our definition of the entropic measure by some heuristic argument let us assume that $\mathbb{P}^\beta$ were given as in (1.1). The identity $\mathbb{Q}^\beta = \mathcal{C}_* \mathbb{P}^\beta$ then
defines a probability measure which satisfies
\[ dQ^\beta(\nu) = \frac{1}{Z} \exp \left( -\beta \cdot \text{Ent}(m|\nu) \right) \cdot dQ^0(\nu). \] (1.2)

Given a measurable partition \( M = \bigcup_{i=1}^N M_i \) and approximating arbitrary probability measures \( \nu \) by measures with constant density on each of the sets \( M_i \) of the partition the previous ansatz (1.2) yields – after some manipulations –

\[ Q^\beta_{M_1,\ldots,M_N}(dx) = \frac{\Gamma(\beta)}\prod_{i=1}^N \Gamma(\beta m(M_i)) \times \delta_{\left(1 - \sum_{i=1}^{N-1} x_i \right)}(dx_N)dx_{N-1}\ldots dx_1. \]

These are, indeed, the finite dimensional distributions of the Dirichlet-Ferguson process.

### 2. Spaces of Convex Functions and Monotone Maps

Throughout this paper, \( M \) will be a compact subset of a complete Riemannian manifold \( \hat{M} \) with Riemannian distance \( d \) and \( m \) will denote a probability measure with support \( M \), absolutely continuous with respect to the volume measure. We assume that it satisfies a Poincaré inequality:

\[ \exists c > 0, \quad \int_M |\nabla u|^2 \, dm \geq c \int_M u^2 \, dm \]

for all weakly differentiable \( u : M \to \mathbb{R} \) with \( \int_M u \, dm = 0 \).

For compact Riemannian manifolds, there is a canonical choice for \( m \), namely, the normalized Riemannian volume measure. The freedom to choose \( m \) arbitrarily might be of advantage in view of future extensions: For Finsler manifolds and for non-compact Riemannian manifolds there is no such canonical probability measure.

The main ingredient of our construction below will be the Brenier-McCann representation of optimal transport in terms of gradients of convex functions.

**Definition 2.1.** A function \( \varphi : M \to \mathbb{R} \) is called \( d^2/2 \)-convex if there exists a function \( \psi : M \to \mathbb{R} \) such that

\[ \varphi(x) = -\inf_{y \in M} \left[ \frac{1}{2} d^2(x,y) + \psi(y) \right] \]

for all \( x \in M \). In this case, \( \varphi \) is called generalized Legendre transform of \( \psi \) or conjugate of \( \psi \) and denoted by \( \varphi = \psi^\circ \).
Let us summarize some of the basic facts on $d^2/2$-convex functions. See [Ro70], [Rü96], [Mc01] and [Vi08] for details.

Lemma 2.2. (i) A function $\varphi$ is $d^2/2$-convex if and only if $\varphi^{cc} = \varphi$.

(ii) Every $d^2/2$-convex function is bounded, Lipschitz continuous and differentiable almost everywhere with gradient bounded by $D = \sup_{x,y \in M} d(x, y)$.

In the sequel, $\mathcal{K} = \mathcal{K}(M)$ will denote the set of $d^2/2$-convex functions on $M$ and $\tilde{\mathcal{K}} = \tilde{\mathcal{K}}(M)$ will denote the set of equivalence classes in $\mathcal{K}$ with $\varphi_1 \sim \varphi_2$ iff $\varphi_1 - \varphi_2$ is constant. $\mathcal{K}$ will be regarded as a subset of the Sobolev space $H^1(M, m)$ with norm

$$\| u \|_{H^1} = \left[ \int_M |\nabla u|^2 \, dm + \int_M u^2 dm \right]^{1/2},$$

and $\tilde{\mathcal{K}} = \mathcal{K}/\text{const}$ will be regarded as a subset of the space $\tilde{H}^1 = H^1/\text{const}$ with norm

$$\| u \|_{\tilde{H}^1} = \left[ \int_M |\nabla u|^2 \, dm \right]^{1/2}.$$

Proposition 2.3. For each Borel map $g : M \to M$ the following are equivalent:

(i) $\exists \varphi \in \tilde{\mathcal{K}} : g = \exp(\nabla \varphi)$ a.e. on $M$;

(ii) $g$ is an optimal transport map from $m$ to $f \ast m$ in the sense that it is a minimizer of $h \mapsto \int_M d^2(x, h(x))m(dx)$ among all Borel maps $h : M \to M$ with $h \ast m = g \ast m$.

In this case, the function $\varphi \in \tilde{\mathcal{K}}$ in (i) is defined uniquely. Moreover, in (ii) the map $f$ is the unique minimizer of the given minimization problem.

A Borel map $g : M \to M$ satisfying the properties of the previous proposition will be called monotone map or optimal Lebesgue transport. The set of $m$-equivalence classes of such maps will be denoted by $\mathcal{G} = \mathcal{G}(M)$. Note that $\mathcal{G}(M)$ does not depend on the choice of $m$ (as long as $m$ is absolutely continuous with full support)! $\mathcal{G}(M)$ will be regarded as a subset of the space of maps $L^2((M, m)(M, d))$ with metric $d_2(f, g) = \left[ \int_M d^2(f(x), g(x))m(dx) \right]^{1/2}$.

According to our definitions, the map $\Upsilon : \varphi \mapsto \exp(\nabla \varphi)$ defines a bijection between $\tilde{\mathcal{K}}$ and $\mathcal{G}$. Recall that $\mathcal{P} = \mathcal{P}(M)$ denotes the set of probability measures $\mu$ on $M$ (equipped with its Borel $\sigma$-field).

Proposition 2.4. The map $\chi : g \mapsto g \ast m$ defines a bijection between $\mathcal{G}$ and $\mathcal{P}(M)$. That is, for each $\mu \in \mathcal{P}$ there exists a unique $g \in \mathcal{G}$ - called Brenier map of $\mu$ - with $\mu = g \ast m$.

