TALAGRAND-TYPE TRANSPORT INEQUALITIES FOR PATH SPACES OVER CARNOT GROUPS

PETER K. FRIZ[1,2], HELENA KREMP[1,2], VAIOS LASCHOS[2], MATTHIAS LIERO[2], AND BENJAMIN A. ROBINSON[3]

Abstract

We consider Talagrand-type transportation inequalities for the law of Brownian motion on Carnot groups. An important example is the lift of standard Brownian motion to the Brownian rough path. We present a direct proof on enhanced path space, which also yields equality when restricting to adapted couplings in the transport problem. Moreover, we prove a Talagrand inequality for the heat kernel measure on Carnot groups and deduce the inequality for the law of Brownian motion on Carnot groups via a bottom-up argument. Our study of this enhanced Wiener measure contributes to a longstanding programme to extend key properties of Wiener measure to the non-commutative setting of the enhanced Wiener measure, which is of central importance in Lyons' rough path theory. With a non-commutative sub-Riemannian state space, we observe phenomena that differ from the Euclidean case. In particular, while a top-down projection argument recovers Talagrand's inequality on Euclidean space from the corresponding inequality on the path space, such a projection argument breaks down in the Carnot group setting. We further study a Riemannian approximation of the Heisenberg group, in which case the failure of the top-down projection can be partially overcome. Finally, we show that the cost function used in the Talagrand inequality is a natural choice, in that it arises as a limit of discretised costs in the sense of Γ-convergence.

Contents

[1]Technische Universität Berlin, Berlin, Germany
[2]Weierstrass-Institut für Angewandte Analysis und Stochastik, Berlin, Germany
[3]Department of Statistics, University of Klagenfurt, Austria
E-mail addresses: friz@math.tu-berlin.de, kremp@wias-berlin.de, artnoage@gmail.com, liero@wias.berlin.de, benjamin.robinson@aau.at.

Date: February 6, 2026.
2020 *Mathematics Subject Classification.* 39B62, 49Q22, 53C17, 60E15, 60L20.
Key words and phrases. Carnot groups, log-Sobolev inequality, optimal transport, rough paths, Talagrand inequality, transportation inequalities.

1. Introduction

Let $\mu \in \mathcal{P}(E)$ be a Borel probability measure on a Polish space E. Given a measurable cost function $c\colon E \times E \to [0,\infty]$, we say that μ satisfies Talagrand's $\mathcal{T}_2$ transport inequality with constant $\alpha > 0$, and write $\mu \in \mathcal{T}_2(E, c, \alpha)$, if for every $\nu \in \mathcal{P}(E)$ it holds that

$$\mathrm{T}^2_{c,2}(\mu,\nu) \coloneqq \inf_{\lambda\in\Pi(\mu,\nu)} \int_{E\times E} c^2(x,y)\,\mathrm{d}\lambda(x,y) \;\le\; \frac{2}{\alpha}\,H(\nu\|\mu),$$

where $\Pi(\mu,\nu)$ denotes the set of couplings between μ and ν and $H(\nu\|\mu)$ is the relative entropy of ν with respect to μ. If the cost is induced by a metric d on E, that is $c(x,y) = d(x,y)$, the above definition reduces to the classical 2-Wasserstein formulation,

$$W_2^2(\mu,\nu) \;\le\; \frac{2}{\alpha}\,H(\nu\|\mu), \qquad \forall \nu \in \mathcal{P}(E).$$

We note, without going into details, that there is an important connection to concentration of measure and the log-Sobolev inequalities by results from [OV00]. Talagrand [Tal96] first proved a $\mathcal{T}_2$ inequality for the standard Gaussian measure on $\mathbb{R}^d$ with Euclidean cost. A $\mathcal{T}_2$ inequality for $\mathbb{R}^d$-valued Brownian motion with a cost given in terms of the Cameron–Martin distance first appeared in [FÜ02]. Later [Leh13] gave a similar proof, using the intrinsic drift from [Föl86, Föl88] and Girsanov's theorem to prove the $\mathcal{T}_2$ inequality directly on Wiener space. Alternative proofs using Girsanov's theorem also appeared in [DGW04] and [FÜ04]. On the other hand, the $\mathcal{T}_2$ inequality on Wiener space can also be derived as a consequence of the Gaussian product case. In fact, [Tal96] already considered the infinite Gaussian product case; cf. [Rie17] and reference therein for explicit constructions. This so-called *bottom-up* approach uses the tensorisation property of the $\mathcal{T}_2$ inequality and a truncated expansion of the Brownian motion. As observed in [Leh13] and [Föl22], one can also recover Talagrand's $\mathcal{T}_2$ inequality on $\mathbb{R}^d$ from the $\mathcal{T}_2$ inequality on path space by considering a Brownian bridge. This gives a so-called *top-down* approach to Talagrand's $\mathcal{T}_2$ inequality. In this paper, a first study connecting aspects of optimal transport with rough analysis, we investigate the validity of the $\mathcal{T}_2$ inequality, as well as the bottom-up and top-down approaches, when $\mathbb{R}^d$ is replaced by a certain Carnot group.

The advent of rough path theory (see, e.g. [Lyo98, FV10]) has highlighted the fundamental importance of (d-dimensional) Brownian motion B lifted to the free step-2 nilpotent group (over $\mathbb{R}^d$), which is an example of Brownian motion with values in a step-2 Carnot group $\mathbb{G}$. Denoted by

$$\boldsymbol{B}_t = \Big(B_t,\ \mathrm{Anti}\Big(\int_0^t B_s \otimes \mathrm{d}B_s\Big)\Big),$$

this process is also known as horizontal Brownian motion, enhanced Brownian motion, or Brownian rough path, depending on authors and context. When $d = 2$, the relevant group is nothing but the classical $(2+1)$-dimensional Heisenberg group $\mathbb{H} \cong \mathbb{R}^3$ with group law

$$\big((x, y, z), (x', y', z')\big) \mapsto \big(x + x', y + y', z + z' + (xy' - x'y)/2\big).$$

Though not directly related to this work, we note that the interplay of optimal transport and Heisenberg groups was pioneered in [AR04]; see also [AS20] for recent work in the context of Carnot groups.

Let us agree on some notation. Unless otherwise stated, $\mu = \mathrm{Law}(B)$ denotes Wiener measure on $\Omega = C_0([0, T], \mathbb{R}^d)$, with Gaussian unit time marginal $\mu_1 = \mathcal{N}(0, I_d)$. Similarly, call $\boldsymbol{\mu} = \mathrm{Law}(\boldsymbol{B})$ the *enhanced Wiener measure* on

$$\boldsymbol{\Omega}_{\mathbb{G}} = C_0([0, T], \mathbb{G}),$$

with (non-Gaussian) unit time marginal $\boldsymbol{\mu}_1$, which we call the *heat kernel measure* on $\mathbb{G}$. Over the last 20 years, starting with [LQZ02], numerous properties of Wiener measure (including sample path regularity, Cameron–Martin shifts, Schilder's large deviations, Stroock–Varadhan support theorem) have been extended from μ to $\boldsymbol{\mu}$, with significant benefits to stochastic analysis (see, e.g. [FV10, Lyo14, FH20] and references therein). See also [CF26] for an abstract view. This naturally raises the question of whether Talagrand's $\mathcal{T}_2$ inequality for Gaussian measures (respectively, Wiener measure) extends to heat kernel measures (respectively, enhanced Wiener measure) on $\mathbb{G}$, and to what extent the bottom-up and top-down approaches remain valid. In this article, we provide a reasonably complete answer to these questions.

We prove the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$ with the cost function $C_{\mathcal{H}}$ on $\boldsymbol{\Omega}_{\mathbb{G}}$,

$$C_{\mathcal{H}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) := \begin{cases} \|h\|_{\mathcal{H}}, & \text{if } \overline{\boldsymbol{\omega}} = T_h \boldsymbol{\omega}, \text{ for some } h \in \mathcal{H}, \\ +\infty, & \text{otherwise,} \end{cases}$$

where $\mathcal{H}$ is the Cameron–Martin space of μ and T_h is (essentially[1]) the translation (or shift) operator known from rough path theory [FV10]. We give multiple strategies to prove the $\mathcal{T}_2$ inequality for our cost $C_{\mathcal{H}}$ on $\boldsymbol{\Omega}_{\mathbb{G}}$, offering both a bottom-up strategy, as well as a direct approach via an application of a contraction principle [DGW04, Rie17] or a lifting of the result from [Leh13] to the Carnot group setting.

The bottom-up approach consists of showing a $\mathcal{T}_2$ inequality for the heat kernel measure $\boldsymbol{\mu}_1$ on the Carnot group and inferring the result for the enhanced Wiener measure $\boldsymbol{\mu}$ by using the tensorisation property of the Talagrand inequality. Our approach is to discretise the Brownian motion in time, rather than to consider an expansion as in [Rie17, Föl22]. By an Otto–Villani argument [OV00, GL13], the $\mathcal{T}_2$ inequality for the heat kernel measure on a Carnot group follows from a log-Sobolev inequality. The latter is only partially available: from [Li06, Eld10] we have certain heat semigroup estimates on the Heisenberg (and so-called H-type) groups, which imply the required log-Sobolev inequalities. Given the correct heat semigroup estimate, our proof does not rely on an H-type setting and holds true for general step-2 Carnot groups.

We further prove the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$ directly on the path space, via two different approaches. First, we apply a contraction principle to the lift of a standard Brownian motion to deduce the result from the $\mathcal{T}_2$ inequality for Wiener measure. This approach also extends to the lifts of more general Gaussian

[1] Contrary to the standard rough path setting, we deal here with general step-2 Carnot groups.

processes; cf. [FV10, Chapter 15]. Alternatively, we exploit Föllmer's intrinsic drift from [Föl86, Föl88] to prove the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$, following the strategy of [Leh13]. The latter approach gives additional insights into so-called *adapted* transport inequalities. As noted in [Ald81, Las18, BBBE20a, BBBE20b], in the case of optimal transport problems involving laws of stochastic processes, it is desirable to consider *adapted* couplings rather than general couplings between the laws. We identify the optimal adapted transport plan for the cost $C_{\mathcal{H}}$ and show that equality holds in the $\mathcal{T}_2$ inequality when restricting to adapted couplings.

A key difference from the Euclidean setting is that the top-down approach fails in the non-commutative sub-Riemannian setting of Carnot groups. Indeed, in the Euclidean setting, given the $\mathcal{T}_2$ inequality for Wiener measure μ, a contraction principle can be applied to deduce the $\mathcal{T}_2$ inequality for μ_1. However, for the enhanced Wiener measure $\boldsymbol{\mu}$, this contraction principle argument breaks down and we cannot deduce the $\mathcal{T}_2$ inequality for the heat kernel measure $\boldsymbol{\mu}_1$ from the corresponding inequality for $\boldsymbol{\mu}$. Considering a Riemannian approximation (still non-commutative) to the Carnot group, we find that the validity of the top-down approach is partially recovered. Given the $\mathcal{T}_2$ inequality for the law $\boldsymbol{\mu}^\varepsilon$ of Brownian motion on the approximating Riemannian manifold, the contraction principle implies that $\boldsymbol{\mu}_1^\varepsilon$ satisfies a $\mathcal{T}_p$ inequality for $p \in [1,2)$, but not for $p = 2$.

We remark that our cost $C_{\mathcal{H}}$ differs from the one considered in [Rie17, Corollary 1.4], which is defined in terms of the Cameron–Martin norm of the difference of the path in the group projected onto its first component, and which turns out to be suboptimal (see the discussion in Section 2.3). Our cost $C_{\mathcal{H}}$ is a natural choice in the following sense: $C_{\mathcal{H}}$ can be obtained as the variational limit (more precisely, the Γ-limit; see Section 6.4) of "finite-dimensional costs" C_n that arise in our bottom-up approach:

$$C_n^2(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) = 2^n \sum_{k=1}^{2^n} d_{\mathrm{CC}}^2(\boldsymbol{\omega}_{t_{k-1}^n, t_k^n}, \overline{\boldsymbol{\omega}}_{t_{k-1}^n, t_k^n}), \quad \boldsymbol{\omega}, \overline{\boldsymbol{\omega}} \in \boldsymbol{\Omega}_{\mathbb{G}},$$

where d_{CC} denotes the Carnot–Caratheodory metric on $\mathbb{G}$. We prove in Section 6.4 that the Γ-convergence of the cost functions C_n also leads to the Γ-convergence of the optimal transport costs $\mathrm{T}_{C_n,2}(\boldsymbol{\mu}, \cdot)$ to $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \cdot)$.

Note that, while our direct approach to proving the $\mathcal{T}_2$ inequality gives an elegant and short proof that holds in greater generality, the bottom-up approach, and in particular the Γ-convergence, yields clear information on the choice of the most natural cost function.

For the reader's convenience we summarise our findings as concise statements.

