Transportation–cost inequalities for diffusions driven by Gaussian processes

Sebastian Riedel*

Abstract

We prove transportation–cost inequalities for the law of SDE solutions driven by general Gaussian processes. Examples include the fractional Brownian motion, but also more general processes like bifractional Brownian motion. In case of multiplicative noise, our main tool is Lyons' rough paths theory. We also give a new proof of Talagrand's transportation–cost inequality on Gaussian Fréchet spaces. We finally show that establishing transportation–cost inequalities implies that there is an easy criterion for proving Gaussian tail estimates for functions defined on that space. This result can be seen as a further generalization of the "generalized Fernique theorem" on Gaussian spaces [FH14, Theorem 11.7] used in rough paths theory.

Introduction

Transportation–cost inequalities can be seen as a functional approach to the concentration of measure phenomenon (cf. Ledoux's work [Led01] for an introduction to the theory of measure concentration and the work [GL10] by Gozlan and Léonard for an overview to transport inequalities). They are usually of the following form: Let (E, d) be a metric space and let $P(E)$ denote the set of probability measures on the Borel sets of E. We say that a *p-transportation–cost inequality* holds for a probability measure $\mu \in P(E)$ if there is a constant C such that

$$\mathcal{W}_p(\nu, \mu) \leq \sqrt{CH(\nu \,|\, \mu)} \tag{0.1}$$

holds for all $\nu \in P(E)$. Here $\mathcal{W}_p(\nu, \mu)$ denotes the Wasserstein p-distance

$$\mathcal{W}_p(\nu, \mu) = \inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{E \times E} d(x, y)^p \, d\pi(x, y) \right)^{\frac{1}{p}}$$

---

*Mathematisches Institut, Universität zu Köln, Germany and Institut für Mathematik, Technische Universität Berlin, Germany. E-mail: riedel@math.tu-berlin.de

where $\Pi(\nu, \mu)$ is the set of all probability measures on the product space $E \times E$ with marginals ν resp. μ, and $H(\nu \,|\, \mu)$ is the relative entropy (or Kullback–Leibler divergence) of ν with respect to μ, i.e.

$$H(\nu \,|\, \mu) = \begin{cases} \int \log \left(\frac{d\nu}{d\mu} \right) \, d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases}$$

If (0.1) holds, we will say that $T_p(C)$ holds for the measure μ.

Inequalities of type (0.1) were first considered by Marton (cf. [Mar86], [Mar96]). The cases "$p = 1$" and "$p = 2$" are of special interest: The 1-transportation–cost inequality, i.e. the weakest form of (0.1), is actually equivalent to Gaussian concentration as it was shown by Djellout, Guillin and Wu in [DGW04] (using preliminary results by Bobkov and Götze obtained in [BG99]). The 2-transportation–cost inequality was first proven by Talagrand for the Gaussian measure on $\mathbb{R}^d$ in [Tal96] with the sharp constant $C = 2$ (for this reason it is also called *Talagrand's transportation–cost inequality*). $T_2(C)$ is particularly interesting since it has the *dimension–free tensorization property*: If $T_2(C)$ holds for two probability measures μ_1 and μ_2, it also holds for the product measure $\mu_1 \otimes \mu_2$ *for the same constant C* (see also [GL07] for a general account on tensorization properties for transportation–cost inequalities). One says that a probability measure μ has the *dimension–free concentration property* if measure concentration holds for μ and all product measures μ^n with the same parameters (a precise definition can be found e.g. in [Goz09]). It follows that the 2-transportation–cost inequality implies the dimension–free Gaussian concentration property for the underlying measure. Gozlan realized in [Goz09] that also the converse is true: If μ possesses the dimension–free Gaussian concentration property, $T_2(C)$ holds for μ. We also remark that the 2–transportation-cost inequality gained much attention because it is intimately linked to other famous concentration inequalities, notably to the logarithmic Sobolev inequality: In their celebrated paper [OV00], Otto and Villani showed that in a smooth Riemannian setting, the logarithmic Sobolev inequality implies the 2-transportation–cost inequality. Since then, this result has been generalized in several directions, see e.g. the recent work of Gigli and Ledoux [GL13] and the references therein.

In this work, we will mainly study transportation–cost inequalities for the law of a continuous diffusion Y induced by a stochastic differential equation (SDE) driven by a general Gaussian process, i.e. $Y \colon [0, T] \to \mathbb{R}^d$ solves

$$dY_t = b(Y_t) \, dt + \sum_{i=1}^m \sigma_i(Y_t) \circ dX_t^i; \qquad Y_0 = \xi \in \mathbb{R}^d, \quad t \in [0, T] \tag{0.2}$$

where $X = (X^1, \ldots, X^m) \colon [0, T] \to \mathbb{R}^m$ is a continuous Gaussian process and $b, \sigma_1, \ldots, \sigma_m$ are vectorfields in $\mathbb{R}^d$. Of course, the equation (0.2) needs an interpretation in a non-martingale setting in which Itō's theory is not applicable. However, in the case when X is a Brownian motion, the equation (0.2) can be solved using Itō's framework, and transportation–cost inequalities were studied in many works: In this context, $T_1(C)$ was first established for the law of Y with respect to the uniform metric by Djellout, Guillin and Wu in [DGW04]. In the same work, also $T_2(C)$ was proven, but for the weaker L^2-metric only. Under stronger assumptions on the equation (which guarantee in particular the existence of a unique invariant probability measure and exponential convergence towards it), Wu and Zhang proved in [WZ04] that also $T_2(C)$ holds for the uniform metric. Üstünel finally proved $T_2(C)$ for the uniform metric in [Üst12] in the most general form. However, replacing the Brownian motion by another Gaussian process, not much is known. To the authors knowledge, the only process which was

studied, up to a certain extend, is the *fractional Brownian motion* (fBm). By definition, a fBm with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process with covariance

$$R(s,t) = \frac{1}{2} \left(|t|^{2H} + |s|^{2H} - |t - s|^{2H} \right),$$

and it is easily seen that we obtain the usual Brownian motion for $H = 1/2$. However, for $H \neq 1/2$ this process is neither a semimartingale nor a Markov process. Guendouzi shows $T_1(C)$ for the L^1-metric for a mixed SDE involving a fBm with Hurst parameter $H > 1/2$ in [Gue12]. Saussereau studies more general equations in [Sau12] and shows $T_1(C)$ and $T_2(C)$ also for the uniform metric in particular situations. However, all equations he considers are either driven by a fBm with Hurst parameter $H > 1/2$, have additive noise or are one-dimensional. In fact, all these examples have something in common. Namely, it is known that in these cases, the solution to (0.2) is a continuous function of the driving process *path-by-path*. This is not true in the general case of (0.2) (and already fails, for instance, for the usual Brownian motion). For studying the equation (0.2) in full generality, one needs further ingredients, and we will use Lyon's *rough paths theory* to achieve this goal. Let us mention that our results imply those obtained in [Sau12] in case of fBm.

There is a further challenge when studying transportation–cost inequalities for solutions to (0.2) for general Gaussian processes X. The standard tool to establish transportation-cost inequalities, following [FÜ04] and [DGW04], is to use the *Girsanov transformation*. In a non-martingale framework, this argument completely breaks down. In case of the fBm, it can still be applied up to a certain point due to the Mandelbrot–van Ness representation of the fBm as a stochastic integral with respect to standard Brownian motion [MVN68]. However, there are many Gaussian processes (and we will encounter a class of them in the forthcoming Example 2.7) where such a representation is simply not known. Our approach can be seen as an attempt to prove concentration inequalities for diffusions *avoiding* the Girsanov transformation.

Let us explain our strategy and the contribution of this work. In Section 1, we consider transportation–cost inequalities on infinite dimensional Gaussian spaces. In turns out that in this framework, the quadratic transport inequality even holds for the *Cameron–Martin metric*, which is defined as follows: If $\mathcal{H}$ denotes the Cameron–Martin space associated to a Gaussian measure γ, set

$$d_{\mathcal{H}}(x, y) = \begin{cases} |x - y|_{\mathcal{H}} & \text{if } x - y \in \mathcal{H} \\ +\infty & \text{otherwise.} \end{cases} \tag{0.3}$$

The fact that a transport inequality holds for γ and this metric should be surprising at first sight since it is known that for infinite dimensional spaces, the Hilbert space $\mathcal{H}$ has γ-measure 0; in other words, $d_{\mathcal{H}}(x, y) = \infty$ "very often". In this form, the quadratic transport inequality was first proven by Feyel and Üstünel on Gaussian Banach spaces in [FÜ04, Theorem 3.1] using the Girsanov transformation (cf. also Gentil's PhD thesis [Gen01]). The proof we give does not rely on the Girsanov transformation and holds even in Fréchet spaces (cf. Theorem 1.2). Our main tool for proving transport inequalities for solutions to (0.2) will be a *contraction principle*, first proven[1] by Djellout, Guillin and Wu in [DGW04, Lemma 2.1] (we state a slightly more general version in the appendix, cf. Lemma 4.1), which states that transport inequalities are stable under Lipschitz maps. Together with our result about transport inequalities on Gaussian spaces, all we need to establish is Lipschitzness of the solution map $X(\omega) \mapsto Y(\omega)$ for equation

---

[1]In the context of measure concentration, this contraction principle already appeared earlier in a work by Maurey for of infimal-convolution inequalities, see [Mau91, Lemma 2].

(0.2). This is usually true for additive noise, and we study this case in Section 2.1 first. Interestingly, due to the strong form of the Gaussian transportation–cost inequality, we obtain such inequalities for the law of Y for metrics which are much larger than the uniform metric (cf. Theorem 2.2 and the discussion in Example 2.7) in the case of b in (0.2) being Lipschitz continuous. We further study the case where b only satisfies a one-sided Lipschitz condition in Theorem 2.6. We proceed with the multiplicative noise case in Section 2.2. As already mentioned, here we cannot expect the solution map to be Lipschitz continuous anymore in the usual topologies. The key idea is to use the *rough path factorization*: Instead of studying the map $X(\omega) \mapsto Y(\omega)$ directly, we consider an intermediate step; namely, we decompose this map as

$$X(\omega) \overset{S}{\mapsto} \mathbf{X}(\omega) \overset{\mathbf{I}}{\mapsto} Y(\omega). \tag{0.4}$$

The map S is called *lift map*, and it takes a Gaussian trajectory and maps it to a rough path. It is not continuous, but easy to analyze. The map $\mathbf{I}$ is called *Itô-Lyons map*, and it is known to be continuous, and even locally Lipschitz continuous in rough paths topology (in fact, this result can be seen as the main theorem in rough paths theory). The point now is that S can be shown to be locally Lipschitz continuous *from $\mathcal{H}$ to a rough paths space*, hence the decomposition, seen as a map from $\mathcal{H}$ to the space of continuous paths, *is* locally Lipschitz continuous. The contraction principle allows us to conclude $T_{2-\varepsilon}(C)$ for any $\varepsilon > 0$, cf. Theorem 2.14 (the ε-correction stems from the fact the we only have *local* Lipschitzness). Finally, we discuss the link between $T_p(C)$ and tail estimates for functions in Section 3 and establish a link between $T_p(C)$ and the *generalized Fernique theorem* (cf. [DOR15, Theorem 17], [FH14, Theorem 11.7] and [FO10]) which is of fundamental importance in rough paths theory (cf. [FH14, Chapter 11]). This section does not depend on the former ones and may be of independent interest.

Let us finally mention that we think that our approach can be carried over to SDEs in infinite dimensions, i.e. to *stochastic partial differential equations*, in particular to those considered in Hairer's theory of *regularity structures* [Hai14] or Gubinelli-Imkeller-Perkowski's approach using *paracontrolled distributions* [GIP15]. Indeed, in both theories, it was understood that (after a possible renormalization), singular equations like the KPZ-equation (cf. also [Hai13]) often have a similar factorization as in (0.4), and this was the basic ingredient we needed for ordinary SDEs as well.