1A function $\varphi$ is $d^2/2$-convex in our sense if and only if the function $-\varphi$ is $c$-concave in the sense of [Ro70], [Rü96], [Mc01], [Vi08] with cost function $c(x, y) = d^2(x, y)/2$. In our presentation, the $c$ stands for 'conjugate'. For the relation between $d^2/2$-convexity and usual convexity on Euclidean space we refer to chapter 4.
The map \( \chi \) of course strongly depends on the choice of the measure \( m \). (If there is any ambiguity we denote it by \( \chi_m \).)

Due to the previous observations, there exist canonical bijections \( \Upsilon \) and \( \chi \) between the sets \( \hat{K} \), \( \mathcal{G} \) and \( \mathcal{P} \). Actually, these bijections are even homeomorphisms with respect to the natural topologies on these spaces.

**Proposition 2.5.** Consider any sequence \( \{\varphi_n\}_{n \in \mathbb{N}} \) in \( \hat{K} \) with corresponding sequences \( \{g_n\}_{n \in \mathbb{N}} = \{\Upsilon(\varphi_n)\}_{n \in \mathbb{N}} \) in \( \mathcal{G} \) and \( \{\mu_n\}_{n \in \mathbb{N}} = \{\chi(g_n)\}_{n \in \mathbb{N}} \) in \( \mathcal{P} \) and let \( \varphi \in \hat{K} \), \( g = \Upsilon(\varphi) \in \mathcal{G} \), \( \mu = \chi(g) \in \mathcal{P} \). Then the following are equivalent:

1. \( \varphi_n \rightarrow \varphi \) in \( H_1 \)
2. \( g_n \rightarrow g \) in \( L^2((M,m),(M,d)) \)
3. \( g_n \rightarrow g \) in \( m \)-probability on \( M \)
4. \( \mu_n \rightarrow \mu \) in \( L^2 \)-Wasserstein distance \( d_W \)
5. \( \mu_n \rightarrow \mu \) weakly.

**Proof.** (i) \( \Leftrightarrow \) (ii) Compactness of \( M \) and smoothness of the exponential map imply that there exists \( \delta > 0 \) such that \( \forall x \in M \), \( \forall v_1, v_2 \in T_x M \) with \( |v_1|,|v_2| \leq D \) and \( |v_1 - v_2| < \delta \):

\[
\frac{1}{2} \leq d(\exp_x v_1, \exp_x v_2)/ |v_1 - v_2|_{T_x M} \leq 2.
\]

Hence, \( \varphi_n \rightarrow \varphi \) in \( H^1 \), that is \( \int_M |\nabla \varphi_n(x) - \nabla \varphi(x)|^2_{g,M} m(dx) \rightarrow 0 \), is equivalent to \( \int_M d^2(g_n(x),g(x))m(dx) \rightarrow 0 \), that is, to \( g_n \rightarrow g \) in \( L^2((M,m),(M,d)) \).

(ii) \( \Leftrightarrow \) (iii) Standard fact from integration theory (taking into account that \( d(g_n,g) \) is uniformly bounded due to compactness of \( M \)).

(iii) \( \Leftrightarrow \) (iv) If \( \mu_n = (g_n)_* m \) and \( \mu_n = g_* m \) then \( (g_n,g)_* m \) is a coupling of \( \mu_n \) and \( \mu \). Hence,

\[
d^2_W(\mu_n,\mu) \leq \int_M d^2(g_n(x),g(x))m(dx). \tag{2.1}
\]

(iv) \( \Leftrightarrow \) (v) Trivial.

(ii) \( \Leftrightarrow \) (iv) \[Vi08\], Corollary 5.21.

**Remark 2.6.** Since \( M \) is compact, assertion (ii) of the previous Proposition is equivalent to

(\[\text{ iii'}\]) \( g_n \rightarrow g \) in \( L^p((M,m),(M,d)) \)

for any \( p \in [1,\infty) \) and similarly, assertion (iv) is equivalent to

(\[\text{ iv'}\]) \( \mu_n \rightarrow \mu \) in \( L^p \)-Wasserstein distance.

**Remark 2.7.** In \( n = 1 \), the inequality in (2.1) is actually an equality. In other words, the map

\( \chi : (\mathcal{G},d_3) \rightarrow (\mathcal{P},d_W) \)

is an isometry. This is no longer true in higher dimensions.

The well-known fact (Prohorov’s theorem) that the space of probability measures on a compact space is itself compact, together with the previous continuity results immediately implies compactness of \( \hat{K} \) and \( \mathcal{G} \).
Corollary 2.8. (i) $\tilde{K}$ is a compact subset of $\tilde{H}^1$.
(ii) $G$ is a compact subset of $L^2((M,m),(M,d))$.

3. The Conjugation Map

Let us recall the definition of the conjugation map $C_K : \varphi \mapsto \varphi^c$ acting on functions $\varphi : M \to \mathbb{R}$ as follows

$$\varphi^c(x) = -\inf_{y \in M} \left[ \frac{1}{2} d^2(x,y) + \varphi(y) \right].$$

The map $C_K$ maps $K$ bijective onto itself with $C_K^2 = I$. For each $\lambda \in \mathbb{R}$, $C_K(\varphi + \lambda) = C_K(\varphi) - \lambda$. Hence, $C_K$ extends to a bijection $C_K : \tilde{K} \to \tilde{K}$. Composing this map with the bijections $\chi : G \to P$ and $\Upsilon : \tilde{K} \to G$ we obtain involutive bijections $C_G = \Upsilon \circ C_{\tilde{K}} \circ \Upsilon^{-1} : G \to G$

and

$$C_P = \chi \circ C_G \circ \chi^{-1} : P \to P,$$

called conjugation map on $G$ or on $P$, respectively. Given a monotone map $g \in G$, the monotone map

$$g^c := C_G(g)$$

will be called conjugate map or generalized inverse map; given a probability measure $\mu \in P$ the probability measure

$$\mu^c := C_P(\mu)$$

will be called conjugate measure.