Theorem 1.1 (*Direct approach*, cf. Theorem 3.3, and extensions in Section 4)**.** *The measure $\boldsymbol{\mu}$ on $\boldsymbol{\Omega}_{\mathbb{G}}$ satisfies the $\mathcal{T}_2$ inequality $\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{G}}, C_{\mathcal{H}}, 1)$.*

Theorem 1.2 (*Bottom-up*, cf. Theorem 6.6)**.** *Suppose that there exists $\alpha > 0$ such that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$. Then $\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{G}}, C_{\mathcal{H}}, \alpha)$.*

Theorem 1.3 ($\mathcal{T}_2$ *on group*, cf. Theorem 5.3)**.** *Let $\mathbb{G}$ be an H-type group. Then there exists $\alpha > 0$ such that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$.*

Theorem 1.4 (*Adapted couplings*, cf. Theorem 3.3)**.** *Let $\boldsymbol{\nu}$ be a probability measure on $\boldsymbol{\Omega}_{\mathbb{G}}$ with $\boldsymbol{\nu} \ll \boldsymbol{\mu}$. Then the optimal adapted coupling between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ is given explicitly and*

$$\mathrm{T}^{\mathrm{ad}}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \boldsymbol{\nu})^2 = 2H(\boldsymbol{\nu} \| \boldsymbol{\mu}).$$

Theorem 1.5 (*Cost approximation*, cf. Corollary 6.22 and Theorem 6.28)**.**

(i) Pointwise convergence $C_n \to C_{\mathcal{H}}$ fails (by example),

(ii) Γ-convergence[2] $C_n \xrightarrow{\Gamma} C_{\mathcal{H}}$ holds with respect to the uniform topology on $\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$,

(iii) $\mathrm{T}_{C_n,2}(\boldsymbol{\mu}, \cdot) \xrightarrow{\Gamma} \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \cdot)$ with respect to the weak topology on $\mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$.

Theorem 1.6 (*Top-down – validity vs. failure*)**.** *A contraction principle*

(i) gives the implication $\mu \in \mathcal{T}_2(\Omega, \|. - .\|_{\mathcal{H}}, \alpha) \implies \mu_1 \in \mathcal{T}_2(\mathbb{R}^d, |. - .|, \alpha)$,

(ii) does not give $\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}, C_{\mathcal{H}}, \alpha) \implies \boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha')$, no matter $\alpha, \alpha' > 0$,

(iii) gives a weak implication $\boldsymbol{\mu}^\varepsilon \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{H}_\varepsilon}, C^\varepsilon_{\mathcal{H}}, \alpha) \implies \boldsymbol{\mu}^\varepsilon_1 \in \mathcal{T}_p(\mathbb{H}_\varepsilon, d_\varepsilon, \widetilde{\alpha}(\varepsilon, p))$, for $p \in [1, 2)$, where $(\mathbb{H}_\varepsilon, d_\varepsilon)$ is a Riemannian approximation to the $(2+1)$-dimensional Heisenberg group.

The paper is structured as follows. Section 2 contains preliminary results on Talagrand inequalities and introduces our setting of step-2 Carnot groups. In Section 3, we present a direct approach to proving the $\mathcal{T}_2$ inequality for the law of Brownian motion on a Carnot group via Föllmer's intrinsic drift, and we show equality for the case of adapted transport plans. In Section 4, we prove the $\mathcal{T}_2$ inequality for general Gaussian rough paths by a direct approach using a contraction principle. Section 5 studies the Talagrand inequality for the heat kernel measure $\boldsymbol{\mu}_1$ on Carnot groups, as well as its connection to log-Sobolev inequalities and heat semigroup estimates. Section 6.1 presents a bottom-up approach to proving the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$ as a consequence of the results of Section 5. In Section 6.2, we show by example that we cannot project the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$ down to a $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}_1$, and that the cost functions C_n blow up pointwise. In Section 6.3, we study a Riemannian approximation of the Heisenberg group, for which we can partially overcome the issues of the cost blow-up and failure of projection. In Section 6.4, we prove the Γ-convergence of the costs C_n to $C_{\mathcal{H}}$. We conclude in Section 7 by commenting on the extension of our results to higher order Carnot groups.

2. Setting

In this section, we first collect relevant definitions and results related to Talagrand inequalities. Next, we define our setting of step-2 Carnot groups and introduce a Brownian motion with paths in a Carnot group, as well as the lift and shift operation on paths. Using the shift operator, we define a suitable cost function $C_{\mathcal{H}}$, which appears in our $\mathcal{T}_2$ inequality.

2.1. Preliminaries on Talagrand inequalities.

Given two Borel probability measures μ, ν on a Polish space E, let $\Pi(\mu, \nu)$ denote the set of probability measures on $E \times E$ with marginals μ, ν. Such measures are called couplings (or transport plans). The relative entropy of ν with respect to μ is defined as

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

[2] Recall that Γ-convergence is a natural notion of convergence from the theory of calculus of variations for sequences of functionals, which guarantees the convergence of minimisers and minima.

Definition 2.1. Let E be a Polish space and let $c\colon E\times E\to[0,\infty]$ be a measurable function. We say that a Borel probability measure μ satisfies the *cost-information inequality* on E with cost c, parameter $\alpha>0$, and exponent $p\in[1,\infty)$ if, for any Borel probability measure ν on E,

$$\mathrm{T}_{c,p}(\mu,\nu)\le\sqrt{\frac{2}{\alpha}H(\nu\|\mu)},\quad\text{where}\quad \mathrm{T}_{c,p}(\mu,\nu):=\Big(\inf_{\lambda\in\Pi(\mu,\nu)}\int\int c^p(x,y)\,\mathrm{d}\lambda(x,y)\Big)^{\frac{1}{p}}.$$

We write $\mu\in\mathcal{T}_p(E,c,\alpha)$ and say that μ satisfies a $\mathcal{T}_p$ inequality.[3]

In particular, we are interested in the case of $p=2$. On $\mathbb{R}^d$, Talagrand [Tal96] proved that the standard Gaussian measure satisfies a $\mathcal{T}_2$ inequality with Euclidean cost. Talagrand's result has since been lifted to the Wiener measure on path space.

Let B denote a standard Brownian motion on $\mathbb{R}^d$, let $\mu_1=\mathrm{Law}(B_1)$ denote the standard Gaussian measure on $\mathbb{R}^d$, and let $\mu=\mathrm{Law}(B)$ denote the Wiener measure on $\Omega:=C_0([0,1],\mathbb{R}^d)$. The Cameron–Martin space for μ is defined as

$$\mathcal{H}:=\{\,h\colon[0,1]\to\mathbb{R}^d\text{ absolutely continuous}:\dot h\in L^2,\ h_0=0\,\}=W_0^{1,2}([0,1],\mathbb{R}^d),\tag{2.1}$$

and the Cameron–Martin norm $\|\cdot\|_{\mathcal{H}}$ is defined by $\|h\|_{\mathcal{H}}^2=\int_0^1|\dot h_t|^2\,\mathrm{d}t$, for $h\in\mathcal{H}$. Throughout the text, for $p\in[1,\infty]$, $W^{1,p}([0,1],\mathbb{R}^d)$ denotes the usual Sobolev space, and $W_0^{1,p}([0,1],\mathbb{R}^d)$ the subspace such that $x_0=0$ for $x\in W_0^{1,p}([0,1],\mathbb{R}^d)$.

Define the Cameron–Martin cost $c_{\mathcal{H}}\colon\Omega\times\Omega\to[0,\infty]$ by

$$c_{\mathcal{H}}(x,y):=\begin{cases}\|y-x\|_{\mathcal{H}}, & y-x\in\mathcal{H},\\ +\infty & \text{otherwise.}\end{cases}\tag{2.2}$$

Then the Wiener measure μ satisfies the $\mathcal{T}_2$ inequality (cf. [FÜ02, DGW04, FÜ04, Leh13]):

$$\mu\in\mathcal{T}_2(\Omega,c_{\mathcal{H}},1).\tag{2.3}$$

Contraction principle. The following contraction principle for $\mathcal{T}_p$ inequalities is a special case of [Rie17, Lemma 4.1].

Lemma 2.2. *Let (E,d), (S,ρ) be metric spaces, with (E,d) a Polish space, let $c\colon E\times E\to[0,\infty]$ and $\widetilde{c}\colon S\times S\to[0,\infty]$ be Borel-measurable functions, and let η be a Borel probability measure on E. Let $\psi\colon E\to S$ and $L\colon E\to[0,\infty]$ be measurable functions such that*

$$\widetilde{c}(\psi(x),\psi(\overline{x}))\le L(x)c(x,\overline{x}),$$

for all $x,\overline{x}\in E_0$, where $E_0\subseteq E$ satisfies $\eta(E_0)=1$.

Suppose that $\eta\in\mathcal{T}_2(E,c,\alpha)$, for some $\alpha\in(0,\infty)$. Then, for any $p\in[1,2]$ such that $L\in L^q(\eta)$ for $q=\frac{2p}{2-p}\in[2,\infty]$, we have $\psi_\sharp\eta\in\mathcal{T}_p(S,\widetilde{c},\alpha\|L\|^{-2}_{L^q(\eta)})$.

Remark 2.3. In particular, the contraction principle in Lemma 2.2 allows us to upgrade the topology used in Theorem 1.1 from the uniform topology to the β-Hölder topology for $\beta\in(\frac13,\frac12)$ (cf. [FV10, Section 8] for the definition of this topology). Indeed, for $E=C_0([0,1],\mathbb{G})$ and $\boldsymbol{\mu}\in\mathcal{P}_2(E)$ the law of Brownian

[3]Since c^p is just another instance of a measurable function on $E\times E$, there is no loss of generality in taking $p=1$. However, we find this definition useful later in the paper.

motion on $\mathbb{G}$, we have that $\boldsymbol{\mu}(\widetilde{E}) = 1$, where $\widetilde{E} = C_0^{\beta}([0,1], \mathbb{G})$ for some $\beta \in (\frac{1}{3}, \frac{1}{2})$. Thus, defining $\widetilde{c} = c|_{\widetilde{E}\times\widetilde{E}}$, the result of Theorem 1.1 that $\boldsymbol{\mu} \in \mathcal{T}_2(E, c, \alpha)$ extends to $\boldsymbol{\mu} \in \mathcal{T}_2(\widetilde{E}, \widetilde{c}, \alpha)$ by Lemma 2.2.

We remark that such a direct upgrade of the topology is not observed in other settings. For example, showing that a large deviation principle can be lifted from the uniform topology to the Hölder topology is significantly more involved; see [FV05, Theorem 39], whose proof is based on the inverse contraction principle for large deviations [DZ10, Theorem 4.2.4].

Adapted $\mathcal{T}_p$ inequalities. For a metric space (S, d), the p-Wasserstein distance $\mathrm{T}_{d,p}$ metrises the weak topology on $\mathcal{P}_p(S)$. When elements of S should be regarded as stochastic processes, however, this topology is not sufficient to capture the flow of information encoded in the filtrations associated to the processes. The *adapted weak topology* and *adapted Wasserstein distance* have been shown to be more suitable; see, e.g. [Ald81, Las18, BBBE20a, BBBE20b]. The adapted Wasserstein distance is a special case of the adapted (also called bicausal) optimal transport problem, defined as follows.

Definition 2.4. Let E be a Polish space and $\mu, \nu \in \mathcal{P}(C([0,1], E))$. Let $\lambda \in \Pi(\mu, \nu)$ and let X, Y be $C([0,1], E)$-valued random variables with $\lambda = \mathrm{Law}(X, Y)$. Write $\mathcal{F}^X$ (resp. $\mathcal{F}^Y$) for the completion of the natural filtration of X (resp. Y) with respect to μ (resp. ν). We say that λ is an *adapted coupling* if the following conditional independence holds under λ: for all $t \in [0,1]$,

$$\mathcal{F}^Y_t \text{ is independent of } \mathcal{F}^X_1 \text{ given } \mathcal{F}^X_t \quad \text{and} \quad \mathcal{F}^X_t \text{ is independent of } \mathcal{F}^Y_1 \text{ given } \mathcal{F}^Y_t.$$

We denote the set of all such couplings by $\Pi_{\mathrm{ad}}(\mu, \nu)$. For a measurable function $c\colon C([0,1], E) \times C([0,1], E) \to [0, \infty]$, define the *adapted optimal transport problem*

$$\mathrm{T}^{\mathrm{ad}}_{c,p}(\mu, \nu) := \Big(\inf_{\lambda \in \Pi_{\mathrm{ad}}(\mu,\nu)} \int \int c^p(x, y) \, \mathrm{d}\lambda(x, y) \Big)^{\frac{1}{p}}.$$

We say that $\mu \in \mathcal{P}(C([0,1], E))$ satisfies an *adapted T_p inequality* for some $p \in [1, \infty)$ if there exists $\alpha > 0$ such that

$$\mathrm{T}^{\mathrm{ad}}_{c,p}(\mu, \nu) \leq \sqrt{\frac{2}{\alpha} H(\nu \| \mu)}.$$

In this adapted setting, [Las18, Lemma 5] and [Föl22, Theorem 3] show that Wiener measure satisfies an adapted $\mathcal{T}_2$ inequality with $\alpha = 1$ and that equality holds; i.e.

$$\mathrm{T}^{\mathrm{ad}}_{c_{\mathcal{H}},2}(\mu, \nu) = \sqrt{2H(\nu \| \mu)}. \tag{2.4}$$

Remark 2.5. For continuous-time stochastic processes, [BBP+25] give an alternative definition of the adapted optimal transport problem and adapted Wasserstein distance that has additional desirable topological properties. The value of this problem is defined such that it lies between $\mathrm{T}_{c,p}$ and $\mathrm{T}^{\mathrm{ad}}_{c,p}$. Thus, it is immediate that an adapted $\mathcal{T}_2$ inequality still holds in this setting. However, equality has not been studied in this case, and we leave this to future work, choosing to focus on the definition given in Definition 2.4 in the present paper.