Notation

If $(X, \mathcal{F})$ is a measurable space, $P(X)$ denotes the set of all probability measures defined on $\mathcal{F}$. If X is a topological space, $\mathcal{F}$ will be usually be the Borel σ-algebra $\mathcal{B}(X)$. If X and Y are measurable spaces and $\nu \in P(X)$, $\mu \in P(Y)$, then $\Pi(\nu, \mu)$ denotes the set of all product measures on $X \times Y$ with marginals ν resp. μ. If $[S, T]$ is any interval in $\mathbb{R}$, we write $\mathcal{P}([S, T])$ for the set of all finite partitions of $[S, T]$ of the form $S = t_0 < t_1 < \ldots < t_M = T$, $M \in \mathbb{N}$. If $x, y \colon [S, T] \to (B, \|\cdot\|)$ are paths with values in a normed space and $p \geq 1$, we define p-variation seminorm and pseudometric as

$$\|x\|_{p-\mathrm{var};[S,T]} := \sup_{D \in \mathcal{P}([S,T])} \left(\sum_{t_i \in D} \|x_{t_{i+1}} - x_{t_i}\|^p \right)^{\frac{1}{p}}; \tag{0.5}$$

$$d_{p-\mathrm{var};[S,T]}(x, y) := \|x - y\|_{p-\mathrm{var};[S,T]}. \tag{0.6}$$

If the time horizon is clear from the context, we sometimes omit the subindex $[S, T]$ in the notation. The set of all continuous paths $x \colon [S, T] \to B$ with $\|x\|_{p-\mathrm{var};[S,T]} < \infty$ is denoted by $C^{p-\mathrm{var}}([S, T]; B)$ and we also define $C^{p-\mathrm{var}}_\xi([S, T]; B) := \{x \in C^{p-\mathrm{var}}([S, T]; B) : x_S = \xi\}$ for some $\xi \in B$. If B is a Banach space, $C^{p-\mathrm{var}}_0([S, T]; B)$ is also a Banach space with the norm $\|\cdot\|_{p-\mathrm{var}}$.

1   Transportation inequality on a Gaussian space

In this section, we give a proof of $T_2(2)$ on Gaussian spaces for the Cameron–Martin metric defined in (0.3), a result which was first proven on Banach spaces by Feyel and Üstünel [FÜ04, Theorem 3.1] using the Girsanov transformation. Our strategy will be to "approximate" the infinite dimensional space by finite dimensional ones on which we know from Talagrand's original result that $T_2(2)$ holds.

We start with an abstract approximation result.

Lemma 1.1. *Let X and Y be Polish spaces and let (μ_n) and (ν_n) be sequences of probability measures on X resp. Y which converge weakly to some probability measures μ resp. ν. Let $c_n \colon X \times Y \to [0, \infty)$ be a nondreasing sequence of bounded, continuous functions such that $c_n \nearrow c$ pointwise where $c \colon X \times Y \to [0, \infty]$. Then, along a subsequence,*

$$\liminf_{k \to \infty} \inf_{\pi \in \Pi(\mu_{n_k}, \nu_{n_k})} \int_{X \times Y} c_{n_k}(x,y)\, d\pi(x,y) \geq \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x,y)\, d\pi(x,y).$$

Proof. For $\pi \in P(X \times Y)$, set

$$I_n(\pi) := \int_{X \times Y} c_n(x,y)\, d\pi(x,y) \quad \text{and} \quad I(\pi) = \int_{X \times Y} c(x,y)\, d\pi(x,y).$$

From continuity of the c_n, we know from [Vil03, Theorem 1.3] that there are measures $\pi_n \in \Pi(\mu_n, \nu_n)$ such that

$$I_n(\pi_n) = \inf_{\pi \in \Pi(\mu_n, \nu_n)} I_n(\pi)$$

for all $n \geq 1$. We claim that the sequence (π_n) is tight in $P(X \times Y)$. Indeed: Let $\varepsilon > 0$ be given. From Prokhorov's theorem, the sequences (μ_n) and (ν_n) are tight, therefore we can find compact sets $K_1 \subset X$ and $K_2 \subset Y$ such that

$$\mu_n(K_1) \geq 1 - \varepsilon/2 \quad \text{and} \quad \nu_n(K_2) \geq 1 - \varepsilon/2$$

for all $n \geq 1$. This implies that

$$\pi_n(K_1 \times K_2) \geq 1 - \varepsilon$$

for all $n \geq 1$ which shows tightness. Using agains Prokhorov's theorem, we know that there is a subsequence for which $\pi_{n_k} \to \pi^*$ weakly in $P(X \times Y)$ for $k \to \infty$. Let $f \colon X \to \mathbb{R}$ be a bounded, continuous function. From weak convergence,

$$\lim_{k \to \infty} \int_X f(x)\, d\mu_{n_k}(x) = \int_X f(x)\, d\mu(x).$$

Furthermore,

$$\lim_{k \to \infty} \int_{X \times Y} f(x)\, d\pi_{n_k}(x,y) = \int_{X \times Y} f(x)\, d\pi^*(x,y) = \int_X f(x)\, d\pi^*(x, Y)$$

which implies that $\pi^*(\cdot, Y) = \mu$. Similarly, $\pi^*(X, \cdot) = \nu$, and we have thus shown that $\pi^* \in \Pi(\mu, \nu)$. From monotonicity, whenever $n \geq m$, $I_n(\pi_n) \geq I_m(\pi_n)$, and therefore

$$\liminf_{k \to \infty} I_{n_k}(\pi_{n_k}) \geq \liminf_{k \to \infty} I_m(\pi_{n_k}) \geq I_m(\pi_*)$$

for every $m \geq 1$. Monotone convergence gives

$$\lim_{m \to \infty} I_m(\pi_*) = I(\pi_*),$$

and thus

$$\liminf_{k \to \infty} I_{n_k}(\pi_{n_k}) \geq \lim_{m \to \infty} I_m(\pi_*) = I(\pi_*) \geq \inf_{\pi \in \Pi(\mu, \nu)} I(\pi). \qquad \square$$

In the following, we aim to consider Gaussian measures on linear spaces. Typically, one assumes that the space should be *locally convex*, i.e. its topology is generated by family of seminorms separating points (cf. [Bog98, Chapter 2 and Appendix A]). It will be convenient for us to assume that the space is also *Polish*, i.e. separable and completely metrizable. Such spaces are also called separable *Fréchet spaces*. A *Gaussian Fréchet spaces* is a triplet $(F, \mathcal{H}, \gamma)$ where F is a separable Fréchet spaces, γ is a Gaussian measure on the Borel σ-field $\mathcal{B}(F)$ and $\mathcal{H}$ denotes the *Cameron-Martin space* which is a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ lying in F (cf. [Bog98, Section 2.4] for the precise definition of the Cameron Martin space and [Bog98, Chapter 2 and 3] for further properties). The induced norm on $\mathcal{H}$ will be denoted by $|\cdot|_{\mathcal{H}}$. Recall the definition of the Cameron–Martin metric given in (0.3).

The following theorem is the main result from this section.

Theorem 1.2. *Let* $(F, \mathcal{H}, \gamma)$ *be a Gaussian Fréchet space. Then for any* $\nu \in P(F)$,

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{F \times F} d_{\mathcal{H}}(x, y)^2 \, d\pi(x, y) \leq 2 \, H(\nu \,|\, \gamma).$$

Proof. Note that for every $h, k \in \mathcal{H}$, there are elements $\hat{h}, \hat{k} \in F^*$ such that

$$\langle h, k \rangle_{\mathcal{H}} = \langle \hat{h}, \hat{k} \rangle_{L^2(\gamma)} = \int_F \hat{h}(x) \hat{k}(x) \, d\gamma(x),$$

cf. [Bog98, Section 2.4 and 3.2.3 Theorem]. Let (e_n) be an orthonormal basis of $\mathcal{H}$. Define

$$H_n := \operatorname{span}\{e_1, \ldots, e_n\}$$

and $p_n \colon F \to H_n$ by

$$p_n(x) = \sum_{k=1}^{n} \hat{e}_k(x) e_k.$$

Note that for $h \in \mathcal{H}$, $\hat{e}_k(h) = \langle e_k, h \rangle_{\mathcal{H}}$ by [Bog98, 2.10.5 Lemma], thus

$$p_n(h) = \sum_{k=1}^{n} \langle e_k, h \rangle_{\mathcal{H}} e_k.$$

We equip the space H_n with the scalar product

$$\langle v, w \rangle_{H_n} = \sum_{k=1}^{n} \lambda_k \mu_k \quad \text{if} \quad v = \sum_{k=1}^{n} \lambda_k e_k, \ w = \sum_{k=1}^{n} \mu_k e_k.$$

Note that with this definition, $\langle p_n(h), p_n(k) \rangle_{H_n} = \langle p_n(h), p_n(k) \rangle_{\mathcal{H}}$. Consider the image measure $\tilde{\gamma}_n := \gamma \circ p_n^{-1}$. Then $(H_n, \tilde{\gamma}_n)$ is a finite dimensional Gaussian space, and we know from Talagrand's result that $T_2(2)$ holds here. Consider the inclusion maps $\iota_n \colon H_n \hookrightarrow F$ and set $\gamma_n := \tilde{\gamma}_n \circ \iota_n^{-1}$. By the contraction principle in Lemma 4.1, we see that for every $\nu \in P(F)$

$$\inf_{\pi \in \Pi(\nu, \gamma_n)} \int_{F \times F} \tilde{d}_n(x, y)^2 \, d\pi(x, y) \leq 2 \, H(\nu \,|\, \gamma_n)$$

holds for all $n \geq 1$ where

$$\tilde{d}_n(x, y) = |p_n(x) - p_n(y)|_{\mathcal{H}}.$$

Set $d_n(x, y) := \tilde{d}_n(x, y) \wedge n$. Since $d_n \leq \tilde{d}_n$, also

$$\inf_{\pi \in \Pi(\nu, \gamma_n)} \int_{F \times F} d_n(x, y)^2 \, d\pi(x, y) \leq 2 \, H(\nu \,|\, \gamma_n) \tag{1.1}$$

holds for every $\nu \in P(F)$ and $n \geq 1$. We collect some facts about the functions d_n. First, it is clear by definition that all $d_n \colon F \times F \to [0, \infty)$ are bounded and continuous. Furthermore, for fixed $x, y \in F$,

$$\tilde{d}_n(x, y)^2 = \sum_{k=1}^{n} |\hat{e}_k(x - y)|^2 \leq \sum_{k=1}^{n+1} |\hat{e}_k(x - y)|^2 = \tilde{d}_{n+1}(x, y)^2$$

which shows that the sequence (d_n) is nondecreasing. We claim that $d_n \nearrow d_{\mathcal{H}}$ pointwise for $n \to \infty$. Indeed, if $x - y \in \mathcal{H}$,

$$\lim_{n \to \infty} d_n(x, y)^2 = \sum_{k=1}^{\infty} |\langle e_k, x - y \rangle_{\mathcal{H}}|^2 = |x - y|_{\mathcal{H}}^2 = d_{\mathcal{H}}(x, y)^2$$

by Parseval's identity. Conversely, if $\lim_{n \to \infty} d_n(x, y) < \infty$ for some $x, y \in F$, we may define

$$\sum_{k=1}^{\infty} \hat{e}_k(x - y) e_k =: z \in \mathcal{H}.$$

This implies that

$$\sum_{k=1}^{\infty} \hat{e}_k(x - y) e_k = \sum_{k=1}^{\infty} \hat{e}_k(z) e_k$$

and applying $\hat{e}_k$ on both sides shows that $\hat{e}_k(x - y) = \hat{e}_k(z)$ holds for every $k \in \mathbb{N}$. Hence $x - y = z \in \mathcal{H}$ and we have shown the claim. Next, we show that $\gamma_n \to \gamma$ weakly for $n \to \infty$. Let $g \colon F \to \mathbb{R}$ be a bounded, continuous function. Then

$$\int_F g(x) \, d\gamma_n(x) = \int_F g\left(\sum_{k=1}^{n} \hat{e}_k(x) e_k \right) d\gamma(x).$$