Example 3.1. (i) Let $M = S^n$ be the $n$-dimensional sphere, and $m$ be the normalized Riemannian volume measure. Put

$$\mu = \lambda \delta_a + (1 - \lambda) m$$

for some point $a \in M$ and $\lambda \in \]0,1[$. Then

$$\mu^c = \frac{1}{1 - \lambda} 1_{M \setminus B_r(a)} \cdot m$$

where $r > 0$ is such that $m(B_r(a)) = \lambda$.

[ Proof. The optimal transport map $g = \exp(\nabla \varphi)$ which pushes $m$ to $\mu$ is determined by the $d^2/2$-convex function

$$\varphi = \left\{ \begin{array}{ll} \frac{1}{2} \frac{r^2 - d^2(a',x)}{r^2 - r^2} & \text{in } B_r(a) \\
\frac{d^2(a',x) - (\pi - r)^2}{2(\pi - r)} & \text{in } B_{\pi - r}(a') = M \setminus B_r(a) \end{array} \right.$$.

Its conjugate is the function

$$\varphi^c(y) = -\frac{r}{2\pi} d^2(a',y) + \frac{1}{2} r(\pi - r).$$]
(ii) Let $M = S^n$, the $n$-dimensional sphere, and $\mu = \delta_a$ for some $a \in M$. Then $\mu^\ell = \delta_{a'}$ with $a' \in M$ being the antipodal point of $a$.

[Proof. Limit of (i) as $\lambda \to 1$. Alternatively: explicit calculations with $\varphi(x) = \frac{1}{2}[\pi^2 - d^2(a, x)]$]

\[ \varphi(x) = \frac{1}{2}[\pi^2 - d^2(a, x)] \]

and

\[ \varphi(x) = \sup_x \left( -\frac{1}{2}d^2(x, y) + \frac{1}{2}d^2(a, x) - \frac{1}{2}\pi^2 \right) = -\frac{1}{2}d^2(a', y). \]

(iii) Let $M = S^n$, the $n$-dimensional sphere, and $\mu = \frac{1}{2}\delta_a + \frac{1}{2}\delta_{a'}$ with north and south pole $a, a' \in M$. Then $\mu^\ell$ is the uniform distribution on the equator, the $(n - 1)$-dimensional set $Z$ of points of equal distance to $a, a'$.

(iv) Let $M = S^1$ be the circle of length 1, $m = \text{uniform distribution and}$

\[ \mu = \sum_{i=1}^{k} \alpha_i \delta_{x_i} \]

with points $x_1 < x_2 < \ldots < x_k < x_1$ in cyclic order on $S^1$ and numbers $\alpha_i \in [0, 1]$, $\sum \alpha_i = 1$. Then

\[ \mu^\ell = \sum_{i=1}^{k} \beta_i \delta_{y_i} \]

with $\beta_i = |x_{i+1} - x_i|$ and points $y_1 < y_2 < \ldots < y_k < y_{k+1} = y_1$ on $S^1$ satisfying

\[ |y_{i+1} - y_i| = \alpha_{i+1}. \]

[Proof. Embedding in $\mathbb{R}^1$ and explicit calculation of distribution and inverse distribution functions.]

Remark 3.2. The conjugation map

\[ \mathcal{C}_P : P \to P \]
depends on the choice of the reference measure $m$ on $M$. Actually, we can choose two different probability measures $m_1, m_2$ and consider $\mathcal{E}_\mathcal{P} = \chi_{m_2} \circ \mathcal{E}_G \circ \chi_{m_1}^{-1}$.

**Proposition 3.3.** Let $\mu = g_* m \in \mathcal{P}$ be absolutely continuous with density $\eta = \frac{d\mu}{dm}$. Put $f = g^c$ and $\nu = f_* m = \mu^c$.

(i) If $\eta > 0$ a.s. then the measure $\nu$ is absolutely continuous with density $\rho = \frac{d\nu}{dm} > 0$ satisfying

$$
\eta(x) \cdot \rho(f(x)) = \rho(x) \cdot \eta(g(x)) = 1 \quad \text{for a.e. } x \in M.
$$

(ii) If $\nu$ is absolutely continuous then $f(g(x)) = g(f(x)) = x$ for a.e. $x \in M$.

(iii) Under the previous assumption the Jacobian $\det Df(x)$ and $\det Dg(x)$ exist for almost every $x \in M$ and satisfy

$$
\det Df(g(x)) \cdot \det Dg(x) = \det Df(x) \cdot \det Dg(f(x)) = 1,
$$

$$
\sigma(x) \cdot \eta(x) = \sigma(f(x)) \cdot \det Df(x), \quad \sigma(x) \cdot \rho(x) = \sigma(g(x)) \cdot \det Dg(x)
$$

for almost every $x \in M$ where $\sigma = \frac{dm}{dvol}$ denotes the density of the reference measure $m$ with respect to the Riemannian volume measure $\text{vol}$.

**Proof.** (i) For each Borel function $v : M \rightarrow \mathbb{R}_+$

$$
\int_M v \, d\nu = \int_M v \circ f \, d\mu = \int_M v \circ f \cdot \frac{1}{\eta} \, d\mu = \int_M v \circ f \cdot \frac{1}{\eta(g \circ f)} \, d\mu = \int_M v \cdot \frac{1}{\eta \circ g} \, dm.
$$

Hence, $\nu$ is absolutely continuous with respect to $m$ with density $\frac{1}{\eta \circ g}$. Interchanging the roles of $\mu$ and $\nu$ (as well as $f$ and $g$) yields the second claim.