Remark 2.6. For discrete-time processes taking values in some Polish space E with n time steps, one can also consider their laws, which are probability measures on E^n, and define an adapted optimal transport problem analogously to Definition 2.4. In this setting, [Par26, Corollary 1.8] shows that the $\mathcal{T}_1$ inequality is equivalent to its adapted counterpart. Moreover, [Par26, Corollary 1.9] shows that, for probability measures with finite exponential moment, an adapted $\mathcal{T}_p$ inequality holds for all $p > 1$, with constant given explicitly in terms of the exponential moment and number of time steps, thus extending the results of [BV05] to the adapted setting. For a standard Gaussian on $\mathbb{R}^n$, [BBLZ17, Proposition 5.10] prove an adapted $\mathcal{T}_2$ inequality using a dynamic programming argument. As noted in [BBLZ17, Remark 5.11], equality cannot generally be expected in the discrete-time setting.

2.2. **Step-2 Carnot groups.** Let $\mathbb{G}$ be a step-2 Carnot group, i.e. a connected, simply connected nilpotent Lie group whose Lie algebra $\mathfrak{g}$ of left-invariant vector fields has dimension $m = d_1 + d_2$ and admits a stratification $\mathfrak{g} = \mathcal{V}_1 \oplus \mathcal{V}_2$ with $\mathcal{V}_2 = [\mathcal{V}_1, \mathcal{V}_1]$, $[\mathcal{V}_1, \mathcal{V}_2] = \{0\}$. Fix an adapted basis $(V_1, \dots, V_m)$ such that $(V_1, \dots, V_{d_1})$ is a basis of $\mathcal{V}_1$. Using exponential coordinates, we can and will identify $\mathbb{G}$ with $\mathbb{R}^m$,

$$\mathbb{G} \ni \boldsymbol{x} = (x_1, \dots, x_{d_1}, x_{d_1+1}, \dots, x_m) = (x^{(1)}, x^{(2)}) \in \mathbb{R}^{d_1} \oplus \mathbb{R}^{d_2} \cong \mathbb{R}^m,$$

with group law in Baker–Campbell–Hausdorff form,

$$(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x}\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{y} + \frac{1}{2}[\boldsymbol{x}, \boldsymbol{y}]. \tag{2.5}$$

It is not restrictive to assume that $V_i(0) = e_i$, the canonical basis vectors of $\mathbb{R}^m$. For $\boldsymbol{x} \in \mathbb{G}$, let $\ell_{\boldsymbol{x}} \colon \mathbb{G} \to \mathbb{G}$ denote the left multiplication map defined by $\ell_{\boldsymbol{x}}\boldsymbol{y} = \boldsymbol{x}\boldsymbol{y}$, for $\boldsymbol{y} \in \mathbb{G}$, and let $\mathrm{d}\ell_{\boldsymbol{x}} \colon T\mathbb{G} \to T\mathbb{G}$ denote its differential. By left invariance, $V_i(\boldsymbol{x}) = \mathrm{d}\ell_{\boldsymbol{x}} e_i$, $i = 1, \dots, m$, $\boldsymbol{x} \in \mathbb{G}$.

Endow $\mathfrak{g}$ with a left-invariant metric $\langle \cdot, \cdot \rangle \colon \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ that makes the V_i orthonormal. Define the *structure constants* $w_{ij} \in \mathbb{R}^{d_2}$, for $i, j \in \{1, \dots, d_1\}$, by $w_{ij}^k \coloneqq \langle [V_i, V_j], V_k \rangle = -w_{ji}^k$, for $k \in \{d_1{+}1, \dots, m\}$; cf. [BLU07, Section 3.2]. The group law in (2.5) can then be written as

$$(\boldsymbol{x}, \boldsymbol{y}) = ((x^{(1)}, x^{(2)}),\, (y^{(1)}, y^{(2)})) \mapsto \boldsymbol{x}\boldsymbol{y} = \Big(x^{(1)} + y^{(1)},\, x^{(2)} + y^{(2)} + \frac{1}{2}\sum_{i<j} w_{ij}(x_i^{(1)} y_j^{(1)} - x_j^{(1)} y_i^{(1)})\Big).$$

For notational brevity, we introduce the operator $\mathsf{W} \colon \mathbb{R}^{d_1 \times d_1} \to \mathbb{R}^{d_2}$ given in terms of the structure constants by

$$\mathsf{W}A = \sum_{i,j=1}^{d_1} w_{ij} A_{ij} = \frac{1}{2}\sum_{i,j=1}^{d_1} w_{ij}(A_{ij} - A_{ji}) = \sum_{i<j} w_{ij}(A_{ij} - A_{ji}), \quad A \in \mathbb{R}^{d_1 \times d_1}.$$

With this definition, we can rewrite the group law for $\boldsymbol{x} = (x^{(1)}, x^{(2)})$, $\boldsymbol{y} = (y^{(1)}, y^{(2)})$ as

$$\boldsymbol{x}\boldsymbol{y} = \Big(x^{(1)} + y^{(1)},\, x^{(2)} + y^{(2)} + \frac{1}{2}\mathsf{W}(x^{(1)} \otimes y^{(1)})\Big).$$

Let $\Delta_{\mathbb{G}} \coloneqq \frac{1}{2}\sum_{i=1}^{d_1} V_i^2$ denote the sub-Laplacian on $\mathbb{G}$ and define the horizontal gradient $\nabla_{\mathbb{G}}$ by its action

$$\nabla_{\mathbb{G}} f \coloneqq \sum_{i=1}^{d_1} (V_i f) V_i \in \mathcal{V}_1, \quad \text{for } f \colon \mathbb{G} \to \mathbb{R}.$$

Let $H\mathbb{G} \subset T\mathbb{G}$ be the horizontal tangent bundle of the group $\mathbb{G}$, i.e. the left-invariant sub-bundle of the tangent bundle $T\mathbb{G}$ such that $H_{\boldsymbol{e}}\mathbb{G} = \{\, V(0) : V \in \mathcal{V}_1 \,\}$, where $\boldsymbol{e}$ is the identity element of $\mathbb{G}$. For $i \in \{1, 2\}$, define the projection operator $\pi_i \colon \mathbb{G} \to \mathbb{R}^{d_i}$ by

$$\pi_i(x^{(1)}, x^{(2)}) = x^{(i)}. \tag{2.6}$$

Dilation on $\mathbb{G}$ by a factor $s > 0$ takes the form

$$\delta_s((x^{(1)}, x^{(2)})) \coloneqq (sx^{(1)}, s^2 x^{(2)}). \tag{2.7}$$

The Haar measure on $\mathbb{G}$ coincides with Lebesgue measure $\mathcal{L}^m$ on $\mathbb{R}^m$. For measurable $E \subseteq \mathbb{R}^m$, we have

$$\mathcal{L}^m(\delta_s E) = s^Q \mathcal{L}^m(E), \tag{2.8}$$

where $Q = d_1 + 2d_2$ is called the *homogeneous dimension* of $\mathbb{G}$.

We endow $\mathbb{G}$ with the Carnot–Carathéodory structure induced by $H\mathbb{G}$, as follows. An absolutely continuous curve $\gamma\colon [0,1] \to \mathbb{G}$ is called horizontal if $\dot\gamma_t \in H_{\gamma_t}\mathbb{G}$ for almost every $t \in [0,1]$. The *Carnot–Carathéodory distance* between $x, y \in \mathbb{G}$ is then defined as

$$d_{\mathrm{CC}}(x,y) = \inf\left\{ \int_0^1 |\dot\gamma_t| \,\mathrm{d}t : \gamma \text{ horizontal}, \gamma_0 = x, \gamma_1 = y \right\}, \tag{2.9}$$

where $|\cdot| = \sqrt{\langle \cdot, \cdot \rangle}$.

We remark that horizontal paths necessarily satisfy, for almost every $t \in [0,1]$,

$$\dot\gamma_t = \sum_{i=1}^{d_1} V_i(\gamma_t)\langle \dot\gamma_t, V_i(\gamma_t)\rangle =: \sum_{i=1}^{d_1} V_i(\gamma_t)\dot h_t^i, \tag{2.10}$$

and hence are in one-to-one correspondence with absolutely continuous $h \in \mathrm{AC}([0,1], \mathbb{R}^{d_1})$. We have that $h = \pi_1\gamma$ for the first-level projection π_1 from (2.6).

Definition 2.7 (Canonical lift)**.** The *canonical lift* $\Psi\colon \mathrm{AC}([0,1], \mathbb{R}^{d_1}) \to C([0,1], \mathbb{G})$ is defined by $\Psi(h) \coloneqq \gamma$, where $\gamma \in C([0,1], \mathbb{G})$ and $h \in \mathrm{AC}([0,1], \mathbb{R}^{d_1})$ are related by (2.10). Explicitly, we have

$$\dot\gamma_t^{(1)} = \dot h_t, \quad \dot\gamma_t^{(2)} = \frac{1}{2}\sum_{i,j} w_{ij} h_t^i \dot h_t^j = \frac{1}{2}\mathsf{W}\big(h_t \otimes \dot h_t\big), \quad \text{for } t \in [0,1].$$

By the Chow–Rashevskii theorem, d_{CC} is in fact a distance, which is also left-invariant and homogeneous with respect to the dilations defined in (2.7). The metric space $(\mathbb{G}, d_{\mathrm{CC}})$ is a Polish and geodesic space (see, e.g. [AS20, Section 2.4]). We let $|\cdot|_{\mathbb{G}}$ denote the norm induced by d_{CC} on $\mathbb{G}$. One can also equip $\mathbb{G}$ with the *gauge distance* d_g defined by

$$d_g(x,y) = |(y^{-1}x)^{(1)}| + |(y^{-1}x)^{(2)}|^{\frac{1}{2}},$$

for $x, y \in \mathbb{G}$. All homogeneous norms on $\mathbb{G}$ are equivalent. In particular, there exists a constant $\kappa \in (0,\infty)$ such that

$$\frac{1}{\kappa} d_g(x,y) \le d_{\mathrm{CC}}(x,y) \le \kappa\, d_g(x,y); \tag{2.11}$$

see, e.g. [BLU07, Proposition 5.1.4].

Remark 2.8 (Metric derivative)**.** The metric derivative of a curve $\gamma\colon [0,1] \to \mathbb{G}$ at $t \in [0,1]$ is defined by

$$|\dot\gamma_t|_{d_{\mathrm{CC}}} \coloneqq \lim_{s \to t} \frac{d_{\mathrm{CC}}(\gamma(s), \gamma(t))}{|s-t|}.$$

If γ is absolutely continuous, the metric derivative exists for almost every $t \in [0,1]$, and $|\dot\gamma_t|_{d_{\mathrm{CC}}} = |\dot\gamma_t|$; see [Mon01, Theorem 1.3.5]. Moreover, the metric derivative is minimal in the sense that $|\dot\gamma_t|_{d_{\mathrm{CC}}} \le m$ for all $m \in L^1([0,1])$ with $d_{\mathrm{CC}}(\gamma(s), \gamma(t)) \le \int_s^t m(r)\,\mathrm{d}r$, $0 \le s < t \le 1$; see [AGS08, Theorem 1.1.2].

Remark 2.9 (Free step-2 Carnot groups)**.** The free step-2 nilpotent case $\mathbb{G} = \mathbb{F}^{d_1,2}$ amounts to $\mathbb{G} \cong \mathbb{R}^{d_1} \oplus \mathfrak{so}(d_1)$ (after identifying the exterior algebra $\wedge^2\mathbb{R}^{d_1}$ with $\mathfrak{so}(d_1)$). The space $\mathfrak{so}(d_1)$ is spanned by $\{\, e_{[i,j]} : 1 \le i < j \le d_1 \,\}$, where $e_{[i,j]} \coloneqq \frac{1}{2}(e_i \otimes e_j - e_j \otimes e_i)$, and has dimension $d_2^* = d_1(d_1{-}1)/2$. Writing the bracket as $[e_i, e_j] = e_{[i,j]}$, the structure constants $w_{ij}^{[p,q]}$ reduce to Kronecker symbols.

All other step-2 Carnot groups can be seen as quotient groups of the free group, captured by $d_2 \le d_2^*$ and the structure constants. For instance, the $(2n+1)$-dimensional Heisenberg group $\mathbb{H}^n \cong \mathbb{R}^{2n} \oplus \mathbb{R}$ has $d_1 = 2n$, $d_2 = 1$, and $w_{1,2} = w_{3,4} = \cdots = w_{2n-1,2n} = 1$ (flip sign upon interchanging indices, zero otherwise). For $d_1 = 2$, we recover the familiar example $\mathbb{H} = \mathbb{H}^1 \cong \mathbb{F}^{2,2}$. Letting $(x, y, z) \in \mathbb{H}$ denote a canonical element of $\mathbb{H}$, the left-invariant vector fields are given by

$$V_1 = \partial_x + \tfrac{1}{2} y \partial_z, \quad V_2 = \partial_y - \tfrac{1}{2} x \partial_z, \quad V_3 = [V_1, V_2] = \partial_z.$$

Remark 2.10 (Heisenberg-type groups)**.** A special class of step-2 Carnot groups is the class of *Heisenberg-type groups*, or H-type groups for short, which enjoy additional properties. Most importantly for us, Talagrand inequalities are known to hold on H-type groups; see Section 5.3. We refer the interested reader to [BLU07, Chapter 18].