We know from [Bog98, 3.5.1 Theorem] that $\sum_{k=1}^{\infty} \hat{e}_k(x) e_k = x$ γ-almost surely, hence by dominated convergence,

$$\int_F g(x) \, d\gamma_n(x) \to \int_F g(x) \, d\gamma(x)$$

for $n \to \infty$ which shows weak convergence. Choose any $\nu \in P(F)$ with $\nu \ll \gamma$. Set $f := \frac{d\nu}{d\gamma}$ and define $d\nu_n := f \, d\gamma_n$. From (1.1), we have

$$\inf_{\pi \in \Pi(\nu_n, \gamma_n)} \int_{F \times F} d_n(x, y)^2 \, d\pi(x, y) \leq 2 \int_F f \log f \, d\gamma_n$$

for every $n \in \mathbb{N}$. Assume first that f is bounded and continuous. In this case, we have $\nu_n \to \nu$ weakly for $n \to \infty$ and we can use Lemma 1.1 for the left hand side and weak convergence for the right hand side of the above inequality to conclude that indeed

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{F \times F} d_{\mathcal{H}}(x, y)^2 \, d\pi(x, y) \leq 2 \int_F f \log f \, d\gamma$$

holds for every $\nu \in P(F)$ with bounded, continuous density. Next, we extend this result to arbitrary density functions. By a result of Wiśniewski [Wiś94, Theorem 1], for every measurable map $f\colon F \to \mathbb{R}$ there exists a sequence of continuous functions (f_n) such that $f_n \to f$ γ-almost surely. Assume first that the density f is bounded by some $C > 0$. Let (f_n) be a sequence of continuous functions converging γ-a.s. to f. We may assume w.l.o.g. that $0 \le f_n \le C$ for all f_n, otherwise we replace each f_n by $(f_n \wedge C) \vee 0$. Set $\alpha_n := \|f_n\|_{L^1(\gamma)}$ and $d\nu_n := (f_n/\alpha_n) \, d\gamma$. We have shown that for every $n \in \mathbb{N}$,

$$\inf_{\pi \in \Pi(\nu_n, \gamma)} \int_{F \times F} d_{\mathcal{H}}(x,y)^2 \, d\pi(x,y) \le 2 \int_F (f_n/\alpha_n) \log(f_n/\alpha_n) \, d\gamma.$$

The above inequality implies that also

$$\inf_{\pi \in \Pi(\nu_n, \gamma)} \int_{F \times F} d_m(x,y)^2 \, d\pi(x,y) \le 2 \int_F (f_n/\alpha_n) \log(f_n/\alpha_n) \, d\gamma$$

holds for every fixed $n, m \in \mathbb{N}$. From Lebesgue's dominated convergence theorem, we can conclude that $\nu_n \to \nu$ weakly and

$$\int_F (f_n/\alpha_n) \log(f_n/\alpha_n) \, d\gamma \to \int_F f \log f \, d\gamma$$

for $n \to \infty$. We can use Lemma 1.1 again (now for the fixed cost function d_m) to see that

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{F \times F} d_m(x,y)^2 \, d\pi(x,y) \le 2 \int_F f \log f \, d\gamma \tag{1.2}$$

holds for every $m \in \mathbb{N}$ and every bounded density f. Now let f be an arbitrary density function. Set $f_n := f \wedge n$, $\alpha_n := \|f_n\|_{L^1(\gamma)}$ and $d\nu_n := (f_n/\alpha_n) \, d\gamma$. Using monotone convergence, we see that $\nu_n \to \nu$ weakly and $H(\nu_n \,|\, \gamma) \to H(\nu \,|\, \gamma)$ for $n \to \infty$. As before, Lemma 1.1 shows that (1.2) holds for every $\nu \ll \gamma$ with density function f and every $m \in \mathbb{N}$. Taking the limes inferior along a subsequence of m in (1.2), we can use Lemma 1.1 a fourth time to conclude the assertion of our theorem. $\qquad\square$

1.1 Banach spaces

Let $(B, \|\cdot\|)$ be a separable Banach space and set $d_B(x,y) := \|x - y\|$. As an immediate corollary of Theorem 1.2 we obtain:

Corollary 1.3. *Let $(B, \mathcal{H}, \gamma)$ be a Gaussian Banach space. Then for any $\nu \in P(B)$,*

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{B \times B} d_B(x,y)^2 \, d\pi(x,y) \le 2\sigma^2 \, H(\nu \,|\, \gamma)$$

where

$$\sigma^2 = \sup_{l \in B^*, \|l\| \le 1} \int l(x)^2 \, d\gamma(x) < \infty. \tag{1.3}$$

Proof. It is well known that $\sigma < \infty$ and that for every $h \in \mathcal{H}$ one has $\|h\| \le \sigma |h|_{\mathcal{H}}$, cf. [Led96, Chapter 4], which gives the claim. $\qquad\square$

1.2 Rough paths spaces

In the case of $B = C_0([0,T], \mathbb{R}^d)$, Theorem 1.2 immediately generalizes to rough paths spaces. Let γ be a Gaussian measure on B with corresponding Cameron–Martin space $\mathcal{H}$. For the sake of simplicity, we will assume that $\mathcal{H}$ is continuously embedded in C_0, otherwise we could have used a smaller space lying in C_0 instead. Let $\mathcal{D}$ be a

rough paths space (which could either be geometric or non-geometric, a p-variation or an α-Hölder rough paths space, cf. [LCL07], [FV10b] or [FH14] for a precise definition) and assume that there is a measurable map $S\colon C_0 \to \mathcal{D}$ such that $\pi_1 \circ S = \mathrm{Id}_{C_0}$ holds where $\pi_1\colon \mathcal{D} \to C_0$ is the projection map. The map S is called a *lift map*. Set $\boldsymbol{\gamma} = \gamma \circ S^{-1}$. Abusing notation, we define $d_{\mathcal{H}}\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R} \cup \{+\infty\}$ as

$$d_{\mathcal{H}}(\mathbf{x}, \mathbf{y}) = \begin{cases} |\pi_1(\mathbf{x}) - \pi_1(\mathbf{y})|_{\mathcal{H}} & \text{if } \pi_1(\mathbf{x}) - \pi_1(\mathbf{y}) \in \mathcal{H} \\ +\infty & \text{otherwise.} \end{cases}$$

Corollary 1.4. *For any* $\boldsymbol{\nu} \in P(\mathcal{D})$,

$$\inf_{\boldsymbol{\pi} \in \Pi(\boldsymbol{\nu}, \boldsymbol{\gamma})} \int_{\mathcal{D} \times \mathcal{D}} d_{\mathcal{H}}(\mathbf{x}, \mathbf{y})^2 \, d\boldsymbol{\pi}(\mathbf{x}, \mathbf{y}) \le 2\, H(\boldsymbol{\nu} \,|\, \boldsymbol{\gamma}).$$

Proof. By definition, $d_{\mathcal{H}}(S(x), S(y)) = d_{\mathcal{H}}(x, y)$, hence S is (in particular) 1-Lipschitz and the result follows from Theorem 1.2 and Lemma 4.1. $\qquad\square$

2 Applications to diffusions

2.1 SDEs with additive noise

In this section, we will consider SDEs of the form

$$Y_t = \xi + \int_0^t b(Y_s)\, ds + \sum_{i=1}^m \int_0^t \sigma_i(s)\, dX_s^i \qquad (2.1)$$

with $\xi \in \mathbb{R}^d$. Here, $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is a continuous vector field, $\sigma_1, \ldots, \sigma_m\colon [0, T] \to \mathbb{R}^d$ are continuous functions and $X = (X^1, \ldots, X^m)\colon [0, T] \to \mathbb{R}^m$ is a Gaussian process with continuous trajectories. The stochastic integrals in (2.1) can either be defined pathwise (e.g. as Young integrals, cf. [You36] or [FV10b, Section 6]) or by probabilistic means (e.g. as Wiener integrals). At this stage, we only assume that the stochastic integrals are defined in such a way that they introduce a bounded linear map from the Gaussian space $C([0, T], \mathbb{R}^m)$ to the space $C([0, T], \mathbb{R}^d)$ which implies that the sum of the integrals is again a Gaussian processes. Therefore, there is no loss of generality to consider SDEs of the form

$$Y_t = \xi + \int_0^t b(Y_s)\, ds + X_t \qquad (2.2)$$

instead of (2.1) where $X\colon [0, T] \to \mathbb{R}^d$ is an $\mathbb{R}^d$-valued Gaussian process starting at 0 with continuous sample paths. Under mild regularity assumptions on b (e.g. continuous, locally Lipschitz continuous and linear growth), the equation (2.2) can be solved pathwise for every continuous trajectory of the Gaussian process. We aim to establish concentration inequalities for the law of the solution $Y\colon [0, T] \to \mathbb{R}^d$. Our strategy will be to show that the solution map $X(\omega) \mapsto Y(\omega)$ is Lipschitz continuous, which implies the concentration inequality by the contraction principle stated in Lemma 4.1.

In the case of X being a Wiener process, Djellout, Guillin and Wu show in [DGW04, Proposition 5.4] that the quadratic transportation inequality holds even for the metric $d_{\mathcal{H}}$. However, their analysis relies on the fact that in case of the Wiener process, the Cameron Martin space is explicitly known; it is the Sobolev space $H_0^1 = W_0^{1,2}$. For a general Gaussian process, the Cameron Martin space is usually only implicitly defined, and showing Lipschitz continuity for the corresponding metric is not obvious. On the other hand, there are often continuous embeddings available for the Cameron Martin

space into the space of paths with finite p-variation. Showing Lipschitz continuity for the p-variation metric is a much easier task which will immediately yield concentration inequalities in p-variation topology.

We start with a simple calculation.

Proposition 2.1. *Let $x_1, x_2 \colon [0,T] \to \mathbb{R}^d$ be two continuous paths and choose $\xi^1, \xi^2 \in \mathbb{R}^d$. Consider the equations*

$$y_t^i = \xi^i + \int_0^t b(y_s^i)\,ds + x_t^i; \quad t \in [0,T],\ i = 1,2 \tag{2.3}$$

where $b \colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous with Lipschitz constant L.

Then the equations (2.3) have unique, continuous solutions $y^1, y^2 \colon [0,T] \to \mathbb{R}^d$ and the estimate

$$\|y^1 - y^2\|_{q-var} \le 2^{1-\frac{1}{q}} \exp\left(2^{1-\frac{1}{q}} LT\right) \left(LT|\xi^1 - \xi^2| + \|x^1 - x^2\|_{q-var}\right) \tag{2.4}$$

holds for every $q \in [1, \infty)$.

Proof. Existence and uniqueness is classical, we only need to prove the estimate (2.4). Fix some $t \in [0,T]$. Let (t_i) be a partition of $[0,t]$. Using the equations shows that

$$|y_{t_{i+1}}^1 - y_{t_i}^1 - y_{t_{i+1}}^2 + y_{t_i}^2| \le L \int_{t_i}^{t_{i+1}} |y_s^1 - y_s^2|\,ds + |x_{t_{i+1}}^1 - x_{t_i}^1 - x_{t_{i+1}}^2 + x_{t_i}^2|$$

for every $t_i < t_{i+1}$. Taking both sides to the power q and summing over all increments gives

$$\sum_{t_i} |y_{t_{i+1}}^1 - y_{t_i}^1 - y_{t_{i+1}}^2 + y_{t_i}^2|^q \le 2^{q-1} L^q \left(\int_0^t |y_s^1 - y_s^2|\,ds\right)^q$$
$$+ 2^{q-1} \sum_{t_i} |x_{t_{i+1}}^1 - x_{t_i}^1 - x_{t_{i+1}}^2 + x_{t_i}^2|^q.$$

Taking now the supremum over all partitions implies

$$\|y^1 - y^2\|_{q-\mathrm{var};[0,t]} \le 2^{1-\frac{1}{q}} L \int_0^t |y_s^1 - y_s^2|\,ds + 2^{1-\frac{1}{q}} \|x^1 - x^2\|_{q-\mathrm{var};[0,t]}.$$

The integral can be estimated by

$$\int_0^t |y_s^1 - y_s^2|\,ds \le \int_0^t \|y^1 - y^2\|_{q-\mathrm{var};[0,s]}\,ds + t|\xi^1 - \xi^2|.$$

Gronwall's inequality implies the claim. $\qquad\qquad\qquad\qquad\qquad\square$

Set $C_\xi = C_\xi([0,T]; \mathbb{R}^d)$.