(ii), (iii) Part of Brenier- McCann representation result of optimal transports. □

**Corollary 3.4.** Under the assumption $\eta > 0$ of the previous Proposition:

$$
\text{Ent}(\mu^c | m) = \text{Ent}(m | \mu).
$$

**Proof.** With notations from above

$$
\text{Ent}(\mu^c | m) = \int \rho \log \rho \, dm = \int \frac{1}{\eta \circ g} \log \frac{1}{\eta \circ g} \, dm = \int \frac{1}{\eta} \log \frac{1}{\eta} \, d\mu = \text{Ent}(m | \mu).
$$

□

**Lemma 3.5.** The conjugation map

$$
\mathcal{E}_\mathcal{K} : \mathcal{K} \rightarrow \mathcal{K}
$$

is continuous.

**Proof.** To simplify notation denote $\mathcal{E}_\mathcal{K}$ by $\mathcal{E}$. Choose a countable dense set $\{y_i\}_{i \in \mathbb{N}}$ in $M$ and for $k \in \mathbb{N}$ define $\mathcal{E}_k : \varphi \mapsto \varphi_k$ on $\mathcal{K}$ by $\varphi_k(x) = - \inf_{i=1,...,k} \left[ \frac{1}{2} d^2(x, y_i) + \langle \varphi(y_i) \rangle \right]$. Then as $k \rightarrow \infty$

$$
\varphi_k \not\rightarrow \varphi^c \quad \text{pointwise on } M.
$$

Recall that each $\varphi \in \mathcal{K}$ is Lipschitz continuous with Lipschitz constant $D$. 
For each $\varepsilon > 0$ choose $k = k(\varepsilon) \in \mathbb{N}$ such that the set $\{y_i\}_{i=1, \ldots, k(\varepsilon)}$ is an $\varepsilon$-covering of the compact space $M$. Then

$$|C_k(\varphi)(x) - C(\varphi)(x)| \leq \sup_{y \in M} \inf_{i=1, \ldots, k} \left| \frac{1}{2} d^2(x, y) - \frac{1}{2} d^2(x, y_i) + \varphi(y) - \varphi(y_i) \right|$$

$$\leq \sup_{y \in M} \inf_{i=1, \ldots, k} 2D \cdot d(y, y_i) \leq 2D\varepsilon$$

uniformly in $x \in M$ and $\varphi \in \mathcal{K}$.

Now let us consider a sequence $(\varphi_l)_l \in \mathbb{N}$ in $\mathcal{K}$ with $\varphi_l \to \varphi$ in $H^1(M)$. Then for each $k \in \mathbb{N}$ as $l \to \infty$

$$C_k(\varphi_l) \to C_k(\varphi)$$

pointwise on $M$ and thus also in $L^2(M)$. Together with the previous uniform convergence of $C_k \to C$ it implies

$$C(\varphi_l) \to C(\varphi)$$

in $L^2(M)$ as $l \to \infty$. Moreover, we know that $\{C(\varphi_l)_l \in \mathbb{N}$ is bounded in $H^1(M)$ (since all gradients are bounded by $D$). Therefore, finally

$$C(\varphi_l) \to C(\varphi)$$

in $H^1(M)$ as $l \to \infty$. This proves the continuity of $C : \mathcal{K} \to \mathcal{K}$ with respect to the $H^1$-norm. □

**Theorem 3.6.** The conjugation map

$$C_P : \mathcal{P} \to \mathcal{P}$$

is continuous (with respect to the weak topology).

**Proof.** Let us first prove continuity of the conjugation map $C_\mathcal{K} : \hat{\mathcal{K}} \to \hat{\mathcal{K}}$ (with respect to the $H^1$-norm on $\hat{\mathcal{K}}$). Indeed, this follows from the previous continuity result together with the facts that the embedding $H^1 \to \hat{H^1}$, $\varphi \mapsto \hat{\varphi} = \{\varphi + c : c \in \mathbb{R}\}$ is continuous (trivial fact) and that the map $\hat{H^1} \to H^1$, $\hat{\varphi} = \{\varphi + c : c \in \mathbb{R}\} \mapsto \varphi - \int_M \varphi dm$ is continuous (consequence of Poincaré inequality).

This in turn implies, due to Proposition 2.5, that the conjugation map $C_\mathcal{G} : \hat{\mathcal{G}} \to \hat{\mathcal{G}}$ is continuous (with respect to the $L^2$-metric on $\hat{\mathcal{G}}$). Moreover, due to the same Proposition it therefore also implies that the conjugation map

$$C_P : \mathcal{P} \to \mathcal{P}$$

is continuous (with respect to the weak topology). □

**Remark 3.7.** In dimension $n = 1$, the conjugation map $C_\mathcal{G} : \mathcal{G} \to \mathcal{G}$ is even an isometry from $\mathcal{G}$, equipped with the $L^1$-metric, into itself.
4. Example: The Conjugation Map on $M \subset \mathbb{R}^n$

In this chapter, we will study in detail the Euclidean case. We assume that $M$ is a compact convex subset of $\mathbb{R}^n$. (The convexity assumption is to simplify notations and results.) The probability measure $m$ is assumed to be absolutely continuous with full support on $M$.

A function $\varphi : M \to \mathbb{R}$ is $d^2/2$-convex if and only if the function $\varphi_1(x) = \varphi(x) + |x|^2/2$ is convex in the usual sense:

$$\varphi_1(\lambda x + (1 - \lambda)y) \leq \lambda \varphi_1(x) + (1 - \lambda)\varphi_1(y)$$

(for all $x, y \in M$ and $\lambda \in [0, 1]$) and if its subdifferential lies in $M$:

$$\partial \varphi_1(x) \subset M$$

for all $x \in M$.