A step-2 Carnot group $\mathbb{G} \cong \mathbb{R}^{d_1} \oplus \mathbb{R}^{d_2}$ is an *H-type group* if, for each $z \in \mathbb{R}^{d_2}$, there exists a linear map $\mathsf{J}_{\mathbb{G}}(z) \colon \mathbb{R}^{d_1} \to \mathbb{R}^{d_1}$ such that

$$\mathsf{J}_{\mathbb{G}}(z)^2 = -|z|^2 \mathrm{id} \quad \text{and} \quad \big\langle \mathsf{W}(x \otimes y), z \big\rangle_{\mathbb{R}^{d_2}} = \langle \mathsf{J}_{\mathbb{G}}(z) x, y \rangle_{\mathbb{R}^{d_1}} \quad \forall\, x, y \in \mathbb{R}^{d_1}.$$

Note that, necessarily, $d_1 \in 2\mathbb{N}$ and $d_2 \le d_1/2$. As the name suggests, the Heisenberg group $\mathbb{H}^n \cong \mathbb{R}^{2n} \oplus \mathbb{R}$ is the canonical example of an H-type group, where the map $\mathsf{J}_{\mathbb{H}^n} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is given by

$$\mathsf{J}_{\mathbb{H}^n}(z) = \begin{pmatrix} 0 & -zI_n \\ zI_n & 0 \end{pmatrix}, \quad z \in \mathbb{R}.$$

Brownian motion on $\mathbb{G}$. Let $B = (B_t)_{t \in [0,1]}$ be a d_1-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \ge 0}, \mathbb{P})$. We define the Brownian motion $\boldsymbol{B}$ on $\mathbb{G}$ as the continuous $\mathbb{G}$-valued Markov process with generator $\Delta_{\mathbb{G}}$, that is obtained by solving the SDE

$$\mathrm{d}\boldsymbol{B}_t = \sum_{i=1}^{d_1} V^i(\boldsymbol{B}_t)\, \mathrm{d}B_t^i, \tag{2.12}$$

more explicitly written as

$$\mathrm{d}\boldsymbol{B}_t^{(1)} = \mathrm{d}B_t, \quad \mathrm{d}\boldsymbol{B}_t^{(2)} = \frac{1}{2} \mathsf{W}(B_t \otimes \mathrm{d}B_t).$$

Note that, since $B = (B_1, \ldots, B_{d_1})$ is a standard Brownian motion, there is no difference between Itô and Stratonovich integration here.

The Brownian motion $\boldsymbol{B}$ takes values in the space

$$\boldsymbol{\Omega}_{\mathbb{G}} \coloneqq C_0([0,1], \mathbb{G})$$

of continuous $\mathbb{G}$-valued paths started from the origin. We write $\boldsymbol{\mu} = \mathrm{Law}(\boldsymbol{B})$ and $\mu = \mathrm{Law}(B)$, and $\boldsymbol{\mu}_t = \mathrm{Law}(\boldsymbol{B}_t)$, $\mu_t = \mathrm{Law}(B_t)$, for $t > 0$. For a path $\boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbb{G}}$, let $\boldsymbol{\omega}_{s,t} = \boldsymbol{\omega}_s^{-1} \boldsymbol{\omega}_t$, $s \le t$, denote its increments.

We equip the space $\boldsymbol{\Omega}_{\mathbb{G}}$ with the uniform topology induced by the metric d_∞ defined by

$$d_\infty(\boldsymbol{\omega},\overline{\boldsymbol{\omega}}) \coloneqq \sup_{t\in[0,1]} d_{\mathrm{CC}}(\boldsymbol{\omega}_t,\overline{\boldsymbol{\omega}}_t). \tag{2.13}$$

Note that $(\boldsymbol{\Omega}_{\mathbb{G}}, d_\infty)$ is a Polish space.

By Hörmander's theorem, $\Delta_{\mathbb{G}}$ is a hypoelliptic operator, and so the associated heat kernel $\mathfrak{p}\colon (0,\infty)\times\mathbb{G}\to(0,\infty)$ is smooth [Hör67, Koh73, Hai11, BB15]. Note that, for all $t>0$, the density of $\boldsymbol{\mu}_t$ is $\mathfrak{p}_t\colon \mathbb{G}\to(0,\infty)$. We also define the heat semigroup $P_t = e^{t\Delta_{\mathbb{G}}}$, for $t>0$, by

$$P_t f(\boldsymbol{x}) = \int_{\mathbb{G}} f(\boldsymbol{x}\boldsymbol{y}^{-1})\mathfrak{p}_t(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_{\mathbb{G}} f(\boldsymbol{y})\mathfrak{p}_t(\boldsymbol{y}^{-1}\boldsymbol{x})\,\mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x}\in\mathbb{G}, \tag{2.14}$$

for any $f\in L^1(\boldsymbol{\mu}_t)$, with P_0 equal to the identity operator.

Shifting $\mathbb{G}$-valued paths. Let $\mathrm{AC}_0([0,1],\mathbb{R}^d)$ denote the space of absolutely continuous curves started from the origin, and recall the canonical lift $\Psi\colon \mathrm{AC}_0([0,1],\mathbb{R}^{d_1})\to\boldsymbol{\Omega}_{\mathbb{G}}$ from Definition 2.7. Since $(\mathbb{G}, d_{\mathrm{CC}})$ is a geodesic space, the following approximation lemma is immediate (cf. [FV10, Lemma 5.19, Theorem 7.32]).

Lemma 2.11 (Geodesic approximations)**.** *Every continuous $\mathbb{G}$-valued path $\boldsymbol{\omega}$ on $[0,T]$ is the uniform limit of absolutely continuous horizontal curves; i.e. $\boldsymbol{\omega}^n = \Psi(\omega^n)$, with $\omega^n\in\mathrm{AC}_0([0,1],\mathbb{R}^{d_1})$ and Ψ defined in Definition 2.7.*

We now extend the canonical lift to a lift map on the space $\Omega = C_0([0,1],\mathbb{R}^{d_1})$ of continuous curves started from the origin. Note that the geodesic approximation of $\boldsymbol{\omega} = (\boldsymbol{\omega}^{(1)},\boldsymbol{\omega}^{(2)})$ from Lemma 2.11 depends on both $\boldsymbol{\omega}^{(1)}$ and $\boldsymbol{\omega}^{(2)}$. Thus we also introduce an approximation based only on $\boldsymbol{\omega}^{(1)}$ in order to extend the canonical lift. For a continuous path $\omega\in\Omega$, let $\widehat{\omega}^n\in\mathrm{AC}_0([0,1],\mathbb{R}^{d_1})$ denote the piecewise-linear approximation of ω on the dyadic grid $(k2^{-n})_{k\in\{0,\dots,2^n\}}$, for $n\in\mathbb{N}$, and note that $\widehat{\omega}^n\to\omega$ with respect to the uniform topology on Ω.

Definition 2.12. (Lift) Extend the canonical lift $\Psi\colon \mathrm{AC}_0([0,1],\mathbb{R}^{d_1})\to\boldsymbol{\Omega}_{\mathbb{G}}$ to the *lift map* $\Psi\colon\Omega\to\boldsymbol{\Omega}_{\mathbb{G}}$ by

$$\Psi(\omega) \coloneqq \begin{cases} \lim_{n\to\infty}\Psi(\widehat{\omega}^n), & \text{if the limit exists,}\\ 0, & \text{otherwise,}\end{cases} \tag{2.15}$$

for $\omega\in\Omega$. Define the *domain* of Ψ as $\mathrm{Dom}(\Psi) \coloneqq \{\,\omega\in\Omega \;:\; \lim_{n\to\infty}\Psi(\widehat{\omega}^n) \text{ exists}\,\}\subset\Omega$.

The following statement holds by a minor modification to the proof of [FV10, Corollary 13.19] in the general step-2 Carnot setting.

Proposition 2.13. *Let B be a d_1-dimensional Brownian motion. Then $\boldsymbol{B}$ as defined in* (2.12) *satisfies $\boldsymbol{B} = \Psi(B) = \lim_{n\to\infty}\Psi(\widehat{B}^n)$ almost surely. In particular, it follows that $\mu(\mathrm{Dom}(\Psi)) = 1$.*

Having defined the lift Ψ for general curves $\omega\in C_0([0,1],\mathbb{R}^{d_1})$, we can now formulate the following lemma relating absolutely continuous measures $\boldsymbol{\nu}\ll\boldsymbol{\mu}$ on $\boldsymbol{\Omega}_{\mathbb{G}}$ with absolutely continuous measures $\nu\ll\mu$ on Ω.

Lemma 2.14. *Let $\boldsymbol{\nu}$ be a Borel probability measure on $\boldsymbol{\Omega}_{\mathbb{G}}$. Then $\boldsymbol{\nu}\ll\boldsymbol{\mu}$ if and only if there exists a Borel probability measure ν on Ω such that $\nu\ll\mu$ and $\boldsymbol{\nu} = \Psi_\sharp\nu$.*

Proof. First suppose that there exists $\nu \ll \mu$ with $\boldsymbol{\nu} = \Psi_\sharp \nu$. Then, for any Borel $A \subseteq \boldsymbol{\Omega}_{\mathbb{G}}$ with $\boldsymbol{\mu}(A) = 0$, we have that $\mu(\Psi^{-1}(A)) = \boldsymbol{\mu}(A) = 0$, and so

$$\boldsymbol{\nu}(A) = \nu(\Psi^{-1}(A)) = 0.$$

Now suppose that $\boldsymbol{\nu} \ll \boldsymbol{\mu}$. Let $\boldsymbol{\rho} \coloneqq \frac{d\boldsymbol{\nu}}{d\boldsymbol{\mu}}$ and define the measure ν via $\frac{d\nu}{d\mu} = \boldsymbol{\rho} \circ \Psi$ so that $\nu \ll \mu$. By Proposition 2.13, $\boldsymbol{\mu}(\boldsymbol{\Omega}_{\mathbb{G}} \setminus \Psi(\Omega)) = 0$ and $\boldsymbol{\mu} = \Psi_\sharp \mu$. Therefore

$$\nu(\Omega) = \int_\Omega \boldsymbol{\rho}(\Psi(\omega))\mu(\mathrm{d}\omega) = \int_{\Psi(\Omega)} \boldsymbol{\rho}(\boldsymbol{\omega})\boldsymbol{\mu}(\mathrm{d}\boldsymbol{\omega}) = \int_{\boldsymbol{\Omega}_{\mathbb{G}}} \boldsymbol{\rho}(\boldsymbol{\omega})\boldsymbol{\mu}(\mathrm{d}\boldsymbol{\omega}) = 1,$$

and so ν is a probability measure on Ω. Moreover, for any Borel measurable $A \subset \boldsymbol{\Omega}_{\mathbb{G}}$,

$$\boldsymbol{\nu}(A) = \int_A \boldsymbol{\rho}(\boldsymbol{\omega})\boldsymbol{\mu}(\mathrm{d}\boldsymbol{\omega}) = \int_{\Psi^{-1}(A)} \boldsymbol{\rho}(\Psi(\omega))\mu(\mathrm{d}\omega) = \int_{\Psi^{-1}(A)} \nu(\mathrm{d}\omega) = \nu(\Psi^{-1}(A))$$

Hence $\boldsymbol{\nu} = \Psi_\sharp \nu$. □

We show that the following shift map is well defined in Proposition 2.16 below.

Definition 2.15. (Shift map) For $h \in \mathrm{AC}_0([0,1], \mathbb{R}^{d_1})$ define the shift map $T_h \colon \boldsymbol{\Omega}_{\mathbb{G}} \to \boldsymbol{\Omega}_{\mathbb{G}}$ by

$$T_h \boldsymbol{\omega} = \lim_{n\to\infty} \Psi(\omega^n + h), \quad \boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbb{G}}, \tag{2.16}$$

where $(\omega^n)_{n\in\mathbb{N}}$ denotes the geodesic approximation from Lemma 2.11.

Proposition 2.16. *The shift map defined in Definition 2.15 satisfies the following:*

(i) For $h \in \mathcal{H}$, the shift map T_h is well defined.

(ii) For $\boldsymbol{X} = (\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) \in \boldsymbol{\Omega}_{\mathbb{G}}$ and $h \in \mathcal{H}$ the shift map $T_h\boldsymbol{X}$ is explicitly given by

$$(T_h\boldsymbol{X})_t^{(1)} = \boldsymbol{X}_t^{(1)} + h_t, \quad (T_h\boldsymbol{X})_t^{(2)} = \boldsymbol{X}_t^{(2)} + \frac{1}{2}\mathsf{W}\left(\mathbb{X}_t^1 + \mathbb{X}_t^2 + \mathbb{X}_t^3\right),$$

where

$$\mathrm{d}\mathbb{X}_t^1 = \boldsymbol{X}_t^{(1)} \otimes \mathrm{d}h_t, \quad \mathrm{d}\mathbb{X}_t^2 = h_t \otimes \mathrm{d}\boldsymbol{X}_t^{(1)}, \quad \mathrm{d}\mathbb{X}_t^3 = h_t \otimes \mathrm{d}h_t.$$

(iii) If $\boldsymbol{X}$ is given by $\boldsymbol{X} = \Psi(X)$ for $X \in \mathrm{Dom}(\Psi)$, then, for $h \in \mathcal{H}$,

$$T_h\boldsymbol{X} = T_h\Psi(X) = \Psi(X + h). \tag{2.17}$$

(iv) The map $\mathrm{AC}_0([0,1], \mathbb{R}^{d_1}) \times \boldsymbol{\Omega}_{\mathbb{G}} \to \boldsymbol{\Omega}_{\mathbb{G}}$, $(h, \boldsymbol{X}) \mapsto T_h\boldsymbol{X}$ is continuous.