Theorem 2.2. *Assume that $b \colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous with Lipschitz constant $L > 0$. Let $X \colon [0,T] \to \mathbb{R}^d$ be a continuous Gaussian process with Cameron Martin space $\mathcal{H}$, and assume that there is a continuous embedding*

$$\iota \colon \mathcal{H} \hookrightarrow C^{q-var} \tag{2.5}$$

for some $q \in [1, \infty)$. Let Y be the solution to the SDE (2.2) and let μ be the law of Y.

Then for every $\nu \in P(C_\xi)$,

$$\inf_{\pi \in \Pi(\nu,\mu)} \int_{C_\xi \times C_\xi} d_{q-var}(x,y)^2\,d\pi(x,y) \le 2^{3-\frac{2}{q}} \exp\left(2^{2-\frac{1}{q}} LT\right) \|\iota\|_{\mathcal{H} \hookrightarrow C^{q-var}}^2 H(\nu \,|\, \mu). \tag{2.6}$$

Proof. Follows from Theorem 1.2, the contraction principle in Lemma 4.1 and Proposition 2.1. □

Remark 2.3. Embeddings of the form (2.5) play a crucial role in Gaussian rough paths theory and we will revisit them also in the next section. Sufficient conditions for such embeddings, as well as many examples of Gaussian processes for which they hold, are given in [FGGR16].

Next, we aim to relax the assumptions on $b \colon \mathbb{R}^d \to \mathbb{R}^d$. In case of the Brownian motion, it is well known (cf. [PR07]) that (2.2) has a unique solution provided b is continuous and satisfies the following *one sided Lipschitz condition*:

(i) There exists a constant C_1 such that

$$\langle b(\xi) - b(\zeta), \xi - \zeta \rangle \leq C_1 |\xi - \zeta|^2 \quad \text{for every } \xi, \zeta \in \mathbb{R}^d. \tag{2.7}$$

However, one has to be careful when solving (2.2) pathwise: In [CHJ13, p. 43], the authors show that there are trajectories which lead to explosion in finite time of solutions to (2.2) although the vector field b satisfies (2.7). In [RS17] and [SS17], a further condition on b was introduced. Together with (2.7), this condition prevents explosion, even in the more general case of multiplicative noise. This condition takes the following form:

(ii) There exists a constant C_2 such that

$$\left| b(\xi) - b(\zeta) - \frac{\langle b(\xi) - b(\zeta), \xi - \zeta \rangle (\xi - \zeta)}{|\xi - \zeta|^2} \right| \leq C_2 |\xi - \zeta|$$

$$\text{for every } \xi, \zeta \in \mathbb{R}^d \text{ with } \xi - \zeta \neq 0. \tag{2.8}$$

In the following, we will assume both (2.7) and (2.8).

Let $x \colon [0, T] \to \mathbb{R}^d$ be continuous. It is shown in [RS17, Lemma 4.1 and Lemma 4.2] that for b continuous and satisfying (2.7) and (2.8), the equation

$$\dot{z}_t = b(z_t + x_t)$$

generates a continuous two-parameter flow.

Lemma 2.4. *Let $x^1, x^2 \colon [S, T] \to \mathbb{R}^d$ be continuous. Consider the solutions*

$$z^1, z^2 \colon [S, T] \to \mathbb{R}^d$$

to the equations

$$\dot{z}_t^i = b(z_t^i + x_t^i); \quad t \in [S, T]$$
$$z_S^i = \xi^i$$

for $i = 1, 2$ and two initial conditions $\xi^1, \xi^2 \in \mathbb{R}^d$. Assume that

$$|z_t^1 - z_t^2| \geq |x_t^1 - x_t^2| \quad \text{for all } t \in [S, T].$$

Then

$$\sup_{t \in [S,T]} |z_t^1 - z_t^2| \leq e^{\frac{3}{2}|T-S|(C_1+C_2)} \left(|\xi^1 - \xi^2| + \sqrt{(C_1 + C_2)|T - S|} \sup_{t \in [S,T]} |x_t^1 - x_t^2| \right).$$

Proof. For all $t \in [S, T]$,

$$|z_t^1 - z_t^2|^2 = |\xi^1 - \xi^2|^2 + 2\int_S^t \langle z_s^1 - z_s^2, b(z_s^1 + x_s^1) - b(z_s^2 + x_s^2)\rangle \, ds.$$

Fix $s \in [S, t]$. Choose $\alpha, \beta \in \mathbb{R}$ such that

$$b(z_s^1 + x_s^1) - b(z_s^2 + x_s^2) = \alpha(z_s^1 + x_s^1 - z_s^2 - x_s^2) + \beta v$$

where $v \perp (z_s^1 + x_s^1 - z_s^2 - x_s^2)$ in the case $z_s^1 + x_s^1 - z_s^2 - x_s^2 \neq 0$ and v arbitrary but $\beta = 0$ otherwise. By (2.7), $\alpha \leq C_1$ and (2.8) implies that

$$|\beta| \leq C_2|z_s^1 + x_s^1 - z_s^2 - x_s^2|.$$

Note that

$$\langle z_s^1 - z_s^2, z_s^1 + x_s^1 - z_s^2 - x_s^2\rangle \geq |z_s^1 - z_s^2|^2 - |z_s^1 - z_s^2||x_s^1 - x_s^2| \geq 0$$

by assumption, therefore

$$\alpha\langle z_s^1 - z_s^2, z_s^1 + x_s^1 - z_s^2 - x_s^2\rangle \leq \frac{C_1}{2}(3|z_s^1 - z_s^2|^2 + |x_s^1 - x_s^2|^2).$$

Furthermore,

$$\beta\langle z_s^1 - z_s^2, v\rangle \leq \frac{C_2}{2}(3|z_s^1 - z_s^2|^2 + |x_s^1 - x_s^2|^2).$$

This implies that

$$|z_t^1 - z_t^2|^2 \leq |\xi^1 - \xi^2|^2 + (C_1 + C_2)|T - S|\sup_{u \in [S,T]}|x_u^1 - x_u^2|^2 + 3(C_1 + C_2)\int_S^t |z_s^1 - z_s^2|^2 \, ds$$

holds for all $t \in [S, T]$. Gronwall's Lemma gives the claim. $\square$

Proposition 2.5. *Let $x_1, x_2 \colon [0, T] \to \mathbb{R}^d$ be two continuous paths starting at 0 and choose $\xi^1, \xi^2 \in \mathbb{R}^d$. Consider the equations*

$$y_t^i = \xi^i + \int_0^t b(y_s^i) \, ds + x_t^i; \quad t \in [0, T], \ i = 1, 2 \tag{2.9}$$

where $b \colon \mathbb{R}^d \to \mathbb{R}^d$ is continuous and satisfies (2.7) and (2.8).

Then the equations (2.9) have unique, continuous solutions $y^1, y^2 \colon [0, T] \to \mathbb{R}^d$ and the estimate

$$\|y^1 - y^2\|_\infty \leq 3e^{2T(C_1 + C_2)}(|\xi^1 - \xi^2| + \|x^1 - x^2\|_\infty) \tag{2.10}$$

holds.

Proof. The fact that the equations (2.9) possess unique solutions is a special case of [RS17, Theorem 4.3]. We only need to prove the estimate (2.10). It is easy to see that if $z^i \colon [0, T] \to \mathbb{R}^d$, $i = 1, 2$ denote the solutions to

$$\begin{aligned} \dot{z}_t^i &= b(z_t^i + x_t^i) \\ z_0^i &= \xi^i, \end{aligned} \tag{2.11}$$

the solutions y^i to (2.9) are given by $z^i + x^i$. Set

$$\beta_t := \left| |z_t^1 - z_t^2| - |x_t^1 - x_t^2| \right|.$$

Let $\delta > 0$. We define a sequence of increasing numbers $0 =: \tau_0 < \tau_1 < \ldots$ as follows:

$$\tau_1 := \inf_{t \geq \tau_0} \{\beta_t > \delta\} \wedge T$$

$$\tau_2 := \inf_{t \geq \tau_1} \{\beta_t < \delta/2\} \wedge T$$

$$\tau_3 := \inf_{t \geq \tau_2} \{\beta_t > \delta\} \wedge T$$

$$\vdots$$

Note that there is a minimal number $N \in \mathbb{N}$ such that $\tau_N = T$. Indeed, otherwise we constructed an increasing sequence (τ_n), bounded by T, which therefore converges towards some number τ, but (β_{τ_n}) can clearly not converge although it is continuous, which is a contradiction. By construction, for every $n = 0, \ldots, N-1$, one either has $\beta_t \leq \delta$ or $\beta_t \geq \delta/2$ for every $t \in [\tau_n, \tau_{n+1}]$. In the first case,

$$|z_t^1 - z_t^2| \leq \delta + \sup_{u \in [\tau_n, \tau_{n+1}]} |x_u^1 - x_u^2|$$

for every $t \in [\tau_n, \tau_{n+1}]$. In the case $\beta_t \geq \delta/2$, we either have $|x_t^1 - x_t^2| \geq \delta/2 + |z_t^1 - z_t^2| \geq |z_t^1 - z_t^2|$ which implies

$$|z_t^1 - z_t^2| \leq \sup_{u \in [\tau_n, \tau_{n+1}]} |x_u^1 - x_u^2|$$

for all $t \in [\tau_n, \tau_{n+1}]$, or $|x_t^1 - x_t^2| \leq |z_t^1 - z_t^2|$ for all $t \in [\tau_n, \tau_{n+1}]$. In the second case, we can use Lemma 2.4 to obtain the estimate

$$|z_t^1 - z_t^2| \leq e^{2(\tau_{n+1} - \tau_n)(C_1 + C_2)} \left(|z_{\tau_n}^1 - z_{\tau_n}^2| + \sup_{u \in [\tau_n, \tau_{n+1}]} |x_u^1 - x_u^2| \right)$$

which holds for all $t \in [\tau_n, \tau_{n+1}]$. In the case $n = 0$, we have $|z_{\tau_n}^1 - z_{\tau_n}^2| = |\xi^1 - \xi^2|$. For $n \geq 1$, we know that $\beta_t \leq \delta$ for $t \in [\tau_{n-1}, \tau_n]$, therefore

$$|z_{\tau_n}^1 - z_{\tau_n}^2| \leq \sup_{u \in [\tau_{n-1}, \tau_n]} |z_u^1 - z_u^2| \leq \delta + \sup_{u \in [\tau_{n-1}, \tau_n]} |x_u^1 - x_u^2|.$$

This shows that in all cases we have considered, the estimate

$$|z_t^1 - z_t^2| \leq 2e^{2T(C_1 + C_2)} \left(\delta + |\xi^1 - \xi^2| + \|x^1 - x^2\|_\infty \right)$$

holds true which implies that

$$\|z^1 - z^2\|_\infty \leq 2e^{2T(C_1 + C_2)} \left(\delta + |\xi^1 - \xi^2| + \|x^1 - x^2\|_\infty \right).$$

Note that this is true for any $\delta > 0$, therefore we can conclude that

$$\|z^1 - z^2\|_\infty \leq 2e^{2T(C_1 + C_2)} \left(|\xi^1 - \xi^2| + \|x^1 - x^2\|_\infty \right)$$

holds true. The claim follows from the equality $y^i = z^i + x^i$ and the triangle inequality. $\square$

Theorem 2.6. *Assume* $b: \mathbb{R}^d \to \mathbb{R}^d$ *is continuous and satisfies* (2.7) *and* (2.8)*. Let* $X: [0, T] \to \mathbb{R}^d$ *be a continuous Gaussian process with corresponding Gaussian measure* γ *on the space of continuous functions, and let* σ^2 *be defined as in* (1.3)*. Let* Y *be the solution to the SDE* (2.2) *and let* μ *be the law of* Y*.*