A function $\psi$ is the conjugate of $\varphi$ if and only if the function $\psi_1(y) = \psi(y) + |y|^2/2$ is the Legendre-Fenchel transform of $\varphi_1$:

$$\psi_1(y) = \sup_{x \in M} [(x, y) - \varphi_1(x)].$$

A Borel map $g : M \to M$ is monotone if and only if

$$\langle g(x) - g(y), x - y \rangle \geq 0$$

for a.e. $x, y \in M$. Equivalently, $g$ is monotone if and only if $g = \nabla \varphi_1$ for some convex $\varphi_1 : M \to \mathbb{R}$.

**Lemma 4.1.** (i) If $\mu = \lambda \delta_z + (1 - \lambda) \nu$ then there exists an open convex set $U \subset M$ with $m(U) = \lambda$ such that the optimal transport map $g$ with $g_* m = \mu$ satisfies $g \equiv z$ a.e. on $U$.

(ii) The conjugate measure $\mu^\ast$ does not charge $U$:

$$\mu^\ast(U) = 0.$$

**Proof.** (i) Linearity of the problem allows to assume that $z = 0$. Let $g = \nabla \varphi_1$ denote the optimal transport map with $\varphi_1$ being an appropriate convex function. Let $V$ be the subset of points in $M$ in which $\varphi_1$ is weakly differentiable with vanishing gradient. By the push forward property it follows that $m(V) = \lambda$. Firstly, then convexity of $\varphi_1$ implies that $\varphi_1$ has to be constant on $V$, say $\varphi_1 \equiv \alpha$ on $V$. Secondly, the latter implies that $\varphi_1 \equiv \alpha$ on the convex hull $W$ of $V$. The interior $U$ of this convex set $W$ has volume $m(U) = m(W) \geq m(V) = \lambda$ and $\varphi_1$ is constant on $U$, hence, differentiable with vanishing gradient. Thus finally $U \subset V$ and $m(U) = \lambda$.

(ii) Let $\mu_\epsilon, \epsilon \in [0, 1]$, denote the intermediate points on the geodesic from $\mu_0 = \mu$ to $\mu_1 = m$. Then $\mu_\epsilon = (g_\epsilon)_* m$ with $g_\epsilon = \exp((1 - \epsilon) \nabla \varphi) = \epsilon \cdot Id + (1 - \epsilon) \cdot g$ and each $\mu_\epsilon$ is absolutely continuous w.r. to $m$. Hence, $g_\epsilon^\ast = g_\epsilon^{-1}$ a.e. on $M$. Therefore, the conjugate measure $\mu_\epsilon^\ast$ satisfies

$$\mu_\epsilon^\ast(U) = m((g_\epsilon^\ast)^{-1}(U)) = m(g_\epsilon(U)) = \epsilon^n \cdot m(U) = \epsilon^n \cdot \lambda.$$
Now obviously \( \mu_{\epsilon} \to \mu \) as \( \epsilon \to 0 \). According to Theorem 3.6 this implies \( \mu_{\epsilon}^c \to \mu^c \) and thus (since \( U \) is open)

\[
\mu^c(U) \leq \liminf_{\epsilon \to 0} \mu_{\epsilon}^c(U) = 0.
\]

\[\square\]

**Theorem 4.2.** (i) If \( \mu = \sum_{i=1}^{N} \lambda_i \delta_{z_i} \), with \( N \in \mathbb{N} \cup \{\infty\} \) then there exist disjoint convex open sets \( U_i \subset M \) with \( m(U_i) = \lambda_i \) such that the optimal transport map \( g = \nabla \varphi_1 \) with \( g_* m = \mu \) satisfies \( g \equiv z_i \) on each of the \( U_i, i \in \mathbb{N} \).

The measure \( \mu^c \) is supported by the compact \( m \)-zero set \( M \setminus \bigcup_{i=1}^{N} U_i \).

(ii) Each of the sets \( U_i \) is the interior of \( M \cap A_i \) where

\[
A_i = \{ x \in \mathbb{R}^n : \varphi_1(x) = \langle z_i, x \rangle + \alpha_i \}
\]

and

\[
\varphi_1(x) = \sup_{i=1,\ldots,N} \left[ \langle z_i, x \rangle + \alpha_i \right]
\]

with numbers \( \alpha_i \) to be chosen in such a way that \( m(A_i) = \lambda_i \).

(iii) If \( N < \infty \) then each of the sets \( A_i \subset \mathbb{R}^n, i = 1,\ldots,N \) is a convex polytope.

The decomposition \( \mathbb{R}^n = \bigcup_{i=1}^{N} A_i \) is a Laguerre tessellation (see e.g. [LZ08] and references therein).

The compact \( m \)-zero set \( M \setminus \bigcup_{i=1}^{N} U_i \) which supports \( \mu^c \) has finite \((n-1)\)-dimensional Hausdorff measure.

**Corollary 4.3.** (i) If \( \mu \) is discrete then the topological support of \( \mu^c \) is a \( m \)-zero set.

In particular, \( \mu^c \) has no absolutely continuous part.

(ii) If \( \mu \) has full topological support then \( \mu^c \) has no atoms.

**Proof.** (i) Obvious from the previous theorem.

(ii) If \( \mu^c \) had an atom (of mass \( \lambda > 0 \)) then according to the previous lemma there would be a convex open set \( U \) (of volume \( m(U) = \lambda \)) such that \( \mu(U) = (\mu^c)^c(U) = 0 \).

\[\square\]

5. The Entropic Measure – Heuristics

Our goal is to construct a canonical probability measure \( \mathbb{P}^\beta \) on the Wasserstein space \( \mathcal{P} = \mathcal{P}(M) \) over a compact Riemannian manifold, according to the formal ansatz

\[
\mathbb{P}^\beta(d\mu) = \frac{1}{Z} e^{-\beta \text{Ent}(\mu|m)} \mathbb{P}^0(d\mu).
\]

Here \( \text{Ent}(\cdot \mid m) \) is the relative entropy with respect to the reference measure \( m \), \( \beta \) is a constant \( \beta > 0 \) (‘the inverse temperature’) and \( \mathbb{P}^0 \) should denote a (non-existing) ‘uniform distribution’ on \( \mathcal{P}(M) \). \( Z \) should denote a normalizing constant.