Proof. If $\boldsymbol{X} = \Psi(X)$ is an absolutely continuous horizontal curve, we have $T_h\boldsymbol{X} = \Psi(X + h)$, for $h \in \mathcal{H}$. Then (ii) follows by definition of the lift on $\mathrm{AC}_0([0,1], \mathbb{R}^{d_1})$. The representation of (ii), with cross integrals in $\mathbb{X}^1, \mathbb{X}^2$, remains meaningful when X is only continuous, by basic properties of Riemann–Stieltjes integration. By continuity properties of Riemann–Stieltjes integration, we obtain that the limit in (2.16) exists, so that (i) and (ii) follow.

Now let $X \in \mathrm{Dom}(\Psi)$, so that $\boldsymbol{X} = \Psi(X) = \lim_{n\to\infty} \Psi(\widehat{X}^n)$ for the piecewise linear approximation $(\widehat{X}^n)$. Let $h \in \mathcal{H}$. Then $\boldsymbol{X}^{(1)} = \lim_{n\to\infty} \widehat{X}^n$ and $\boldsymbol{X}^{(2)} = \lim_{n\to\infty} \frac{1}{2}\mathsf{W}\mathbb{X}^{0,n}$, where $\mathrm{d}\mathbb{X}_t^{0,n} = \widehat{X}_t^n \otimes \mathrm{d}\widehat{X}_t^n$. By the definition of Ψ on $\mathrm{AC}_0([0,1], \mathbb{R}^d)$, we have that

$$\Psi(\widehat{X}^n + h) = \Big(\widehat{X}^n + h,\ \frac{1}{2}\mathsf{W}(\mathbb{X}_t^{0,n} + \mathbb{X}_t^{1,n} + \mathbb{X}_t^{2,n} + \mathbb{X}_t^3)\Big),$$

with $\mathbb{X}^{1,n}$, $\mathbb{X}^{2,n}$ defined as in (ii) with X replaced by $\widehat{X}^n$, and $\mathbb{X}^3$ defined as in (ii). By continuity of the Riemann–Stieltjes integral and (ii), we deduce that $\lim_{n\to\infty}\Psi(\widehat{X}^n+h)=T_h\Psi(X)$. This proves (iii).

Finally, (iv) also follows from continuity properties of the Riemann–Stieltjes integral and the representation from (ii). ☐

Remark 2.17. As is plain from Proposition 2.16, part (ii), we can translate any $\boldsymbol{X}=(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})\in\boldsymbol{\Omega}_{\mathbb{G}}$ in the direction of any absolutely continuous h. The situation is more complicated when dealing with Carnot groups of level strictly greater than 2, cf. [FV10, Section 9.4.6], or when h has less regularity, as is the case for Cameron–Martin paths of fractional Brownian motion with Hurst parameter $H<1/2$; cf. Section 4. In these cases, one has to incorporate suitable p-variation or Hölder rough path regularity on the path space of $\boldsymbol{X}$.

Remark 2.18. For any $h\in\mathcal{H}$ and $\boldsymbol{X}\in\boldsymbol{\Omega}_{\mathbb{G}}$, manipulating the expression for $T_h\boldsymbol{X}$ from Proposition 2.16 gives $\boldsymbol{X}_{s,t}^{-1}(T_h\boldsymbol{X})_{s,t}=(\boldsymbol{Z}_{s,t}^{(1)},\boldsymbol{Z}_{s,t}^{(2)})$, for $s,t\in[0,1]$, $s\le t$, where

$$\boldsymbol{Z}_{s,t}^{(1)}=h_{s,t},\quad \boldsymbol{Z}_{s,t}^{(2)}=\frac{1}{2}\Big(\mathbb{X}_{s,t}^{1}+\mathbb{X}_{s,t}^{2}+\mathbb{X}_{s,t}^{3}+2X_t\otimes h_s-(X_s\otimes h_s+X_t\otimes h_t+h_s\otimes h_t)\Big).$$

Also define $\boldsymbol{h}=\Psi(h)$. Then, after integrating by parts, we find that, for any $s,t\in[0,1]$ with $s\le t$,

$$\boldsymbol{\theta}_{s,t}:=\boldsymbol{h}_{s,t}^{-1}\boldsymbol{X}_{s,t}^{-1}(T_h\boldsymbol{X})_{s,t}=\Big(0,\int_s^t \mathsf{W}h_{s,r}\otimes \mathrm{d}X_r\Big).$$

The increments $\boldsymbol{\theta}_{s,t}$ can be interpreted as an error of non-commutativity between the increments of the shifted path $T_h\boldsymbol{X}$ and the increments of the (right-)translation $\boldsymbol{X}\boldsymbol{h}$ by the lifted path $\boldsymbol{h}$.

We now use the shift map to define a cost function on $\boldsymbol{\Omega}_{\mathbb{G}}$.

Definition 2.19. Define a cost function $C_{\mathcal{H}}\colon\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}\to[0,\infty]$ by

$$C_{\mathcal{H}}(\boldsymbol{X},\boldsymbol{Y}):=\begin{cases}\|h\|_{\mathcal{H}}, & \text{if } \boldsymbol{Y}=T_h\boldsymbol{X}, \text{ for some } h\in\mathcal{H},\\ +\infty, & \text{otherwise.}\end{cases}\tag{2.18}$$

Lemma 2.20. *The cost $C_{\mathcal{H}}\colon\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}\to[0,\infty]$ is lower semicontinuous.*

Proof. Let $\boldsymbol{X},\boldsymbol{Y}\in\boldsymbol{\Omega}_{\mathbb{G}}$ and let $(\boldsymbol{X}^n),(\boldsymbol{Y}^n)\subset\boldsymbol{\Omega}_{\mathbb{G}}$ be sequences such that $(\boldsymbol{X}^n,\boldsymbol{Y}^n)\to(\boldsymbol{X},\boldsymbol{Y})$. We may assume that there exists a subsequence $n_k\to\infty$ such that $\boldsymbol{Y}^{n_k}=T_{h^{n_k}}\boldsymbol{X}^{n_k}$, where $h^{n_k}=\pi_1\boldsymbol{Y}^{n_k}-\pi_1\boldsymbol{X}^{n_k}\in\mathcal{H}$, and $\liminf_{n\to\infty}C_{\mathcal{H}}(\boldsymbol{X}^n,\boldsymbol{Y}^n)=\lim_{k\to\infty}C_{\mathcal{H}}(\boldsymbol{X}^{n_k},\boldsymbol{Y}^{n_k})=:I<\infty$. Then we have that

$$I=\lim_{k\to\infty}C_{\mathcal{H}}(\boldsymbol{X}^{n_k},\boldsymbol{Y}^{n_k})=\liminf_{k\to\infty}\|h^{n_k}\|_{\mathcal{H}}\ge\|h\|_{\mathcal{H}},$$

where $h=\pi_1\boldsymbol{Y}-\pi_1\boldsymbol{X}\in\mathcal{H}$. By the continuity of the shift shown in Proposition 2.16 (iv), $\boldsymbol{Y}=T_h\boldsymbol{X}$, and so $C_{\mathcal{H}}(\boldsymbol{X},\boldsymbol{Y})=\|h\|_{\mathcal{H}}$. ☐

2.3. **Discussion on the choice of the cost function $C_{\mathcal{H}}$.** The choice of cost function $C_{\mathcal{H}}$ is natural in the sense that it arises as the Γ-limit of the sequence C_n, as shown in Section 6.4. Moreover, $C_{\mathcal{H}}$ has the crucial property that whenever $H(\boldsymbol{\nu}\|\boldsymbol{\mu})=+\infty$, also $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\mu},\boldsymbol{\nu})=+\infty$. Indeed, supposing that $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\mu},\boldsymbol{\nu})<\infty$, there exists a coupling $\boldsymbol{\lambda}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ such that

$$\boldsymbol{\lambda}(\{\,(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\in\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}:\ \overline{\boldsymbol{\omega}}=T_{\overline{\omega}-\omega}\boldsymbol{\omega},\ \overline{\omega}-\omega\in\mathcal{H}\,\})=1,$$

and $\int C^2_{\mathcal{H}}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}}) \,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}}) < \infty$. We have that $\boldsymbol{\mu} = \Psi_\sharp\mu$, by Proposition 2.13, and combining this with (2.17) from Proposition 2.16 gives

$$\boldsymbol{\lambda}(\{ (\boldsymbol{\omega},\overline{\boldsymbol{\omega}}) \in \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}} :\ \boldsymbol{\omega} = \Psi(\omega),\, \overline{\boldsymbol{\omega}} = \Psi(\overline{\omega}),\, \overline{\omega} - \omega \in \mathcal{H} \}) = 1.$$

Thus there exists $\nu \in \mathcal{P}(\Omega)$ such that $\boldsymbol{\nu} = \Psi_\sharp\nu$. By Itô representation and Girsanov's theorem, we also have that $\nu \ll \mu$. Hence $\boldsymbol{\nu} = \Psi_\sharp\nu \ll \Psi_\sharp\mu = \boldsymbol{\mu}$, and $H(\boldsymbol{\nu}\|\boldsymbol{\mu}) < \infty$.

This is in contrast to the cost function $\widetilde{C}_{\mathcal{H}} \coloneqq c_{\mathcal{H}} \circ (\pi_1 \times \pi_1) \colon \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}} \to [0,\infty]$, which appears in the cost-information inequality in [Rie17, Corollary 1.4]. Indeed, consider the Brownian motion $\boldsymbol{B} = (B, \mathbb{A})$ on the Heisenberg group $\mathbb{G} \cong \mathbb{R}^2 \oplus \mathbb{R}$. Let $\boldsymbol{\nu} = \mathrm{Law}(\boldsymbol{X})$, where $\boldsymbol{X} = (B, \mathbb{X})$ is defined as follows. Let $M \in (0,\infty)$, and define $\mathbb{X}_0 = 0$ and $\mathbb{X}_{s,t} = \mathbb{A}_{s,t} + (t-s)M$, for all $s,t \in [0,1]$ with $s < t$. Since $\boldsymbol{X}$ only differs from $\boldsymbol{B}$ in the second component, we see that

$$0 \le \mathrm{T}^2_{\widetilde{C}_{\mathcal{H}},2}(\boldsymbol{\mu},\boldsymbol{\nu}) \le \mathbb{E}\Big[\widetilde{C}^2_{\mathcal{H}}(\boldsymbol{B},\boldsymbol{X})\Big] = \mathbb{E}\Big[c^2_{\mathcal{H}}(B,B)\Big] = 0.$$

However, $\boldsymbol{\nu}$ is not absolutely continuous with respect to $\boldsymbol{\mu}$. Indeed, supposing that $\boldsymbol{\nu} \ll \boldsymbol{\mu}$, Lemma 2.14 implies that $\boldsymbol{\nu} = \Psi_\sharp\nu$ for some probability measure $\nu \ll \mu$. It follows that $\boldsymbol{\nu}(\Psi(\Omega)) = \nu(\Omega) = 1$. On the other hand, since $\boldsymbol{\nu} = \mathrm{Law}(\boldsymbol{X})$ with $\boldsymbol{X} = (B,\mathbb{X}) \neq \Psi(B)$, we see that $\boldsymbol{\nu}(\Psi(\Omega)) < 1$, which is a contradiction. Hence $H(\boldsymbol{\nu}\|\boldsymbol{\mu}) = +\infty$.

3. Talagrand for Brownian motion on Carnot groups – Direct approach via Föllmer drift

In this section, we give a first proof of Talagrand's $\mathcal{T}_2$ inequality for the law of Brownian motion on a step-2 Carnot group. We follow the strategy of [Leh13] and [Föl22], using Föllmer's intrinsic drift from [Föl86, Föl88]. Moreover, we show that equality is attained in the $\mathcal{T}_2$ inequality when restricting to adapted couplings, as was shown in the classical case in [Las18, Lemma 5] and [Föl22, Theorem 3].

We first give a characterisation of adapted couplings on $\mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$. In particular, we show that our definition of adapted couplings in Definition 2.4 is consistent with that of [Föl22, Definition 1].

Lemma 3.1. *Suppose that $\nu \ll \mu$. Then $\lambda \in \Pi(\mu,\nu)$ is an adapted coupling if and only if there exists a filtered probability space $(\widetilde{\Omega}, \mathcal{F}, (\mathcal{F}_t), \mathbb{Q})$ on which processes X, Y are defined such that X is a Brownian motion, Y is an adapted process, and $\lambda = \mathrm{Law}(X,Y)$ under $\mathbb{Q}$.*

Suppose that $\boldsymbol{\nu} \ll \boldsymbol{\mu}$. Then $\boldsymbol{\lambda} \in \Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ is an adapted coupling if and only if there exists $\nu \ll \mu$ and an adapted coupling $\lambda \in \Pi_{\mathrm{ad}}(\mu,\nu)$ such that $\boldsymbol{\lambda} = (\Psi \times \Psi)_\sharp\lambda$.