Then for every $\nu \in P(C_\xi)$*,*

$$\inf_{\pi \in \Pi(\nu, \mu)} \int_{C_\xi \times C_\xi} \|x - y\|_\infty^2 \, d\pi(x, y) \leq 18\sigma^2 e^{4T(C_1 + C_2)} H(\nu \,|\, \mu).$$

Proof. This is a consequence of Corollary 1.3, the contraction principle in Lemma 4.1 and Proposition 2.5. □

Example 2.7. We finally discuss an example to illustrate our findings. Let $B^{H,K} \colon [0, T] \to \mathbb{R}^m$ be a *bifractional Brownian motion*, i.e. a continuous, centered Gaussian process with independent components and the covariance of each component is given by

$$R(s, t) = \frac{1}{2^K} \left((s^{2H} + t^{2H})^K - |t - s|^{2HK} \right)$$

with $H \in (0, 1)$ and $K \in (0, 1]$. This process was introduced in [HV03] and further studied e.g. in [RT06, KRT07]. Note that for $K = 1$, we obtain a fractional Brownian motion, and for $K = 1$ and $H = 1/2$ we have the usual Brownian motion. In the general form, it is not known whether the process can be written as a stochastic integral with respect to Brownian motion (as for the fractional Brownian motion) or whether it is adapted to a Brownian filtration. This rules out any Girsanov transformation techniques. It can be shown that $B^{H,K}$ has sample paths of α-Hölder regularity for any $\alpha < HK$ (and the sample paths are therefore of finite $1/\alpha$-variation), but not better. In [FGGR16, Example 2.12], it was shown that the corresponding Cameron Martin space can be continuously embedded in the space of paths with finite q-variation for $q = (HK + 1/2)^{-1} \vee 1$. We aim to take $B^{H,K}$ as the driver in equation (2.1). If all σ_i have finite p-variation for $p < (1 - HK)^{-1}$, we can define the stochastic integrals pathwise as Young integrals. We can relax the assumptions on σ if we are only interested in defining the stochastic integral by stochastic means. More precisely, if $HK \geq 1/2$ and if all σ_i are continuous, we can integrate each σ_i against any Cameron Martin path, using Riemann-Stieltjes integrals. If $HK \leq 1/2$ and if all σ_i have finite p-variation for some $p \geq 1$ satisfying

$$\frac{1}{p} > \frac{1}{2} - HK,$$

we can integrate each σ_i against any Cameron Martin path using Young integrals. In these two cases, integration induces a bounded linear map from the associated Cameron Martin space $\mathcal{H}$ to the space $C([0, T], \mathbb{R}^d)$, and therefore also to the Hilbert space $L^2([0, T], \mathbb{R}^d)$. If (e_n) denotes an orthonormal basis of L^2, we define a scalar product

$$\langle x, y \rangle^\sim := \sum_n \frac{1}{n^2} \langle x, e_n \rangle_{L^2} \langle y, e_n \rangle_{L^2}.$$

Let $\tilde{L}^2$ denote the space $L^2([0, T], \mathbb{R}^d)$ equipped with this scalar product. Then integration induces a Hilbert-Schmidt operator from $\mathcal{H}$ to $\tilde{L}^2$, which can therefore be uniquely extended to the whole space $C([0, T], \mathbb{R}^d)$ almost surely and induces a Gaussian measure on $\tilde{L}^2$ (cf. e.g. [Hai09, Theorem 3.44]). It can be shown (using the explicit bounds of this map) that the associated Gaussian process on $\tilde{L}^2$ has actually continuous sample paths almost surely, and that its Cameron Martin space can again be continuously embedded in the space of q-variation paths with the same choice of q. From now on, assume that the σ_i satisfy one of the stated regularity assumptions. In case that the drift b is Lipschitz continuous, we can solve (2.1), and the law μ of the solution Y satisfies the quadratic transport inequality

$$\inf_{\pi \in \Pi(\nu, \mu)} \int_{C_\xi \times C_\xi} d_{q-\mathrm{var}}(x, y)^2 \, d\pi(x, y) \leq C \, H(\nu \, | \, \mu) \tag{2.12}$$

for some constant $C > 0$ and any $\nu \in P(C_\xi)$ by Theorem 2.2. Note that $q < 1/(HK)$ (e.g. $q = 1$ in case of the Brownian motion), and we cannot expect that the sample paths of Y

itself have finite q-variation. Assuming only continuity, (2.7) and (2.8) for b, we can still solve (2.1), and the law μ of the solution Y satisfies the quadratic transport inequality

$$\inf_{\pi \in \Pi(\nu,\mu)} \int_{C_\xi \times C_\xi} \|x - y\|_\infty^2 \, d\pi(x,y) \leq C \, H(\nu \,|\, \mu) \tag{2.13}$$

for another constant $C > 0$ and any $\nu \in P(C_\xi)$ by Theorem 2.6.

In case of the fractional Brownian motion (i.e. $K = 1$), the transport inequalitities (2.12) and (2.13) may be compared to the corresponding results obtained in [Sau12] (namely Theorem 1 and Theorem 3). Note that our results imply those and are even stronger in several regards (quadratic transport inequality instead of simple one, larger metric, less regularity assumptions on the vector fields).

2.2 SDEs with multiplicative noise

Next we will consider SDEs with multiplicative noise, i.e. equations of the form

$$Y_t = \xi + \int_0^t b(Y_s) \, ds + \sum_{i=1}^m \int_0^t \sigma_i(Y_s) \circ dX_s^i \tag{2.14}$$

where $\xi \in \mathbb{R}^d$, $X = (X^1, \ldots, X^m)$ is a continuous m-dimensional Gaussian process and $b, \sigma_1, \ldots \sigma_m \colon \mathbb{R}^d \to \mathbb{R}^d$ are continuous vector fields. The problem in (2.14) is of course to make sense of the stochastic integrals if X is not a martingale.

We start to discuss a simple case; namely, we assume that the driving process is one dimensional. Under further assumptions on the vector fields, we can use the Doss-Sussmann representation to define the solution to (2.14) pathwise for any continuous driving signal[2]. If we further assume that also the solution space is one dimensional we can follow [Sau12] to derive the following result:

Theorem 2.8. *Assume $m = d = 1$ and consider the equation*

$$Y_t = \xi + \int_0^t b(Y_s) \, ds + \int_0^t \sigma(Y_s) \circ dX_s \tag{2.15}$$

where $X \colon [0, T] \to \mathbb{R}$ is a continuous Gaussian process. We further assume that b is bounded by some constant B and Lipschitz continuous with Lipschitz constant L_b. For the diffusion vector field σ, we assume that it is Lipschitz continuous with Lipschitz constant L_σ and that there are constants $0 < \sigma_1 \leq \sigma_2$ such that $\sigma_1 \leq \sigma(x) \leq \sigma_2$ for all $x \in \mathbb{R}$.

Then the equation (2.15) has a unique continuous solution Y and its law μ satisfies the quadratic transport inequality

$$\inf_{\pi \in \Pi(\nu,\mu)} \int_{C_\xi \times C_\xi} \|x - y\|_\infty^2 \, d\pi(x,y) \leq C \, H(\nu \,|\, \mu)$$

for all $\nu \in P(C_\xi)$ where $C > 0$ is a constant depending on the variance of the Gaussian measure given in (1.3), T and all constants above.

Proof. Under the stated conditions, one can show, using the Lamperti transform (cf. [Sau12, proof of Theorem 12 on p. 12]) that the solution map associated to (2.14) is Lipschitz continuous on the space of continuous functions. The result follows from Corollary 1.3 and the contraction principle in Lemma 4.1. $\qquad\square$

---

[2] Note that we can also use rough paths theory for $m = 1$ to make sense of (2.14) since the iterated integrals are canonically given as products in this case.

Note that this result generalizes [Sau12, Theorem 2] to arbitrary Gaussian processes. It is even stronger than [Sau12, Theorem 2] since we can deduce the quadratic transportation inequality, not only the simple one. We can also deduce the quadratic transportation inequality for general Gaussian processes under the conditions stated in [Sau12, Theorem 4] for the uniform metric. Indeed, an inspection of the proof reveals that under these conditions, the solution map associated to (2.14) is again Lipschitz, therefore we can conclude as before.

Now assume that X is an m-dimensional Brownian motion. In contrast to the additive noise case or the one dimensional case, the solution map $I(\cdot, \xi)\colon C_0([0,T], \mathbb{R}^m) \to C_\xi([0,T], \mathbb{R}^d)$ which assigns to each Brownian path the solution path to the SDE (2.14) will in general *not* be (Lipschitz-) continuous. This issue can be overcome using Lyons' rough paths theory. Indeed, rough paths theory shows that there is a Polish space $\mathcal{D}_g^{0,p}$ (cf. [FV10b, Definition 9.15 and Proposition 8.25] for the precise definition) such that the diagram

$$
\begin{array}{ccc}
\mathcal{D}_g^{0,p} & & \\
\uparrow S & \searrow \mathbf{I}(\cdot, \xi) & \\
C_0 & \xrightarrow[I(\cdot, \xi)]{} & C_\xi
\end{array}
$$

(2.16)

commutes almost surely and the map $\mathbf{I}(\cdot, \xi)\colon \mathcal{D}_g^{0,p} \to C_\xi$ is locally Lipschitz continuous. The map $S\colon C_0 \to \mathcal{D}_g^{0,p}$ is constructed w.r.t. the Wiener measure on the path space C_0. Using a pathwise approach, one is not restricted to Wiener measure and it is indeed possible to construct lift maps S w.r.t. more general Gaussian measures γ (cf. [CQ02], [FV10a]). In this case, one *defines* $I(\cdot, \xi) := \mathbf{I}(S(\cdot), \xi)$ which gives rise to solutions of SDEs of the form

$$
Y_t = \xi + \int_0^t b(Y_t) + \sum_{i=1}^m \int_0^t \sigma_i(Y_s) \circ dX_s^i
$$

(2.17)

where $X = (X^1, \ldots, X^m)$ is the canonical process induced by the Gaussian measure γ. Our key result will be that for *Brownian-like* Gaussian processes (we will be more precise later), we have an estimate of the form

$$
\|I(x, \xi) - I(y, \xi)\|_{p-\mathrm{var}} \leq L(y) d_\mathcal{H}(x, y)
$$

almost surely for every $x, y \in C_0$ where L is a random variable which possesses every moment w.r.t. the Gaussian measure γ. Together with Lemma 4.1, this yields a transportation inequality which is stated in Theorem 2.14.

We will not make an attempt to give an overview to rough paths theory since we will use it merely as a tool. Instead, we refer to the monographs [LQ02], [LCL07], [FV10b] and [FH14]. The terms and notation we are using coincides with the one from [FV10b] with the only exception that we use the symbol $\mathcal{D}_g^{0,p}$ to denote the space of geometric p-variation rough paths $C_0^{0,p-\mathrm{var}}([0,T]; G^{[p]}(\mathbb{R}^m))$ equipped with the p-variation metric. By [FV10b, Proposition 8.25], this space is Polish.

We start with some deterministic estimates for rough paths. If ω is a control function and $\alpha > 0$, recall the definition of $N_\alpha(\omega; [s,t])$ resp. of $N_\alpha(\mathbf{x}; [s,t])$ for geometric rough paths $\mathbf{x}$ ([CLL13], [FR13]). The next proposition is a version of [BFRS16, Theorem 4] for the p-variation metric.