Using the conjugation map \( \mathcal{C} : \mathcal{P}(M) \to \mathcal{P}(M) \) and denoting \( \mathbb{Q}^\beta := (\mathcal{C}_\mathbb{P})_* \mathbb{P}^\beta, \mathbb{Q}^0 := (\mathcal{C}_\mathbb{P})_* \mathbb{P}^0 \) the above problem can be reformulated as follows:
Construct a probability measure \( Q^\beta \) on \( \mathcal{P}(M) \) such that – at least formally –

\[
Q^\beta(d\nu) = \frac{1}{Z} e^{-\beta \text{Ent}(m | \nu)} Q^0(d\nu)
\]  

(5.1)

with some ‘uniform distribution’ \( Q^0 \) in \( \mathcal{P}(M) \). Here, we have used the fact that

\[
\text{Ent}(\nu^\delta | m) = \text{Ent}(m | \nu)
\]

(Corollary 3.4), at least if \( \nu \ll m \) with \( \frac{d\nu}{dm} > 0 \) almost everywhere.

Probability measures \( P(d\mu) \) on \( \mathcal{P}(M) \) – so called random probability measures on \( M \) – are uniquely determined by the distributions \( PM_{M_1,\ldots,M_N} \) of the random vectors

\[
(\mu(M_1),\ldots,\mu(M_N))
\]

for all \( N \in \mathbb{N} \) and all measurable partitions

\[
M = \bigcup_{i=1}^N M_i
\]

of \( M \). Conversely, if a consistent family \( PM_{M_1,\ldots,M_N} \) of probability measures on \( [0,1]^N \) (for all \( N \in \mathbb{N} \) and all measurable partitions \( M = \bigcup_{i=1}^N M_i \)) is given then there exists a random probability measure \( P \) such that

\[
P_{M_1,\ldots,M_N}(A) = P((\mu(M_1),\ldots,\mu(M_N)) \in A)
\]

for all measurable \( A \subset [0,1]^N \), all \( N \in \mathbb{N} \) and all partitions \( M = \bigcup_{i=1}^N M_i \).

Given a measurable partition \( M = \bigcup_{i=1}^N M_i \) the ansatz (5.1) yields the following characterization of the finite dimensional distribution on \( [0,1]^N \)

\[
Q^\beta_{M_1,\ldots,M_N}(dx) = \frac{1}{Z_N} e^{-\beta S_{M_1,\ldots,M_N}(x)} q_{M_1,\ldots,M_N}(dx)
\]  

(5.2)

where \( S_{M_1,\ldots,M_N}(x) \) denotes the conditional expectation (with respect to \( Q^0 \)) of \( S(\cdot) = \text{Ent}(m | \cdot) \) under the condition \( \nu(M_1) = x_1,\ldots,\nu(M_N) = x_N \).

Moreover, \( q_{M_1,\ldots,M_N}(dx) = Q^0((\nu(M_1),\ldots,\nu(M_N)) \in dx) \) denotes the distribution of the random vector \((\nu(M_1),\ldots,\nu(M_N))\) in the simplex

\[
\sum_N = \left\{ x \in [0,1]^N : \sum_{i=1}^N x_i = 1 \right\}
\]

According to our choice of \( Q^0 \), the measure \( q_{M_1,\ldots,M_N} \) should be the ‘uniform distribution’ in the simplex \( \sum_N \). In [RS08] we argued that the canonical choice for a ‘uniform distribution’ in \( \sum_N \) is the measure

\[
q_N(dx) = c \cdot \frac{dx_1 \cdots dx_{N-1}}{x_1 \cdot x_2 \cdots x_{N-1} \cdot x_N} \cdot \delta_{1-\frac{1}{\sum_{i=1}^N x_i}}(dx_N).
\]  

(5.3)

It remains to get hands on \( S_{M_1,\ldots,M_N}(x) \), the conditional expectation of \( S(\cdot) = \text{Ent}(m | \cdot) \) under the constraint \( \nu(M_1) = x_1,\ldots,\nu(M_N) = x_N \). We simply replace it by \( S_{M_1,\ldots,M_N}(x) \), the minimum of \( \nu \rightarrow \text{Ent}(m | \nu) \) under the constraint \( \nu(M_1) = x_1,\ldots,\nu(M_N) = x_N \).

Obviously, this minimum is attained at a measure with constant density on each of the sets \( M_i \) of the partition, that is

\[
\nu = \sum_{i=1}^N \frac{x_i}{m(M_i)} 1_{M_i} m.
\]
Hence,
\[ S_{M_1, \ldots, M_N}(x) = - \sum_{i=1}^{N} \log \frac{x_i}{m(M_i)} \cdot m(M_i). \tag{5.4} \]

Replacing \( S_{M_1, \ldots, M_N} \) by \( S_{M_1, \ldots, M_N} \) in (5.2), the latter yields
\[
Q^\beta_{M_1, \ldots, M_N}(dx) = \frac{c \cdot e^{-\beta S_{M_1, \ldots, M_N}(x)}}{\prod_{i=1}^{N} \Gamma(\beta m(M_i))} \cdot x_1^{\beta m(M_1) - 1} \cdot \cdots \cdot x_N^{\beta m(M_N) - 1} \times \delta_{1 - \sum_{i=1}^{N-1} x_i} \]
\[
\Gamma(\beta) \cdot \prod_{i=1}^{N} \Gamma(\beta m(M_i)) \times (dx_1)dx_{N-1} \ldots dx_1.
\]

This, indeed, defines a projective family! Hence, the random probability measure \( Q^\beta \) exists and is uniquely defined. It is the well-known Dirichlet-Ferguson process. Therefore, in turn, also the random probability measure \( P^\beta = (\mathcal{C}_P) \), \( Q^\beta \) exists uniquely.