Proof. Suppose that $\nu \ll \mu$. By [Las18, Lemma 4], our Definition 2.4 of adapted couplings is equivalent to the symmetric counterpart of [Las18, Definition 1] (see [Las18, Section 4.1]). Then the first claim follows from [Las18, Propositions 3 and 4].

Now suppose that $\boldsymbol{\nu} \ll \boldsymbol{\mu}$. By Lemma 2.14, there exists $\nu \ll \mu$ such that $\boldsymbol{\nu} = \Psi_\sharp\nu$. If X is an $\mathbb{R}^{d_1}$-valued process with natural filtration $(\mathcal{F}_t)_{t\in[0,1]}$ completed with respect to the law of X, and $\boldsymbol{X} = \Psi(X)$ is a $\mathbb{G}$-valued process with natural filtration $(\boldsymbol{\mathcal{F}}_t)_{t\in[0,1]}$ completed with respect to the law of $\boldsymbol{X}$, then $\boldsymbol{\mathcal{F}}_t = \Psi(\mathcal{F}_t)$, for all $t \in [0,1]$. Since $\boldsymbol{\mu} = \Psi_\sharp\mu$ and $\boldsymbol{\nu} = \Psi_\sharp\nu$, we have that $\boldsymbol{\lambda} \in \Pi_{\mathrm{ad}}(\boldsymbol{\mu},\boldsymbol{\nu})$ if and only if $\boldsymbol{\lambda} = (\Psi \times \Psi)_\sharp\lambda$ for some $\lambda \in \Pi_{\mathrm{ad}}(\mu,\nu)$. ☐

Remark 3.2. By Lemma 3.1, if $\boldsymbol{\lambda} \in \Pi_{\mathrm{ad}}(\boldsymbol{\mu},\boldsymbol{\nu})$, then there exist $\boldsymbol{X} = \Psi(X)$, $\boldsymbol{Y} = \Psi(Y)$ defined on some filtered probability space $(\widetilde{\Omega}, \mathcal{F}, (\mathcal{F}_t), \mathbb{Q})$ such that $\boldsymbol{\lambda} = \mathrm{Law}_{\mathbb{Q}}(\boldsymbol{X},\boldsymbol{Y})$, where $\boldsymbol{X}$ is a Brownian motion on

$\mathbb{G}$ and $\boldsymbol{Y}$ is an adapted $\mathbb{G}$-valued process, and $\mathrm{Law}_{\mathbb{Q}}$ denotes the law under $\mathbb{Q}$. Then, letting $\mathbb{E}_{\mathbb{Q}}$ denote expectation with respect to $\mathbb{Q}$, we have

$$\int C^2_{\mathcal{H}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}})\,\mathrm{d}\boldsymbol{\lambda} = \mathbb{E}_{\mathbb{Q}}[C^2_{\mathcal{H}}(\boldsymbol{X}, \boldsymbol{Y})].$$

We now prove the main result of this section, showing that $\boldsymbol{\mu}$ satisfies a $\mathcal{T}_2$ inequality, and that equality holds when restricting to adapted couplings.

Theorem 3.3. *Let $\boldsymbol{\nu} \ll \boldsymbol{\mu}$. Then there exists $\nu \ll \mu$ such that $\boldsymbol{\nu} = \Psi_\sharp \nu$, and there exists a predictable process b^ν on $\mathbb{R}^d$ with $b^\nu \in L^2$, ν-almost surely, such that $B^\nu := B - \int_0^\cdot b^\nu_t\,\mathrm{d}t$ is a Brownian motion under ν, and $\boldsymbol{\lambda}^* = \mathrm{Law}_\nu(\Psi(B^\nu), \Psi(B))$ is the unique optimal adapted coupling of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ with*

$$\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\nu}, \boldsymbol{\mu})^2 \le \mathrm{T}^{\mathrm{ad}}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \boldsymbol{\nu})^2 = \mathbb{E}_\nu[C^2_{\mathcal{H}}(\Psi(B^\nu), \Psi(B))] = 2H(\boldsymbol{\nu}\|\boldsymbol{\mu}). \tag{3.1}$$

In particular, $\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{G}}, C_{\mathcal{H}}, 1)$.

Proof. By Lemma 2.14, there exists a probability measure ν on Ω with $\nu \ll \mu$ and $\Psi_\sharp \nu = \boldsymbol{\nu}$. Writing $\boldsymbol{\rho} := \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\boldsymbol{\mu}}$, we have $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = \boldsymbol{\rho} \circ \Psi$. Since $\boldsymbol{\mu}(\boldsymbol{\Omega}_{\mathbb{G}} \setminus \Psi(\Omega)) = 0$, by Proposition 2.13, we obtain that

$$H(\boldsymbol{\nu}\|\boldsymbol{\mu}) = \int_{\Psi(\Omega)} \boldsymbol{\rho}(\boldsymbol{\omega}) \log(\boldsymbol{\rho}(\boldsymbol{\omega}))\,\mathrm{d}\boldsymbol{\mu}(\boldsymbol{\omega}) = \int_\Omega \boldsymbol{\rho}(\Psi(\omega)) \log(\boldsymbol{\rho}(\Psi(\omega))\,\mathrm{d}\mu(\omega) = H(\nu\|\mu).$$

We can apply [Föl22, Proposition 1] to obtain that there exists a predictable process b^ν with $b^\nu \in L^2$, ν-almost surely, such that $B^\nu = B - \int_0^\cdot b^\nu_t\,\mathrm{d}t$ is a Brownian motion under ν with

$$\mathbb{E}_\nu[\|B^\nu - B\|^2_{\mathcal{H}}] = 2H(\nu\|\mu) = 2H(\boldsymbol{\nu}\|\boldsymbol{\mu}). \tag{3.2}$$

Moreover, from [Föl22, Theorem 3] it follows that $\lambda^* = \mathrm{Law}_\nu(B^\nu, B)$ is the unique optimal adapted coupling between μ and ν. By Lemma 3.1, $\boldsymbol{\lambda}^* = \mathrm{Law}_\nu(\Psi(B^\nu), \Psi(B))$ is an adapted coupling of $\boldsymbol{\mu} = \Psi_\sharp\mu$ and $\boldsymbol{\nu} = \Psi_\sharp\nu$. Thus, using the definition of $C_{\mathcal{H}}$ from (2.18),

$$\mathrm{T}^{\mathrm{ad}}_{c_{\mathcal{H}},2}(\mu, \nu)^2 = \mathbb{E}_\nu[\|B^\nu - B\|^2_{\mathcal{H}}] = \mathbb{E}_\nu[C^2_{\mathcal{H}}(\Psi(B^\nu), \Psi(B))] \ge \mathrm{T}^{\mathrm{ad}}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \boldsymbol{\nu})^2. \tag{3.3}$$

On the other hand, since $c_{\mathcal{H}} \circ (\pi_1 \times \pi_1) \le C_{\mathcal{H}}$, applying Lemma 3.1 gives

$$\begin{aligned}
\mathrm{T}^{\mathrm{ad}}_{c_{\mathcal{H}},2}(\mu, \nu)^2 = \inf_{\lambda \in \Pi_{\mathrm{ad}}(\mu,\nu)} \int_{\Omega\times\Omega} c^2_{\mathcal{H}}(\omega, \overline{\omega})\,\mathrm{d}\lambda(\omega, \overline{\omega}) &= \inf_{\boldsymbol{\lambda} \in \Pi_{\mathrm{ad}}(\boldsymbol{\mu},\boldsymbol{\nu})} \int_{\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}} c^2_{\mathcal{H}}(\pi_1\boldsymbol{\omega}, \pi_1\overline{\boldsymbol{\omega}})\,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \\
&\le \inf_{\boldsymbol{\lambda} \in \Pi_{\mathrm{ad}}(\boldsymbol{\mu},\boldsymbol{\nu})} \int_{\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}} C^2_{\mathcal{H}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}})\,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \\
&= \mathrm{T}^{\mathrm{ad}}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \boldsymbol{\nu})^2.
\end{aligned}$$

Hence we have optimality of $\boldsymbol{\lambda}^*$ and

$$\mathrm{T}^{\mathrm{ad}}_{C_{\mathcal{H}},2}(\boldsymbol{\nu}, \boldsymbol{\mu})^2 = \mathbb{E}_\nu[C_{\mathcal{H}^2}(\Psi(B^\nu), \Psi(B))] = \mathbb{E}_\nu[\|B^\nu - B\|^2_{\mathcal{H}}] = \mathrm{T}^{\mathrm{ad}}_{c_{\mathcal{H}},2}(\mu, \nu)^2.$$

Applying (3.2) gives (3.1), and uniqueness of the optimiser follows by Lemma 3.1. ☐

Remark 3.4. Suppose that h is an adapted process with $h \in \mathcal{H}$ almost surely, and let $\boldsymbol{\nu} = \mathrm{Law}(T_h\boldsymbol{B})$. Then we can take $b^\nu = \dot{h}$ in Theorem 3.3. Indeed, by Girsanov's theorem, $B - h$ is a Brownian motion under ν, and $H(\boldsymbol{\nu}\|\boldsymbol{\mu}) = H(\nu\|\mu) = \mathbb{E}_\nu[\log \mathrm{d}\nu/\mathrm{d}\mu] = \mathbb{E}_\nu[\|h\|^2_{\mathcal{H}}]$.

4. Talagrand for Gaussian rough paths – Direct approach via contraction

We now give an alternative proof of Talagrand's $\mathcal{T}_2$ inequality in a more general setting, following the contraction approach of [Rie17].

Let Z be a d-dimensional continuous Gaussian process that admits a level 2 "rough path" lift $\boldsymbol{Z} = \boldsymbol{Z}(\omega)$ with $\boldsymbol{Z} \in \mathcal{D}$ almost surely, where $\mathcal{D}$ is a suitable p-variation (or β-Hölder) rough path space (cf. [FV10, Chapter 15]). We refer to [FV10, Chapter 15] for conditions under which a Gaussian process can be lifted to a Gaussian rough path. Here we simply assume that such a lift exists. Let $\nu \in \mathcal{P}(\Omega)$ denote the law of the Gaussian process Z, $\boldsymbol{\nu} \in \mathcal{P}(\mathcal{D})$ the law of $\boldsymbol{Z}$, and $\mathcal{H}_\nu$ the Cameron–Martin space of Z (cf. [Jan97, Chapter 8, Section 4]).

A $\mathcal{T}_2$ inequality is known to hold for general Gaussian processes with $\alpha = 1$ and cost

$$\widetilde{c}_{\mathcal{H}_\nu}(x, y) = \begin{cases} \|h\|_{\mathcal{H}_\nu}, & \text{if } x - y \in \mathcal{H}_\nu, \\ +\infty, & \text{otherwise;} \end{cases}$$

that is $\nu \in \mathcal{T}_2(\Omega, \widetilde{c}_{\mathcal{H}_\nu}, 1)$; see [FÜ04, Theorem 3.1] and [Rie17, Theorem 1.2].

We work under the following assumption.

Assumption 4.1. Suppose that there exists $\widetilde{\Omega} \subseteq \Omega$ with $\nu(\widetilde{\Omega}) = 1$ such that

(i) There exists a Borel-measurable lift map $\widetilde{\Psi} \colon \Omega \to \mathcal{D}$ with $\pi_1 \widetilde{\Psi}(x) = x$, for $x \in \widetilde{\Omega}$, where π_1 is the projection onto the first component, such that $\widetilde{\Psi}(Z) = \boldsymbol{Z}$ almost surely;

(ii) There exists a continuous shift map

$$\mathcal{H}_\nu \times \mathcal{D} \to \mathcal{D}, \quad (h, \boldsymbol{x}) \mapsto \widetilde{T}_h \boldsymbol{x},$$

such that

$$\widetilde{T}_h \widetilde{\Psi}(x) = \widetilde{\Psi}(x + h), \quad x \in \widetilde{\Omega}, \quad h \in \mathcal{H}_\nu. \tag{4.1}$$

Remark 4.2. By standard results [FV10, Chapter 15], we see that Assumption 4.1 is satisfied for $Z = B$ a fractional Brownian motion and its lift $\boldsymbol{B}$ in the step-2 Carnot group $\mathbb{G}$, with path space $\mathcal{D} = C_0^{p-\mathrm{var}}([0, 1], \mathbb{G})$, for $H \in (1/3, 1/2]$, $p \in (1/H, 3)$. Extensions to $H > 1/4$ are possible, at the price of lifting Z to a step-3 Carnot group; we do not give details for the sake of brevity.

Definition 4.3. Define the cost $\widetilde{C}_{\mathcal{H}_\nu} \colon \mathcal{D} \times \mathcal{D} \to [0, \infty]$, similarly to (2.18), by

$$\widetilde{C}_{\mathcal{H}_\nu}(\boldsymbol{X}, \boldsymbol{Y}) := \begin{cases} \|h\|_{\mathcal{H}_\nu}, & \text{if } \boldsymbol{Y} = \widetilde{T}_h \boldsymbol{X}, \text{ for some } h \in \mathcal{H}_\nu, \\ +\infty, & \text{otherwise,} \end{cases} \tag{4.2}$$

where $\widetilde{T}_h$ is the shift from Assumption 4.1.