Proposition 2.9. *Let* $\mathbf{x}^1$ *and* $\mathbf{x}^2$ *be weakly geometric* p-*rough paths for some* $p \geq 1$. *Consider the rough differential equations (RDEs)*

$$dy_t^j = \sigma^j(y_t^j)\,d\mathbf{x}_t^j;\ y_S^j \in \mathbb{R}^d$$

for $j = 1, 2$ *on some interval* $[S, T]$ *where* $f^1 = (f_i^1)_{i=1,\ldots,m}$ *and* $f^2 = (f_i^2)_{i=1,\ldots,m}$ *are two families of vector fields in* $\mathbb{R}^d$, $\theta > p$ *and* β *is a bound on*[3] $|\sigma^1|_{Lip^\theta}$ *and* $|\sigma^2|_{Lip^\theta}$.
   Then for every $\alpha > 0$ *there is a constant* $C = C(\theta, p, \beta, \alpha)$ *such that*

$$
\begin{aligned}
d_{p-\mathrm{var};[S,T]}(y^1, y^2) &\leq C\left[|y_S^1 - y_S^2| + \left|\sigma^1 - \sigma^2\right|_{Lip^{\theta-1}} + \rho_{p-\mathrm{var};[S,T]}(\mathbf{x}^1, \mathbf{x}^2)\right] \\
&\quad \times (\|\mathbf{x}^1\|_{p-\mathrm{var};[S,T]} + \|\mathbf{x}^2\|_{p-\mathrm{var};[S,T]} + 1) \\
&\quad \times \exp\left\{C\left(N_\alpha(\mathbf{x}^1; [S,T]) + N_\alpha(\mathbf{x}^2; [S,T]) + 1\right)\right\}
\end{aligned}
$$

holds.

Proof. The proof follows [BFRS16, Lemma 7 and Theorem 4]. Let ω be a control function such that $\sup_{s<t} \frac{\|\mathbf{x}^j\|}{\omega(s,t)^{1/p}} \leq 1$ for $j = 1, 2$. Set $\bar{y} := y^1 - y^2$ and

$$\kappa := \frac{\left|\sigma^1 - \sigma^2\right|_{\mathrm{Lip}^{\theta-1}}}{\beta} + \rho_{p-\omega;[S,T]}(\mathbf{x}^1, \mathbf{x}^2).$$

We claim that there is a constant $C = C(\theta, p)$ such that for every $s < t$,

$$\|\bar{y}\|_{p-\mathrm{var};[s,t]} \leq C\beta\omega(s,t)^{\frac{1}{p}}\left(\|\bar{y}\|_{\infty;[s,t]} + \kappa\right)\exp\left\{C\beta^p(N_\alpha(\omega; [s,t]) + 1)\right\}. \tag{2.18}$$

Indeed, as it was shown in the proof of [BFRS16, Lemma 7],

$$2^{1-p}|\bar{y}_{s,v}|^p \leq C\beta^p\omega(u,v)(|\bar{y}_u| + \kappa)^p\exp\left\{C\beta^p\omega(u,v)\right\} + |\bar{y}_{s,u}|^p$$

for every $[u, v] \subseteq [s, t]$. Thus if $s = \tau_0 < \ldots < \tau_M < \tau_{M+1} = v$,

$$|\bar{y}_{s,v}|^p \leq 2^{(M+1)(p-1)}C\beta^p\omega(s,v)(\|\bar{y}\|_{\infty;[s,v]} + \kappa)^p\exp\left\{C\beta^p\sum_{i=0}^M \omega(\tau_i, \tau_{i+1})\right\}$$

for every $s \leq v \leq t$. Choosing $\tau_0 = s$, $\tau_{i+1} = \inf_t\{\omega(\tau_i, t) \geq \alpha\} \wedge v$ gives

$$|\bar{y}_{s,v}|^p \leq C\beta^p\omega(s,v)(\|\bar{y}\|_{\infty;[s,v]} + \kappa)^p\exp\left\{C\beta^p(N_\alpha(\omega; [s,v]) + 1)\right\}$$

for every $s \leq v \leq t$ and (2.18) follows. Now we can use the conclusion from [BFRS16, Lemma 7] to see that

$$\|\bar{y}\|_{p-\mathrm{var};[s,t]} \leq C\beta\omega(s,t)^{\frac{1}{p}}\left(\|\bar{y}_s\| + \kappa\right)\exp\left\{C\beta^p(N_\alpha(\omega; [s,t]) + 1)\right\}$$

holds for every $s < t$. We conclude as in [BFRS16, Theorem 4]. $\qquad\square$

Lemma 2.10. *Let* $\mathbf{x}$ *be a weakly geometric* p-*rough path with* $p \in [2, 3)$ *and* h *a path of finite* q-*variation with* $1 \leq q \leq p$ *and* $\frac{1}{p} + \frac{1}{q} > 1$. *Then there is a constant* $C = C(p, q)$ *such that*

$$\rho_{p-\mathrm{var};[S,T]}(T_h(\mathbf{x}), \mathbf{x}) \leq C_{p,q}(1 \vee \|x\|_{p-\mathrm{var};[S,T]})(\|h\|_{q-\mathrm{var};[S,T]} + \|h\|_{q-\mathrm{var};[S,T]}^2)$$

where $T_h(\mathbf{x})$ *denotes the translation of* $\mathbf{x}$ *by* h, *cf.* [FV10b, Section 9.4.6].

---

[3]We mean Lipschitz it the sense of Stein, cf. [FV10b, Chapter 10]

Proof. Recall that

$$\rho_{p-\text{var};[S,T]}(\mathbf{x},\mathbf{y}) = \sup_{D \in \mathcal{P}([S,T])} \left(\sum_{t_i \in D} |x_{t_i,t_{i+1}} - y_{t_i,t_{i+1}}|^p \right)^{\frac{1}{p}}$$
$$+ \sup_{D \in \mathcal{P}([S,T])} \left(\sum_{t_i \in D} |\mathbf{x}^2_{t_i,t_{i+1}} - \mathbf{y}^2_{t_i,t_{i+1}}|^{p/2} \right)^{\frac{2}{p}}.$$

Therefore, we immediately obtain

$$\rho_{p-\text{var};[S,T]}(T_h(\mathbf{x}),\mathbf{x}) \le \|h\|_{q-\text{var};[S,T]} + \sup_{D \in \mathcal{P}([S,T])} \left(\sum_{t_i \in D} |T_h(\mathbf{x})^2_{t_i,t_{i+1}} - \mathbf{x}^2_{t_i,t_{i+1}}|^{p/2} \right)^{\frac{2}{p}}.$$

Concerning the second term, fix some $D \in \mathcal{P}([S,T])$. We have

$$\sum_{t_i \in D} |T_h(\mathbf{x})^2_{t_i,t_{i+1}} - \mathbf{x}^2_{t_i,t_{i+1}}|^{p/2} = \sum_{t_i \in D} \left| \iint_{\Delta^2_{t_i,t_{i+1}}} d(x+h) \otimes d(x+h) - \int_{\Delta^2_{t_i,t_{i+1}}} dx \otimes dx \right|^{p/2}$$

and

$$\left| \iint_{\Delta^2_{t_i,t_{i+1}}} d(x+h) \otimes d(x+h) - \int_{\Delta^2_{t_i,t_{i+1}}} dx \otimes dx \right|$$
$$\le \left| \iint_{\Delta^2_{t_i,t_{i+1}}} dh \otimes d(x+h) \right| + \left| \iint_{\Delta^2_{t_i,t_{i+1}}} dx \otimes dh \right|$$
$$\le C_{p,q} \|h\|_{q-\text{var};[t_i,t_{i+1}]} \left(\|x+h\|_{p-\text{var};[t_i,t_{i+1}]} + \|x\|_{p-\text{var};[t_i,t_{i+1}]} \right)$$

by the estimates for the Young integral. From Hölder's inequality,

$$\sum_{t_i \in D} |T_h(\mathbf{x})^2_{t_i,t_{i+1}} - \mathbf{x}^2_{t_i,t_{i+1}}|^{p/2}$$
$$\le C_{p,q} \left(\sum_{t_i} \|h\|^q_{q-\text{var};[t_i,t_{i+1}]} \right)^{\frac{p}{2q}} \left(\sum_{t_i} \|x+h\|^p_{p-\text{var};[t_i,t_{i+1}]} + \|x\|^p_{p-\text{var};[t_i,t_{i+1}]} \right)^{\frac{1}{2}}$$
$$\le C_{p,q} \|h\|^{p/2}_{q-\text{var};[S,T]} (\|x+h\|^{p/2}_{p-\text{var};[S,T]} + \|x\|^{p/2}_{p-\text{var};[S,T]})$$

and the result follows from the triangle inequality for the p-variation seminorm and standard estimates. $\qquad\square$

Lemma 2.11. *Let* $\mathbf{x}^1 := \mathbf{x}$ *and* $\mathbf{x}^2 := T_h(\mathbf{x})$ *where* $\mathbf{x}$ *is a weakly geometric* p-*rough path for some* $p \in [1,3)$ *and* h *is a path of finite* q-*variation with* $\frac{1}{p} + \frac{1}{q} > 1$. *Consider the solutions* y^1 *and* y^2 *to the RDEs as in Proposition 2.9 with* $f^1 = f^2$ *and* $y^1_S = y^2_S$. *Then*

$$d_{p-\text{var};[S,T]}(y^1,y^2) \le C \exp\{C(N_1(\mathbf{x};[S,T])+1)\}(\|h\|_{q-\text{var};[S,T]} \vee \|h\|^q_{q-\text{var};[S,T]})$$

where C *is a constant depending on* p,q,θ *and* β.

Proof. We will only consider the case $p \in [2,3)$, the case $p \in [1,2)$ is similar (and easier). Let $\|h\|_{q-\text{var};[S,T]} \le 1$. We claim that

$$d_{p-\text{var};[S,T]}(y^1,y^2) \le C \exp\{C(N_1(\mathbf{x};[S,T])+1)\} \|h\|_{q-\text{var};[S,T]} \qquad (2.19)$$

holds for some constant C. Indeed: From Proposition 2.9 we know that for every $\alpha > 0$,

$$d_{p-\mathrm{var};[S,T]}(y^1, y^2) \leq C(\|\mathbf{x}\|_{p-\mathrm{var};[S,T]} + \|T_h(\mathbf{x})\|_{p-\mathrm{var};[S,T]} + 1)$$
$$\times \exp\left\{C(N_\alpha(\mathbf{x};[S,T]) + N_\alpha(T_h(\mathbf{x});[S,T]) + 1)\right\} \rho_{p-\mathrm{var};[S,T]}(\mathbf{x}, T_h(\mathbf{x})).$$

Using [FV10b, Theorem 9.33], [BFRS16, Lemma 5], [FR13, Lemma 1] and the assumption $\|h\|_{q-\mathrm{var};[S,T]} \leq 1$ shows that

$$d_{p-\mathrm{var};[S,T]}(y^1, y^2) \leq C(\|\mathbf{x}\|_{p-\mathrm{var};[S,T]} + 1) \exp\left\{C(N_1(\mathbf{x};[S,T]) + 1)\right\} \rho_{p-\mathrm{var};[S,T]}(\mathbf{x}, T_h(\mathbf{x}))$$

for a larger constant C and α chosen appropriately. Applying Lemma 2.10 shows (2.19), using the estimate $\|\mathbf{x}\|_{p-\mathrm{var};[S,T]} \leq N_1(\mathbf{x};[S,T]) + 1$ which was proven in [FR13, Lemma 4].

Now let $\|h\|_{q-\mathrm{var};[S,T]} \geq 1$. In this case,

$$d_{p-\mathrm{var};[S,T]}(y^1, y^2) \leq \|y^1\|_{p-\mathrm{var};[S,T]} + \|y^2\|_{p-\mathrm{var};[S,T]}$$
$$\leq C(N_\alpha(\mathbf{x};[S,T]) + N_\alpha(T_h(\mathbf{x});[S,T]) + 1)$$

using the deterministic estimates for the Itō–Lyons map proven in [FR13]. With [FH14, Lemma 11.12], we conclude that

$$d_{p-\mathrm{var};[S,T]}(y^1, y^2) \leq C(N_1(\mathbf{x};[S,T]) + 1)\|h\|_{q-\mathrm{var};[S,T]}^q$$

for a larger constant C. $\qquad\square$

We come back to our original setup. Assume that γ is a Gaussian measure on the Borel sets of the Banach space $C_0([0,T], \mathbb{R}^m)$ induced by a continuous $\mathbb{R}^m$-valued Gaussian process X. As usual, we denote the corresponding Cameron-Martin space by $\mathcal{H}$. Assume that there is some $p \in [1,3)$ and a measurable lift map $S\colon C_0 \to \mathcal{D}_g^{0,p}$ such that the diagram (2.16) commutes on a set of full γ-measure. Will now make further assumptions on our lift map S: Suppose that

(i) There is a continuous embedding

$$\iota\colon \mathcal{H} \hookrightarrow C^{q-\mathrm{var}}([0,T], \mathbb{R}^d); \quad 1 \leq q \leq p$$

with $\frac{1}{p} + \frac{1}{q} > 1$ (note that this implies $1 \leq q < 2$ when $p \geq 2$).