6. The Entropic Measure – Rigorous Definition

Definition 6.1. Given any compact Riemannian space \((M, d, m)\) and any parameter \( \beta > 0 \) the entropic measure
\[
P^\beta := (\mathcal{C}_P), Q^\beta
\]
is the push forward of the Dirichlet-Ferguson process \( Q^\beta \) (with reference measure \( \beta m \)) under the conjugation map \( \mathcal{C}_P : \mathcal{P}(M) \to \mathcal{P}(M) \).

\( P^\beta \) as well as \( Q^\beta \) are probability measures on the compact space \( \mathcal{P} = \mathcal{P}(M) \) of probability measures on \( M \). Recall the definition of the Dirichlet-Ferguson process \( Q^\beta \) [Fe73]: For each measurable partition \( M = \bigcup_{i=1}^{N} M_i \), the random vector \((\nu(M_1), \ldots, \nu(M_N))\) is distributed according to a Dirichlet distribution with parameters \((\beta m(M_1), \ldots, \beta m(M_N))\). That is, for any bounded Borel function \( u : \mathbb{R}^N \to \mathbb{R} \)
\[
\int_{\mathcal{P}(M)} u(\nu(M_1), \ldots, \nu(M_N)) Q^\beta(d\nu) =
\]
\[
\frac{\prod_{i=1}^{N} \Gamma(\beta m(M_i)) \cdot \int_{[0,1]^N} u(x_1, \ldots, x_N) \cdot x_1^{\beta m(M_1) - 1} \cdot \cdots \cdot x_N^{\beta m(M_N) - 1} \times \delta_{1 - \sum_{i=1}^{N-1} x_i} \cdot (dx_1)dx_{N-1} \ldots dx_1.}
\]

The latter uniquely characterizes the ‘random probability measure’ \( Q^\beta \). The existence (as a projective limit) is guaranteed by Kolmogorov’s theorem.
An alternative, more direct construction is as follows: Let \((x_i)_{i \in \mathbb{N}}\) be an iid sequence of points in \(M\), distributed according to \(m\), and let \((t_i)_{i \in \mathbb{N}}\) be an iid sequence of numbers in \([0, 1]\), independent of the previous sequence and distributed according to the Beta distribution with parameters 1 and \(\beta\), i.e. \(\text{Prob}(t_i \in ds) = \beta(1-s)^{\beta-1} \cdot 1_{[0,1]}(s)ds\). Put
\[
\lambda_k = t_k \cdot \prod_{i=1}^{k-1} (1-t_i) \quad \text{and} \quad \nu = \sum_{k=1}^{\infty} \lambda_k \cdot \delta_{x_k}.
\]

Then \(\nu \in \mathcal{P}(M)\) is distributed according to \(Q^\beta \) \([Se94]\).

The distribution of \(\nu\) does not change if one replaces the above ‘stick-breaking process’ \((\lambda_k)_{k \in \mathbb{N}}\) by the ‘Dirichlet-Poisson process’ \((\lambda(k))_{k \in \mathbb{N}}\) obtained from it by ordering the entries of the previous one according to their size: \(\lambda(1) \geq \lambda(2) \geq \ldots \geq 0\). Alternatively, the Dirichlet-Poisson process can be regarded as the sequence of jumps of a Gamma process with parameter \(\beta\), ordered according to size.

Note that \(m(M_0) = 0\) for a given \(M_0 \subset M\) implies that \(\nu(M_0) = 0\) for \(Q^\beta\)-a.e. \(\nu \in \mathcal{P}(M)\). On the other hand, obviously, \(Q^\beta\)-a.e. \(\nu \in \mathcal{P}(M)\) is discrete. In contrast to that, as a corollary to Theorem 4.3 and in analogy to the 1-dimensional case we obtain:

**Corollary 6.2.** If \(M \subset \mathbb{R}^n\) then \(\mathbb{P}^\beta\)-a.e. \(\mu \in \mathcal{P}(M)\) has no absolutely continuous part and no atoms. The topological support of \(\mu^c\) is a \(m\)-zero set.

For \(\mathbb{P}^\beta\)-a.e. \(\mu \in \mathcal{P}(M)\) there exist a countable number of open convex sets \(U_k \subset M\) (‘holes in the support of \(\mu^c\)’) with sizes \(\lambda_k = m(U_k), k \in \mathbb{N}\). The measure \(\mu\) is supported on the complement of all these holes \(M \setminus \bigcup_k U_k\), a compact \(m\)-zero set. The sequence \((\lambda_k)_{k \in \mathbb{N}}\) of sizes of the holes is distributed according to the stick breaking process with parameter \(\beta\). In particular,
\[
\mathbb{E}\lambda_k = \frac{1}{\beta} \left( \frac{\beta}{1 + \beta} \right)^k \quad (\forall k \in \mathbb{N}).
\]

In average, each hole has size \(\leq \frac{1}{1+\beta}\). For large \(\beta\), the size of the \(k\)-th hole decays like \(\frac{1}{\beta} \exp(-k/\beta)\) as \(k \to \infty\). For small \(\beta\), \(\lambda(1)\) the size of the largest hole is of order \(\sim \frac{1}{1+0.73\beta}\) \([Gr88]\).

**Remark 6.3.** In principle, the reference measures in the conjugation map (see Remark 3.2) and in the Dirichlet-Ferguson process could be chosen different from each other.