Due to the assumed continuity of the shift, we recover measurability of $\widetilde{C}_{\mathcal{H}_\nu}$ (cf. Lemma 2.20). By applying the contraction principle from Lemma 2.2, similarly to [Rie17], but for a different cost, we lift the Talagrand inequality to the rough path space $\mathcal{D}$.

Theorem 4.4. *Let Assumption 4.1 hold. Then* $\boldsymbol{\nu} \in \mathcal{T}_2(\mathcal{D}, \widetilde{C}_{\mathcal{H}_\nu}, 1)$.

Proof. The property (4.1) implies that for $h \in \mathcal{H}_\nu$, a path $x \in \widetilde{\Omega}$ satisfies $x = y + h$ if and only if $\widetilde{\Psi}(x) = \widetilde{T}_h\widetilde{\Psi}(y)$. Hence for $x - y \in \mathcal{H}_\nu$, we have that $\widetilde{\Psi}(y + (x - y)) = \widetilde{T}_{x-y}\widetilde{\Psi}(y)$ and thus

$$\|x - y\|^2_{\mathcal{H}_\nu} = \widetilde{C}^2_{\mathcal{H}_\nu}(\widetilde{\Psi}(x), \widetilde{\Psi}(y)).$$

Similarly, if $x - y \notin \mathcal{H}_\nu$, then $\widetilde{\Psi}(x)$ is not a shift of $\widetilde{\Psi}(y)$ and so $\widetilde{C}_{\mathcal{H}_\nu}(\widetilde{\Psi}(x), \widetilde{\Psi}(y)) = +\infty$. Together with $\nu \in \mathcal{T}_2(\Omega, \widetilde{c}_{\mathcal{H}_\nu}, 1)$, $\nu(\widetilde{\Omega}) = 1$, and measurability of $\widetilde{C}_{\mathcal{H}_\nu}$ and $\widetilde{\Psi}$, an application of the contraction principle (Lemma 2.2) with $L = 1$ then yields $\boldsymbol{\nu} \in \mathcal{T}_2(\mathcal{D}, \widetilde{C}_{\mathcal{H}_\nu}, 1)$. $\square$

5. Talagrand for the heat kernel measure on Carnot groups

In this section, we show that a $\mathcal{T}_2$ inequality on the step-2 Carnot group $\mathbb{G}$ follows from a log-Sobolev inequality, which in turn can be deduced from a heat semigroup estimate. In particular, we prove a $\mathcal{T}_2$ inequality for the heat kernel measure $\boldsymbol{\mu}_1$ in the case that $\mathbb{G}$ is an H-type group. We will apply this result in Section 6 to show that a $\mathcal{T}_2$ inequality also holds on the path space by a bottom-up approach.

5.1. **From log-Sobolev to Talagrand.** We follow the approach put forward by Otto–Villani [OV00], namely deducing a $\mathcal{T}_2$ inequality as a consequence of a log-Sobolev inequality. We will make use of the generalisation by Gigli–Ledoux [GL13] of Otto–Villani's result. Whereas the result of Gigli–Ledoux [GL13] depends on the log-Sobolev inequality for Lipschitz test functions, we show via a mollification argument that this can be relaxed to only requiring the log-Sobolev inequality for smooth test functions; see Theorem 5.1. Moreover, for the heat kernel measure, we show that the log-Sobolev inequality for smooth test functions follows from certain heat semigroup estimates; see Theorem 5.2. In the special case of H-type groups, as defined in Remark 2.10, the required heat semigroup estimates are known. Thus, in Theorem 5.3, we show that a $\mathcal{T}_2$ inequality holds for the heat kernel measure on H-type groups and, in particular, on the Heisenberg group.

Let $\mathbb{G}$ be a step-2 Carnot group. Recall that, for the Carnot–Carathéodory metric d_{CC} defined in (2.9), the space $(\mathbb{G}, d_{\mathrm{CC}})$ is a Polish space. Hence, for any Borel probability measure $\boldsymbol{\eta}$ on $\mathbb{G}$, the space $(\mathbb{G}, d_{\mathrm{CC}}, \boldsymbol{\eta})$ is a metric measure space in the sense of [GL13].

For a locally Lipschitz function $f \colon \mathbb{G} \to \mathbb{R}$, define the local Lipschitz constant $\mathrm{Lip}_{\mathbb{G}}(f)$ by

$$\mathrm{Lip}_{\mathbb{G}}(f)(\boldsymbol{x}) := \limsup_{\boldsymbol{y} \to \boldsymbol{x}} \frac{|f(\boldsymbol{y}) - f(\boldsymbol{x})|}{d_{\mathrm{CC}}(\boldsymbol{x}, \boldsymbol{y})}, \quad \boldsymbol{x} \in \mathbb{G}.$$

By an extension of Rademacher's theorem due to Pansu [Pan89] (see also [DPMM$^+$25, PS17], and [LD25, Theorem 11.3.2]), we obtain that every Lipschitz continuous function $f \colon U \subseteq \mathbb{G} \to \mathbb{R}$ is Pansu differentiable Lebesgue-almost everywhere. In particular, its gradient $\nabla_{\mathbb{G}} f$ exists $\mathcal{L}^m$-almost everywhere and $\mathrm{Lip}_{\mathbb{G}}(f)(\boldsymbol{x}) = |\nabla_{\mathbb{G}} f(\boldsymbol{x})|_{\mathbb{G}}$, for $\mathcal{L}^m$-almost every $\boldsymbol{x} \in \mathbb{G}$.

We say that a Borel probability measure $\boldsymbol{\eta}$ on $\mathbb{G}$ with $\boldsymbol{\eta} \ll \mathcal{L}^m$ satisfies the log-Sobolev inequality if there exists $\alpha \in (0, \infty)$ such that

$$\text{(LSI)} \qquad 2\alpha \int_{\mathbb{G}} f \log f \,\mathrm{d}\boldsymbol{\eta} \le \int_{\{f>0\}} \frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f} \,\mathrm{d}\boldsymbol{\eta}, \quad \text{for all } f \in C_c^\infty(\mathbb{G}, [0, \infty)), \int_{\mathbb{G}} f \,\mathrm{d}\boldsymbol{\eta} = 1.$$

We first show that (LSI) implies a $\mathcal{T}_2$ inequality. As an intermediate step, we show that the log-Sobolev inequality also holds for Lipschitz functions, so that we can then apply [GL13, Theorem 5.2].

Theorem 5.1. *Suppose that $\boldsymbol{\eta}$ is a Borel probability measure on $\mathbb{G}$ with $\boldsymbol{\eta} \ll \mathcal{L}^m$ satisfying* (LSI) *for some $\alpha \in (0,\infty)$. Then, for any Lipschitz function $f\colon \mathbb{G} \to [0,\infty)$ with $\int_{\mathbb{G}} f \,\mathrm{d}\boldsymbol{\eta} = 1$,*

$$2\alpha \int_{\mathbb{G}} f \log f \,\mathrm{d}\boldsymbol{\eta} \le \int_{\{f>0\}} \frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f} \,\mathrm{d}\boldsymbol{\eta}. \tag{5.1}$$

Moreover, $\boldsymbol{\eta} \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$.

Proof. We argue as in the proof of [AS20, Theorem 4.8] via mollification. Consider $f\colon \mathbb{G} \to [0,\infty)$ Lipschitz continuous with $\int_{\mathbb{G}} f \,\mathrm{d}\boldsymbol{\eta} = 1$. In particular, its gradient $\nabla_{\mathbb{G}} f$ exists $\mathcal{L}^m$-almost everywhere.

Let $\rho\colon \mathbb{G} \to \mathbb{R}$ be a symmetric smooth mollifier in $\mathbb{G}$, i.e. a function $\rho \in C^\infty_c(\mathbb{R}^m, [0,\infty))$ such that $\operatorname{supp}\rho \subset B_1$, $0 \le \rho \le 1$, $\rho(\boldsymbol{x}^{-1}) = \rho(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathbb{G}$, and $\int_{\mathbb{G}} \rho \,\mathrm{d}\boldsymbol{x} = 1$. For $k \in \mathbb{N}$ and $\boldsymbol{x} \in \mathbb{G}$, set $\rho_k(\boldsymbol{x}) = k^Q \rho(\delta_k \boldsymbol{x})$ and define the mollification $\widehat{f}_k = \rho_k \star f$ by

$$\widehat{f}_k(\boldsymbol{x}) \coloneqq (\rho_k \star f)(\boldsymbol{x}) = \int_{\mathbb{G}} \rho_k(\boldsymbol{x}\boldsymbol{y}^{-1}) f(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} = \int_{\mathbb{G}} \rho_k(\boldsymbol{y}) f(\boldsymbol{y}^{-1}\boldsymbol{x}) \,\mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x} \in \mathbb{G}.$$

For each $k \in \mathbb{N}$, also define a smooth truncation function $\chi_k \in C^\infty_c(\mathbb{R}^m, [0,1])$ such that $\chi_k = 1$ in B_k, $\chi_k = 0$ in B^c_{2k}, and $|\nabla_{\mathbb{G}} \chi_k|_{\mathbb{G}} \le C/k$ for some constant $C \in (0,\infty)$, and define $f_k \coloneqq \widehat{f}_k \, \chi_k$.

Thus, $f_k \in C^\infty_c(\mathbb{G}, [0,\infty))$, for each $k \in \mathbb{N}$. Moreover, $f_k \to f$ in $L^1(\boldsymbol{\eta})$ and thus $\boldsymbol{\eta}$-almost everywhere along a subsequence. By Fatou's lemma, we have that

$$\liminf_{k\to\infty} \int_{\mathbb{G}} f_k \log f_k \,\mathrm{d}\boldsymbol{\eta} \ge \int_{\mathbb{G}} f \log f \,\mathrm{d}\boldsymbol{\eta}.$$

We next show an upper estimate for the lim sup of the right-hand side in (5.1) with $\widehat{f}_k$ in place of f. Note that, by left-invariance of the Carnot–Caratheodory distance, for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{G}$, we have

$$\frac{|\widehat{f}_k(\boldsymbol{x}_1) - \widehat{f}_k(\boldsymbol{x}_2)|}{d_{\mathrm{CC}}(\boldsymbol{x}_1, \boldsymbol{x}_2)} \le \int_{\mathbb{G}} \rho_k(\boldsymbol{y}) \frac{|f(\boldsymbol{y}^{-1}\boldsymbol{x}_1) - f(\boldsymbol{y}^{-1}\boldsymbol{x}_1)|}{d_{\mathrm{CC}}(\boldsymbol{x}_1, \boldsymbol{x}_2)} \,\mathrm{d}\boldsymbol{y} \le \int_{\mathbb{G}} \rho_k(\boldsymbol{y}) \frac{|f(\boldsymbol{y}^{-1}\boldsymbol{x}_1) - f(\boldsymbol{y}^{-1}\boldsymbol{x}_1)|}{d_{\mathrm{CC}}(\boldsymbol{y}^{-1}\boldsymbol{x}_1, \boldsymbol{y}^{-1}\boldsymbol{x}_2)} \,\mathrm{d}\boldsymbol{y}.$$

Thus, after passing to the lim sup for $\boldsymbol{x}_1 \to \boldsymbol{x}_2 = \boldsymbol{x}$, we obtain $\mathrm{Lip}_{\mathbb{G}} \widehat{f}_k(\boldsymbol{x}) \le \rho_k \star \mathrm{Lip}_{\mathbb{G}} f(\boldsymbol{x})$, or equivalently $|\nabla_{\mathbb{G}} \widehat{f}_k|_{\mathbb{G}}(\boldsymbol{x}) \le \rho_k \star |\nabla_{\mathbb{G}} f|_{\mathbb{G}}(\boldsymbol{x})$. With this estimate, the Cauchy–Schwarz inequality gives

$$|\nabla_{\mathbb{G}} \widehat{f}_k|^2_{\mathbb{G}} \le \left[\rho_k \star \left(\chi_{\{f>0\}} \sqrt{f} \frac{|\nabla_{\mathbb{G}} f|_{\mathbb{G}}}{\sqrt{f}}\right)\right]^2 \le f_k \left(\rho_k \star \frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f} \chi_{\{f>0\}}\right).$$

Multiplying f_k by the truncation χ_k and applying the product rule, we estimate

$$\begin{aligned}
\limsup_{k\to\infty} \int_{\{f_k>0\}} \frac{|\nabla_{\mathbb{G}} f_k|^2_{\mathbb{G}}}{f_k} \,\mathrm{d}\boldsymbol{\eta} &\le \limsup_{k\to\infty} \int_{\{\widehat{f}_k>0\}} \frac{|\nabla_{\mathbb{G}} \widehat{f}_k|^2_{\mathbb{G}}}{\widehat{f}_k} \,\mathrm{d}\boldsymbol{\eta} \\
&\le \limsup_{k\to\infty} \int_{\mathbb{G}} \rho_k \star \left(\frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f} \chi_{\{f>0\}}\right) \mathrm{d}\boldsymbol{\eta} \le \int_{\{f>0\}} \frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f} \,\mathrm{d}\boldsymbol{\eta}.
\end{aligned}$$

Finally, define $I_k \coloneqq \int_{\mathbb{G}} f_k \,\mathrm{d}\boldsymbol{\eta}$ and the normalised function $\widetilde{f}_k \coloneqq f_k / I_k$, for each $k \in \mathbb{N}$. Then, by (LSI),

$$\begin{aligned}
2\alpha \int_{\mathbb{G}} f_k \log f_k \,\mathrm{d}\boldsymbol{\eta} &= 2\alpha I_k \int_{\mathbb{G}} \widetilde{f}_k \log \widetilde{f}_k \,\mathrm{d}\boldsymbol{\eta} + 2\alpha I_k \log I_k \\
&\le I_k \int_{\{\widetilde{f}_k>0\}} \frac{|\nabla_{\mathbb{G}} \widetilde{f}_k|^2_{\mathbb{G}}}{\widetilde{f}_k} \,\mathrm{d}\boldsymbol{\eta} + 2\alpha I_k \log I_k = \int_{\{f_k>0\}} \frac{|\nabla_{\mathbb{G}} f_k|^2_{\mathbb{G}}}{f_k} \,\mathrm{d}\boldsymbol{\eta} + 2\alpha I_k \log I_k.
\end{aligned}$$

We conclude that

$$2\alpha \int_{\mathbb{G}} f \log f \, \mathrm{d}\boldsymbol{\eta} \leq \liminf_{k\to\infty} 2\alpha \int_{\mathbb{G}} f_k \log f_k \, \mathrm{d}\boldsymbol{\eta} \leq \limsup_{k\to\infty} \int_{\{f_k>0\}} \frac{|\nabla_{\mathbb{G}} f_k|_{\mathbb{G}}^2}{f_k} \, \mathrm{d}\boldsymbol{\eta} \leq \int_{\{f>0\}} \frac{|\nabla_{\mathbb{G}} f|_{\mathbb{G}}^2}{f} \, \mathrm{d}\boldsymbol{\eta}.$$

Thus, the log-Sobolev inequality for Lipschitz-continuous functions is established.