(ii) The set

$$\{x \in C_0 \mid S(x+h) = T_h(S(x)) \text{ for all } h \in \mathcal{H}\}$$

has full γ-measure.

Remark 2.12. Assumption (i) and (ii) are trivially satisfied for $p \in [1,2)$. More generally, they hold if the covariance of the corresponding $\mathbb{R}^m$-valued Gaussian process has *mixed $(1,\rho)$-variation* for some $\rho \in [1,3/2)$ with $q = 2(1/\rho + 1)^{-1}$, cf. [FGGR16, p. 688] for the definition of mixed variation and [FGGR16, Theorem 1.1] and [FV10b, Lemma 15.58] for the corresponding results. A list of processes which satisfy this condition can also be found in [FGGR16]. In particular, they hold for the Stratonovich lift of the Brownian motion with $q = 1$.

Under these two conditions, the following Proposition is an immediate consequence of Lemma 2.11.

Proposition 2.13. *Consider the RDEs as in Proposition 2.9.*
  Then

$$d_{p-\text{var}}(y^1, y^2) \leq L(x^1)(\|\iota\|_{\mathcal{H} \hookrightarrow C^{q-\text{var}}} \vee \|\iota\|_{\mathcal{H} \hookrightarrow C^{q-\text{var}}}^q)(d_{\mathcal{H}}(x^1, x^2) \vee d_{\mathcal{H}}(x^1, x^2)^q)$$

for all $x^1, x^2 \in C_0$ where

$$L(x) = C \exp\left\{C(N_1(S(x); [0, T]) + 1)\right\}$$

and C is a constant depending on p, q, θ and β.

The next theorem is our main result for the multiplicative case.

Theorem 2.14. *Let Y be the solution to the SDE (2.17) driven by the Gaussian process X defined pathwise via the diagram (2.16) and let μ be the law of Y. Assume that $q = 1$. Then for every $\varepsilon > 0$ there is a constant C depending on ε, p, θ and β such that for every $\nu \in P(C_\xi)$,*

$$\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{C_\xi \times C_\xi} d_{p-\text{var}}(x, y)^{2-\varepsilon} \, d\pi(x, y) \right)^{\frac{1}{2-\varepsilon}} \leq C \|\iota\|_{\mathcal{H} \hookrightarrow C^{1-\text{var}}} \sqrt{H(\nu \mid \mu)}.$$

Proof. From [CLL13, Theorem 6.3] we know that $N_1(S; [0, T])$ has Gaussian tails w.r.t. γ, hence $\| \exp\{C(N_1(S; [0, T]) + 1)\}\|_{L^q(\gamma)} < \infty$ for every $q \in [1, \infty)$. The assertion follows from Theorem 1.2, Proposition 2.13 and the contraction principle Lemma 4.1. $\square$

Example 2.15. Let us come back to the bifractional Brownian motion $B^{H,K} : [0, T] \to \mathbb{R}^m$ already considered in Example 2.7. In [RT06] and [KRT07], it was shown that in the case $2HK = 1$, the process has many similarities to the usual Brownian motion. The same holds true here: From [FGGR16, Example 2.12], we know that the covariance of the bifractional Brownian motion has mixed $(1, \rho)$-variation for $\rho = (2HK)^{-1}$. In particular, we can choose $q = 1$ in the case $2HK = 1$, and Theorem 2.14 applies. Therefore, we can *almost* (i.e. modulo an ε-correction) deduce Talagrand's transport inequality in this case. However, for the Brownian motion (which we obtain for the choice $K = 1$ and $H = 1/2$), it is known that Talagrand's inequality holds for the uniform distance (which is smaller than the p-variation distance) even without ε-correction, cf. [Üst12]. It remains an open problem how to obtain the full 2- transport inequality for diffusions driven by a multidimensional Brownian motion without using the Girsanov transformation.

3 Tail estimates for functionals

In the following, we aim to motivate why it is useful to have p-transportation–cost inequalities for $p > 1$. This section is independent of the former one and may be interesting in its own right.

It is well known that transportation–cost inequalities imply Gaussian measure concentration. This was first disovered by Marton ([Mar86], [Mar96]). In [DGW04], it was shown that for $p = 1$, the converse is true: Gaussian tails imply the 1-transportation–cost inequality. In the case of a Gaussian Banach space $(E, \mathcal{H}, \gamma)$, it is a classical result (cf. [Bog98, 4.5.6. Theorem]) that $\mathcal{H}$-Lipschitz functions on Gaussian spaces have Gaussian tails. This result was further generalized in [DOR15, Theorem 17] (cf. also [FO10] and [FH14, Theorem 11.7]) where it was shown that the linear growth of a function in $\mathcal{H}$-direction already implies that it has Gaussian tails. More precisely, if there is a constant $\sigma > 0$ and a measurable map $g \colon E \to [0, \infty]$ for which $g < \infty$ on a set of positive γ-measure such that $f \colon E \to [0, \infty]$ satisfies

$$f(x + h) \leq g(x) + \sigma|h|_{\mathcal{H}} \tag{3.1}$$

for all x on a set of full γ-measure and all $h \in \mathcal{H}$, then f has Gaussian tails. In the following, we will prove an abstract result which will imply that we may even choose σ random in (3.1) and still obtain Gaussian tails for f.

Theorem 3.1. *Let E be a linear Polish space and let μ be a probability measure defined on its Borel σ–algebra. Assume that there is a normed subspace $\mathcal{U} \subseteq E$ and let $d_{\mathcal{U}} \colon E \times E \to [0, \infty]$ be defined as*

$$
d_{\mathcal{U}}(x, y) = \begin{cases} |x - y|_{\mathcal{U}} & \text{if } x - y \in \mathcal{U} \\ +\infty & \text{otherwise.} \end{cases}
$$

Assume that there is a $p \in [1, \infty)$ and a constant C such that for every $\nu \in P(E)$,

$$
\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{E \times E} d_{\mathcal{U}}(x, y)^p \, d\pi(x, y) \right)^{\frac{1}{p}} \leq \sqrt{CH(\nu \mid \mu)}.
$$

Let (F, d) be some metric space and let $f \colon E \to F$ be measurable w.r.t. the Borel σ–algebra. Choose $r_0 \geq 0$ and some element $e \in E$ such that

$$
\mu \left\{ x \in E \, : \, d(f(x), e) \leq r_0 \right\} =: a > 0.
$$

Assume that there are measurable functions $g, \sigma \colon E \to [0, \infty]$ such that

$$
d(f(x + h), e) \leq g(x) + \sigma(x) |h|_{\mathcal{U}}
$$

holds for every $x \in E$ and every $h \in \mathcal{U}$, and assume that $g \in L^1(\mu)$ and $\sigma \in L^q(\mu)$ where $q \in (1, \infty]$ is chosen such that $\frac{1}{q} + \frac{1}{q} = 1$.
  Then

$$
\mu \left\{ x \in E \, : \, d(f(x), e) > r \right\} \leq \exp \left\{ -\frac{(r - r_1)^2}{C \|\sigma\|_{L^q(\mu)}^2} \right\}
$$

for all $r \geq r_1$ where $r_1 = r_0 + 4\|g\|_{L^1(\mu)} + \|\sigma\|_{L^q(\mu)} \sqrt{2C \log(a^{-1})}$. In particular, the random variable $d(f(\cdot), e) \colon E \to [0, \infty)$ has Gaussian tails.

Proof. For $x, y \in E$ set

$$
d_f(x, y) = d(f(x), f(y)).
$$

For any measurable set $A \subseteq E$ and $r \geq 0$ we define

$$
A^r := \left\{ x \in E \, : \, \text{there is an } \bar{x} \in A \text{ such that } d_f(x, \bar{x}) \leq r \right\}.
$$

Fix some $r \geq 0$ and set $B := (A^r)^c$. Assume first that A and B have positive measure. On E, we define the measures

$$
d\mu_A := \frac{\mathbb{1}_A}{\mu(A)} d\mu \quad \text{and} \quad d\mu_B := \frac{\mathbb{1}_B}{\mu(B)} d\mu.
$$

Then

$$
\begin{aligned}
r &\leq \inf_{\pi \in \Pi(\mu_A, \mu_B)} \int_{E \times E} d_f(x, y) \, d\pi(x, y) \\
&\leq \inf_{\pi \in \Pi(\mu_A, \mu)} \int_{E \times E} d_f(x, y) \, d\pi(x, y) + \inf_{\pi \in \Pi(\mu_B, \mu)} \int_{E \times E} d_f(x, y) \, d\pi(x, y)
\end{aligned}
$$

where we used symmetry and the triangle inequality for optimal transportation costs. (The triangle inequality can be deduced, in our case, exactly as for the usual Wasserstein metric using the Gluing Lemma [Vil03, Lemma 7.6] and the triangle inequality for d_f, cf. [Vil03, Theorem 7.3].) If $x - y = h \in \mathcal{U}$, we obtain by assumption

$$d_f(x, y) \leq 2g(y) + \sigma(y)|h|_{\mathcal{U}}.$$

This implies that for all $x, y \in E$,

$$d_f(x, y) \leq 2g(y) + \sigma(y)d_{\mathcal{U}}(x, y).$$

It follows that

$$r \leq \inf_{\pi \in \Pi(\mu_A, \mu)} \int_{E \times E} \sigma(y)d_{\mathcal{U}}(x, y) \, d\pi(x, y) + \inf_{\pi \in \Pi(\mu_B, \mu)} \int_{E \times E} \sigma(y)d_{\mathcal{U}}(x, y) \, d\pi(x, y) + 4\|g\|_{L^1}$$

$$\leq \|\sigma\|_{L^q} \inf_{\pi \in \Pi(\mu_A, \mu)} \left(\int_{E \times E} d_{\mathcal{U}}(x, y)^p \, d\pi(x, y) \right)^{\frac{1}{p}}$$

$$+ \|\sigma\|_{L^q} \inf_{\pi \in \Pi(\mu_B, \mu)} \left(\int_{E \times E} d_{\mathcal{U}}(x, y)^p \, d\pi(x, y) \right)^{\frac{1}{p}} + 4\|g\|_{L^1}$$

$$\leq \|\sigma\|_{L^q} \left(\sqrt{CH(\mu_A \,|\, \mu)} + \sqrt{CH(\mu_B \,|\, \mu)} \right) + 4\|g\|_{L^1}$$

$$= \|\sigma\|_{L^q} \left(\sqrt{C \log(\mu(A)^{-1})} + \sqrt{C \log(\mu(B)^{-1})} \right) + 4\|g\|_{L^1}.$$

Rearranging terms, we see that

$$\mu(A^r) \geq 1 - \exp\left\{ -\frac{(r - \hat{r})^2}{C\|\sigma\|_{L^q}^2} \right\}$$

for every $r \geq \hat{r}$ where $\hat{r} := \|\sigma\|_{L^q}\sqrt{C \log(\mu(A)^{-1})} + 4\|g\|_{L^1}$. Now set

$$A := \{x \in E \,:\, d(f(x), e) \leq r_0\}.$$

By assumption, $\mu(A) = a > 0$. For every $r \geq 0$, we have

$$A^r \subseteq \{x \in E \,:\, d(f(x), e) \leq r_0 + r\}.$$

If $\mu(B) = 0$, it follows that $\{x \in E \,:\, d(f(x), e) \leq r_0 + r\}$ has full measure. In other words, $d(f(\cdot), e)$ is bounded almost surely and the claimed estimate is trivial. If $\mu(B) > 0$, we can use our calculations above to conclude that

$$1 - \exp\left\{ -\frac{(r - \hat{r})^2}{C\|\sigma\|_{L^q}^2} \right\} \leq \mu\{x \in E \,:\, d(f(x), e) \leq r_0 + r\}$$

holds for every $r \geq \hat{r}$ and the claim follows. $\qquad\square$

In the Gaussian case, Theorem 1.2 immediately implies

Corollary 3.2. *Let $(F, \mathcal{H}, \gamma)$ be a Gaussian Fréchet space and $f \colon F \to [0, \infty]$ be measurable. Assume that there are nonnegative random variables $g \in L^1(\gamma)$ and $\sigma \in L^2(\gamma)$ such that*

$$f(x + h) \leq g(x) + \sigma(x)|h|_{\mathcal{H}}$$

holds for every $x \in F$ and $h \in \mathcal{H}$.
 Then f has Gaussian tails.