Given a diffeomorphism \(h : M \to M\) the challenge for the sequel will be to deduce a change of variable formula for the entropic measure \(\mathbb{P}^\beta(d\mu)\) under the induced transformation
\[
\mu \mapsto h_* \mu
\]
of \(\mathcal{P}(M)\).
Conjecture 6.4. For each $\varphi^2$-diffeomorphism $h : M \to M$ there exists a function $Y_h^\beta : P \to \mathbb{R}$ such that
\[
\int U(h_*\mu)\mathbb{P}^\beta(d\mu) = \int U(\mu)Y_h^\beta(\mu)\mathbb{P}^\beta(d\mu),
\] (6.1)
for all bounded Borel functions $U : P \to \mathbb{R}$. (It suffices to consider $U$ of the form $U(\mu) = u(\mu(M_1), \ldots, \mu(M_N))$ for measurable partitions $M = \bigcup M_i$ and bounded measurable $u : \mathbb{R}^N \to \mathbb{R}$.) The density $Y_h^\beta$ is of the form
\[
Y_h^\beta(\mu) = \exp \left( \beta \int_M \log \det Dh(x)\mu(dx) \right) \cdot Y_h^0(\mu)
\] (6.2)
with $Y_h^0(\mu)$ being independent of $\beta$.

As an intermediate step, in order to derive a more direct representation for the entropic measure $\mathbb{P}^\beta$ on $\mathcal{P}(M)$, we may consider the measure $Q_\beta^\beta := (\chi^{-1})_*\mathbb{P}^\beta = (\mathcal{C}_G \circ \chi^{-1})_*\mathbb{Q}^\beta$
on $\mathcal{G}$.

It is the unique probability measure on the space $\mathcal{G}$ of monotone maps with the property that
\[
\int_{\mathcal{G}} u(m((g')^{-1}(M_1)), \ldots, m((g')^{-1}(M_N))) Q_\beta^\beta(dg) =
\frac{\Gamma(\beta)}{\prod_{i=1}^N \Gamma(\beta m(M_i))} \cdot \int_{[0,1]^N} u(x_1, \ldots, x_N) \cdot x_1^{\beta m(M_1)} \cdot \ldots \cdot x_N^{\beta m(M_N)} dx_1 \times
\delta_{(1 - \sum_{i=1}^N x_i)}(dx_N) dx_{N-1} \ldots dx_1
\]
for each measurable partition $M = \bigcup_{i=1}^N M_i$ and each bounded Borel function $u : \mathbb{R}^N \to \mathbb{R}$. Actually, one may assume without restriction that the partition consists of continuity sets of $m$ (i.e. $m(\partial M_i) = 0$ for all $i = 1, \ldots, N$) and that $u$ is continuous. Note that $(g')^{-1} = g$ almost everywhere whenever $g, m \ll m$.

Moreover, note that in dimension 1, say $M = [0, 1]$, the map $\mathcal{C}_G \circ \chi^{-1} : P \to \mathcal{G}$ assigns to each probability measure $\nu$ its cumulative distribution function $g$.

In dimension 1, the change of variable formula (6.1) allows to prove closability of the Dirichlet form $\mathcal{E}(u,u) = \int_P \|\nabla u\|^2(\mu) \, d\mathbb{P}^\beta(\mu)$ and to construct the Wasserstein diffusion $(\mu_t)_{t \geq 0}$, the reversible Markov process with continuous trajectories (and invariant distribution $\mathbb{P}^\beta$) associated to it [RS08].

The change of variable formula in dimension 1 can also be regarded as a ‘Girsanov type theorem’ for the (normalized) Gamma process [RYZ07]. Until now, no higher dimensional analogue is known.
The Wasserstein diffusion on 1-dimensional spaces satisfies a logarithmic Sobolev inequality [DS07]; it can be obtained as scaling limit of empirical distributions of interacting particle systems [AR07].

References

[AGS05] L. Ambrosio, N. Gigli and G. Savaré (2005): Gradient flows in metric spaces and in the space of probability measures. *Lectures in Mathematics ETH Zürich*. Birkhäuser Verlag, Basel.

[AR07] S. Andres and M. K. von Renesse (2007): Particle Approximation of the Wasserstein Diffusion. *Submitted*.

[Br87] Y. Brenier (1987): Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.* 305 no. 19, 805–808.

[CMS01] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger (2001): A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* 146 no. 2, 219–257.

[DS07] M. Döring and W. Stannat (2007): The logarithmic Sobolev inequality for the Wasserstein Diffusion. *Preprint*.

[Fe73] T. S. Ferguson (1973): A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1, 209–230.

[Gr88] R. C. Griffiths (1988): On the Distribution of Points in a Poisson Dirichlet Process. *Journal of Applied Probability* 25 no. 2, 336-345.

[LZ08] C. Lautensack and S. Zuyev (2008): Random Laguerre tessellations. *Adv. in Appl. Probab.* 40 no. 3, 630-650.

[Mc01] R.J. McCann (2001): Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.* 11 no. 3, 589–608.

[Oh08] S. Ohta (2008): Finsler interpolation inequalities. *Preprint*.

[Ot01] F. Otto (2001): The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* 26 no. 1-2, 101–174.

[Ro70] R.T. Rockafellar (1970): Convex analysis. *Princeton Mathematical Series* 28.

[RS08] M. K. von Renesse and K. T. Sturm (2008): Entropic measure and Wasserstein Diffusion. *To appear in Ann. Probab*.

[RYZ07] M. K. von Renesse, M. Yor and L. Zambotti (2007): Quasi-invariance properties of a class of subordinators. *To appear in Stoch Proc Appl*.

[Rü96] L. Rüschendorf (1996): On c-optimal random variables. *Statist. Probab. Lett.* 27 no. 3, 267–270.

[Se94] Sethuraman, Jayaram (1994): A constructive definition of Dirichlet priors. *Statist. Sinica* 4 no. 2, 639–650.

[Vi03] C. Villani (2003): Topics in Mass Transportation. *Graduate Studies in Mathematics, American Mathematical Society*.

[Vi08] C. Villani (2008): Optimal Transport - Old and new. *Grundlehren. Springer, Berlin*.
Karl-Theodor Sturm
Institut fuer Angewandte Mathematik
Poppelsdorfer Allee 82
D 53115 Bonn
Germany
e-mail: sturm@uni-bonn.de