Remark 3.3. Corollary 3.2 is more universal than the *Generalized Fernique Theorem* proven in [DOR15, Theorem 17] and [FH14, Theorem 11.7] (cf. also [FO10]) since it allows σ to be an $L^2(\gamma)$-random variable. Moreover, our proof does not rely on the Borell-Sudakov-Cirelson inequality ([Bor75], [SC74]) and can be applied in more general frameworks whenenver transport inequalities are available. Non-Gaussian examples include the law of diffusions driven by Gaussian processes, as was shown in this work.

4 Appendix

4.1 A generalized contraction principle

The next Lemma is a generalization of [DGW04, Lemma 2.1].

Lemma 4.1. *Let $(X, \mathcal{F})$ be a measurable space on which regular conditional distributions exist and let $c\colon X \times X \to \mathbb{R}_+ \cup \{+\infty\}$ be a measurable function. Assume that there is a measure $\mu \in P(X)$ such that*

$$\inf_{\pi \in \Pi(\nu,\mu)} \left(\int_{X \times X} c(x,y)^p \, d\pi(x,y) \right)^{\frac{1}{p}} \le \sqrt{C H(\nu \,|\, \mu)}$$

holds for every $\nu \in P(X)$ where C is some constant and $p \in [1,\infty)$. Let $(Y, \mathcal{G})$ be another measurable space, $\tilde{c}\colon Y \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be a measurable function and assume that there is a measurable function $\Psi\colon X \to Y$ for which

$$\tilde{c}(\Psi(x), \Psi(y)) \le L(y) c(x,y)$$

holds for every $x, y \in X_0$ where $X_0 \subseteq X$ has full measure w.r.t. μ and $L\colon X \to \mathbb{R} \cup \{+\infty\}$ is another measurable function. Set $\tilde{\mu} := \mu \circ \Psi^{-1}$. Then for every $\tilde{p} \in [1,p]$,

$$\inf_{\tilde{\pi} \in \Pi(\tilde{\nu},\tilde{\mu})} \left(\int_{Y \times Y} \tilde{c}(x,y)^{\tilde{p}} \, d\tilde{\pi}(x,y) \right)^{\frac{1}{\tilde{p}}} \le \|L\|_{L^q(\mu)} \sqrt{C H(\tilde{\nu} \,|\, \tilde{\mu})}$$

holds for very $\tilde{\nu} \in P(Y)$ where $q \in (1,\infty]$ is chosen such that $\frac{1}{q} + \frac{1}{p} = \frac{1}{\tilde{p}}$.

Proof. W.l.o.g. we may assume $C = 1$. Let $\tilde{\nu} \in P(Y)$ and assume that $H(\tilde{\nu} \,|\, \tilde{\mu}) < \infty$. Choose $\nu \in P(X)$ such that $\tilde{\nu} = \nu \circ \Psi^{-1}$ and $\nu \ll \mu$ (note that there is at least one ν which fulfills this condition; e.g. $\nu_0(dx) := \frac{d\tilde{\nu}}{d\tilde{\mu}}(\Psi(x))\mu(dx)$). Then

$$\inf_{\tilde{\pi} \in \Pi(\tilde{\nu},\tilde{\mu})} \int \tilde{c}(x,y)^{\tilde{p}} \, d\tilde{\pi}(x,y) \le \inf_{\pi \in \Pi(\nu,\mu)} \int_{Y \times Y} \tilde{c}(x,y)^{\tilde{p}} \, d(\pi \circ (\Psi \times \Psi)^{-1})(x,y)$$

$$= \inf_{\pi \in \Pi(\nu,\mu)} \int_{X \times X} \tilde{c}(\Psi(x), \Psi(y))^{\tilde{p}} \, d\pi(x,y).$$

Since $\nu \ll \mu$, $X_0 \times X_0$ has full measure for every $\pi \in \Pi(\nu, \mu)$, therefore

$$\inf_{\pi \in \Pi(\nu,\mu)} \int_{X \times X} \tilde{c}(\Psi(x), \Psi(y))^{\tilde{p}} \, d\pi(x,y) \le \inf_{\pi \in \Pi(\nu,\mu)} \int_{X \times X} (L(y) c(x,y))^{\tilde{p}} \, d\pi(x,y)$$

$$\le \|L\|_{L^q(\mu)}^{\tilde{p}} \inf_{\pi \in \Pi(\nu,\mu)} \left(\int_{X \times X} c(x,y)^p \, d\pi(x,y) \right)^{\frac{\tilde{p}}{p}}$$

by Hölder's inequality. The assertion follows from the identity

$$H(\tilde{\nu} \,|\, \tilde{\mu}) = \inf\{H(\nu \,|\, \mu) \mid \nu \in P(X) \text{ s.t. } \nu \circ \Psi^{-1} = \tilde{\nu}\} \tag{4.1}$$

which holds under the assumption that regular conditional distributions exist on $(X, \mathcal{F})$, see [DGW04, Lemma 2.1]. $\qquad\square$

References

[BFRS16] Christian Bayer, Peter K. Friz, Sebastian Riedel, and John Schoenmakers. From rough path estimates to multilevel Monte Carlo. *SIAM J. Numer. Anal.*, 54(3):1449–1483, 2016. MR-3501366

[BG99]   Sergey G. Bobkov and Friedrich Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999. MR-1682772

[Bog98]   Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998. MR-1642391

[Bor75]   Christer Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975. MR-0399402

[CHJ13]   Sonja Cox, Martin Hutzenthaler, and Arnulf Jentzen. Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv:1309.5595*, pages 1–84, 2013.

[CLL13]   Thomas Cass, Christian Litterer, and Terry J. Lyons. Integrability and tail estimates for Gaussian rough differential equations. *Ann. Probab.*, 41(4):3026–3050, 2013.

[CQ02]   Laure Coutin and Zhongmin Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields*, 122(1):108–140, 2002. MR-1883719

[DGW04]   Hacene Djellout, Arnaud Guillin, and Li-ming Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 2004.

[DOR15]   Joscha Diehl, Harald Oberhauser, and Sebastian Riedel. A Lévy area between Brownian motion and rough paths with applications to robust nonlinear filtering and rough partial differential equations. *Stochastic Process. Appl.*, 125(1):161–181, 2015.

[FGGR16]   Peter K. Friz, Benjamin Gess, Archil Gulisashvili, and Sebastian Riedel. The Jain-Monrad criterion for rough paths and applications to random Fourier series and non-Markovian Hörmander theory. *Ann. Probab.*, 44(1):684–738, 2016.

[FH14]   Peter K. Friz and Martin Hairer. *A Course on Rough Paths with an introduction to regularity structures*, volume XIV of *Universitext*. Springer, Berlin, 2014. MR-3289027

[FO10]   Peter K. Friz and Harald Oberhauser. A generalized Fernique theorem and applications. *Proc. Amer. Math. Soc.*, 138(10):3679–3688, 2010. MR-2661566

[FR13]   Peter K. Friz and Sebastian Riedel. Integrability of (non-)linear rough differential equations and integrals. *Stoch. Anal. Appl.*, 31(2):336–358, 2013.

[FÜ04]   Denis Feyel and Ali S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.

[FV10a]   Peter K. Friz and Nicolas B. Victoir. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(2):369–413, 2010.

[FV10b]   Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[Gen01]   Ivan Gentil. *Inégalités de Sobolev logarithmiques et hypercontractivité en mécanique statistique et en EDP*. PhD thesis, Université Paul Sabatier Toulouse III, 2001.

[GIP15]   Massimiliano Gubinelli, Peter Imkeller, and Nicolas Perkowski. Paracontrolled distributions and singular PDEs. *Forum Math. Pi*, 3:e6, 75, 2015. MR-3406823

[GL07]   Nathael Gozlan and Christian Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields*, 139(1-2):235–283, 2007.

[GL10]   Nathael Gozlan and Christian Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.

[GL13]   Nicola Gigli and Michel Ledoux. From log Sobolev to Talagrand: a quick proof. *Discrete Contin. Dyn. Syst.*, 33(5):1927–1935, 2013.

[Goz09]   Nathael Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.

[Gue12]   Toufik Guendouzi. Transportation inequalities for SDEs involving fractional Brownian motion and standard Brownian motion. *Adv. Model. Optim.*, 14(3):615–634, 2012.

[Hai09]   Martin Hairer. *An Introduction to Stochastic PDEs.* Preprint, Warwick, 2009.

[Hai13]   Martin Hairer. Solving the KPZ equation. *Ann. of Math. (2)*, 178(2):559–664, 2013. MR-3071506

[Hai14]   Martin Hairer. A theory of regularity structures. *Invent. Math.*, 198(2):269–504, 2014.

[HV03]   Christian Houdré and José Villa. An example of infinite dimensional quasi-helix. In *Stochastic models (Mexico City, 2002)*, volume 336 of *Contemp. Math.*, pages 195–201. Amer. Math. Soc., Providence, RI, 2003.

[KRT07]   Ida Kruk, Francesco Russo, and Ciprian A. Tudor. Wiener integrals, Malliavin calculus and covariance measure structure. *J. Funct. Anal.*, 249(1):92–142, 2007. MR-2338856

[LCL07]   Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[Led96]   Michel Ledoux. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Math.*, pages 165–294. Springer, Berlin, 1996.

[Led01]   Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[LQ02]   Terry J. Lyons and Zhongmin Qian. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.

[Mar86]   Katalin Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3):445–446, 1986.

[Mar96]   Katalin Marton. Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996. MR-1404531

[Mau91]   Bernard Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2):188–197, 1991.

[MVN68]   Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968. MR-0242239

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

[PR07]   Claudia Prévôt and Michael Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[RS17]   Sebastian Riedel and Michael Scheutzow. Rough differential equations with unbounded drift term. *J. Differential Equations*, 262(1):283–312, 2017. MR-3567487

[RT06]   Francesco Russo and Ciprian A. Tudor. On bifractional Brownian motion. *Stochastic Process. Appl.*, 116(5):830–856, 2006.

[Sau12]   Bruno Saussereau. Transportation inequalities for stochastic differential equations driven by a fractional Brownian motion. *Bernoulli*, 18(1):1–23, 2012.

[SC74]   Vladimir N. Sudakov and Boris S. Cirel′son. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 41:14–24, 165, 1974. Problems in the theory of probability distributions, II.

[SS17]   Michael Scheutzow and Susanne Schulze. Strong completeness and semi-flows for stochastic differential equations with monotone drift. *J. Math. Anal. Appl.*, 446(2):1555-1570, 2017.

[Tal96]   Michel P. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

[Üst12]   Ali S. Üstünel. Transportation cost inequalities for diffusions under uniform distance. In *Stochastic analysis and related topics*, volume 22 of *Springer Proc. Math. Stat.*, pages 203–214. Springer, Heidelberg, 2012.

[Vil03]   Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[Wiś94]   Andrzej Wiśniewski. The structure of measurable mappings on metric spaces. *Proc. Amer. Math. Soc.*, 122(1):147–150, 1994.

[WZ04]   Li-ming Wu and Zheng-liang Zhang. Talagrand's T_2-transportation inequality w.r.t. a uniform metric for diffusions. *Acta Math. Appl. Sin. Engl. Ser.*, 20(3):357–364, 2004.

[You36]   Laurence C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.