Finally, since $(\mathbb{G}, d_{\mathrm{CC}}, \boldsymbol{\eta})$ is a metric measure space in the sense of [GL13], we conclude that $\boldsymbol{\eta} \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$ by [GL13, Theorem 5.2]. □

5.2. **From heat semigroup estimates to Talagrand.** We now specialise to the case of the heat kernel measure $\boldsymbol{\mu}_1$ and give a sufficient condition for $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$, for some $\alpha \in (0, \infty)$.

As noted in [Eld10, Section 5] and [BBBC08, Remark 6.6], one can deduce the log-Sobolev inequality (LSI) for $\boldsymbol{\mu}_1$ from the following heat semigroup estimate: there exists a constant $K \in (0, \infty)$ such that

$$|\nabla_{\mathbb{G}} P_t f|_{\mathbb{G}} \leq K P_t(|\nabla_{\mathbb{G}} f|_{\mathbb{G}}), \quad \text{for all } f \in C_c^\infty(\mathbb{G}, \mathbb{R}), \ t \geq 0, \tag{5.2}$$

where P_t is the heat semigroup defined in (2.14). Indeed, for the Heisenberg group $\mathbb{G} = \mathbb{H} = \mathbb{H}^1$, [Li06, Théorème 1.1] proves the estimate (5.2), and [Li06, Corollaire 1.2] states that (LSI) holds as a direct consequence, following the arguments in [ABC+00, Théorème 5.4.7]. More generally, for $\mathbb{G} = \mathbb{H}^n$, [HZ10, Theorem 7.3] and [BBBC08, Theorem 6.1] prove that (LSI) holds, again relying on the heat semigroup estimate (5.2). For completeness, we provide a proof in Theorem 5.2 that, for any step-2 Carnot group, the heat semigroup estimate (5.2) implies the log-Sobolev inequality (LSI). Thanks to Theorem 5.1, the $\mathcal{T}_2$ inequality also follows.

Theorem 5.2. *Let $\mathbb{G}$ be a step-2 Carnot group and suppose that there exists $K \in (0, \infty)$ such that the heat semigroup P on $\mathbb{G}$ satisfies the estimate* (5.2) *for all $t \in [0, 1]$. Let $\alpha = \frac{1}{2K^2}$. Then the heat kernel measure $\boldsymbol{\mu}_1$ on $\mathbb{G}$ satisfies the log-Sobolev inequality* (LSI) *with constant α, and $\boldsymbol{\mu}_1$ satisfies the $\mathcal{T}_2$ inequality $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$.*

Proof. Suppose that P_t satisfies (5.2) for all $t \in [0, 1]$. Let $\phi \in C^2(I, \mathbb{R})$ for some interval $I \subset \mathbb{R}$, and suppose moreover that $\phi'' > 0$ and the function $-1/\phi''$ is convex. Let $f \in C_c^\infty(\mathbb{G}, \mathbb{R})$ and let $t \in [0, 1]$. Then, by the heat equation and chain rule for the sub-Laplacian, for any $s \in [0, t]$,

$$\begin{aligned} \partial_s P_s \phi(P_{t-s} f) &= P_s\big(\Delta_{\mathbb{G}} \phi(P_{t-s} f) - \phi'(P_{t-s} f) \Delta_{\mathbb{G}} P_{t-s} f\big) \\ &= P_s\big(\phi''(P_{t-s} f) |\nabla_{\mathbb{G}} P_{t-s} f|_{\mathbb{G}}^2\big). \end{aligned}$$

The heat semigroup estimate (5.2) and the Cauchy–Schwarz inequality imply that

$$\begin{aligned} |\nabla_{\mathbb{G}} P_{t-s} f|_{\mathbb{G}}^2 \leq K^2 \big(P_{t-s}(|\nabla_{\mathbb{G}} f|_{\mathbb{G}})\big)^2 &= K^2 \Big(P_{t-s}\big(|\nabla_{\mathbb{G}} f|_{\mathbb{G}} \sqrt{\phi''(f)} \cdot 1/\sqrt{\phi''(f)}\big)\Big)^2 \\ &\leq K^2 P_{t-s}(|\nabla_{\mathbb{G}} f|_{\mathbb{G}}^2 \phi''(f)) P_{t-s}(1/\phi''(f)), \end{aligned}$$

and, by Jensen's inequality,

$$\phi''(P_{t-s} f) \leq \frac{-1}{P_{t-s}(-1/\phi''(f))} = \frac{1}{P_{t-s}(1/\phi''(f))}.$$

Hence

$$\begin{aligned} \partial_s P_s \phi(P_{t-s} f) &= P_s\big(\phi''(P_{t-s} f) |\nabla_{\mathbb{G}} P_{t-s} f|_{\mathbb{G}}^2\big) \\ &\leq K^2 P_s P_{t-s}(|\nabla_{\mathbb{G}} f|_{\mathbb{G}}^2 \phi''(f)) = K^2 P_t(|\nabla_{\mathbb{G}} f|_{\mathbb{G}}^2 \phi''(f)), \end{aligned}$$

and integrating gives

$$P_t\phi(f) - \phi(P_tf) = \int_0^t \partial_s P_s\phi(P_{t-s}f)\,\mathrm{d}s \le K^2 t P_t(|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}\phi''(f)).$$

Now take $\phi\colon (0,\infty) \to \mathbb{R}$ to be $\phi(x) = x\log x$ for all $x \in (0,\infty)$, and suppose that $f\colon \mathbb{G} \to (0,\infty)$. Then we have the following form of the log-Sobolev inequality:

$$P_t(f\log f) - P_tf\log(P_tf) \le K^2 t P_t\frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f}.$$

To arrive at (LSI), we set $t = 1$, evaluate both sides of the inequality at the identity, and additionally suppose that $\int_{\mathbb{G}} f\,\mathrm{d}\boldsymbol{\mu}_1 = 1$. Then

$$\int_{\mathbb{G}} f\log f\,\mathrm{d}\boldsymbol{\mu}_1 \le K^2\int_{\mathbb{G}}\frac{|\nabla_{\mathbb{G}} f|^2_{\mathbb{G}}}{f}\,\mathrm{d}\boldsymbol{\mu}_1.$$

Note that, allowing $f\colon \mathbb{G} \to [0,\infty)$, we have $\int_{\mathbb{G}} f\log f\,\mathrm{d}\boldsymbol{\mu}_1 = \int_{\{f>0\}} f\log f\,\mathrm{d}\boldsymbol{\mu}_1$. Thus (LSI) holds with constant $\alpha = \frac{1}{2K^2}$.

Applying Theorem 5.1, we further have that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$. ☐

5.3. **Talagrand on Heisenberg-type groups.** For any H-type group $\mathbb{G}$, [Eld10, Theorem 2.4] proves that the heat semigroup estimate (5.2) is satisfied. We thus have the following corollary of Theorem 5.2.

Theorem 5.3. *Let $\mathbb{G}$ be an H-type group. Then there exists $\alpha > 0$ such that the heat kernel measure $\boldsymbol{\mu}_1$ on $\mathbb{G}$ satisfies the log-Sobolev inequality* (LSI) *with constant α, and $\boldsymbol{\mu}_1$ satisfies the $\mathcal{T}_2$ inequality $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{H}, d_{\mathrm{CC}}, \alpha)$.*

Proof. By [Eld10, Theorem 2.4], the estimate (5.2) holds on $\mathbb{G}$ with some constant K. Thus the result follows from Theorem 5.2 with $\alpha = \frac{1}{2K^2}$. ☐

We remark that the best possible constant in Theorem 5.3 is $\alpha \le 1/2$, since [Eld10, Proposition 4.1] shows that the optimal constant in (5.2) satisfies $K \ge \sqrt{\frac{3d_1+5}{3d_1+1}}$.

6. Talagrand for Brownian motion on Carnot groups – Bottom-up approach

In Section 5, we discussed the availability of Talagrand transportation inequalities on Carnot groups, as a consequence of log-Sobolev inequalities and heat kernel estimates. In this section, we demonstrate that we can transfer the $\mathcal{T}_2$ inequality for the heat kernel measure on a Carnot group, via a rescaling and tensorisation argument, to a $\mathcal{T}_2$ inequality on the associated path space; see Section 6.1. We highlight that this approach yields interesting insights into optimal transport problems in the non-commutative sub-Riemannian setting that distinguishes it from the Euclidean case; see Sections 6.2 and 6.3. Finally, we show that the cost function defined in (2.18) on the path space arises naturally as the Γ-limit of discretised cost functions based on the Carnot–Caratheodory distance on the Carnot group; see Section 6.4.

Throughout this section, let $\mathbb{G} \cong \mathbb{R}^{d_1} \oplus \mathbb{R}^{d_2}$ be a step-2 Carnot group and set $d = d_1$; see Section 2.2. Recall that $\boldsymbol{B}$ denotes Brownian motion on $\mathbb{G}$, with law $\boldsymbol{\mu} = \mathrm{Law}(\boldsymbol{B})$ and time marginals $\boldsymbol{\mu}_t = \mathrm{Law}(\boldsymbol{B}_t)$ for $t \in [0,1]$.

6.1. **From Talagrand on Carnot groups to Talagrand on path space.** The main result of this section is that the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}_1$ on the group $\mathbb{G}$ implies the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$ on the space $\boldsymbol{\Omega}_{\mathbb{G}}$ of continuous $\mathbb{G}$-valued paths started from the origin.

We first show that the heat kernel measure satisfies the following scaling property.

Lemma 6.1. *Suppose that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$, for some $\alpha \in (0,\infty)$. Then, for any $t \in (0,1]$, $\boldsymbol{\mu}_t \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha t^{-1})$.*

Proof. We claim that the heat kernel measure on $\mathbb{G}$ satisfies the scaling

$$\boldsymbol{\mu}_t = (\delta_{s^{-1}})_\sharp \boldsymbol{\mu}_{s^2 t}, \tag{6.1}$$

for any $s > 0$, $t \in [0,1]$. To see this, recall that $\mathfrak{p}\colon (0,\infty) \times \mathbb{G} \to (0,\infty)$ denotes the heat kernel on $\mathbb{G}$, Q the homogeneous dimension of $\mathbb{G}$, and $\mathcal{L}^m$ the Lebesgue measure on $\mathbb{G}$. Then, as in (2.8), $(\delta_s)_\sharp \mathcal{L}^m(\mathrm{d}\boldsymbol{x}) = s^{-Q}\mathcal{L}^m(\mathrm{d}\boldsymbol{x})$, for any $s > 0$. Moreover, by [AS20, Theorem 2.3], for any $s > 0$, $t \in [0,1]$, and $\boldsymbol{x} \in \mathbb{G}$, we have that $\mathfrak{p}_{s^2 t}(\delta_s \boldsymbol{x}) = s^{-Q}\mathfrak{p}_t(\boldsymbol{x})$. Therefore, for any Borel set $A \subseteq \mathbb{G}$,

$$\begin{aligned}\mu_t(A) = \int_A \mathfrak{p}_t(\boldsymbol{x})\mathcal{L}^m(\mathrm{d}\boldsymbol{x}) &= s^Q \int_A (\mathfrak{p}_{s^2 t} \circ \delta_s)(\boldsymbol{x})\mathcal{L}^m(\mathrm{d}\boldsymbol{x}) \\ &= s^Q \int_{\delta_{s^{-1}}(A)} \mathfrak{p}_{s^2 t}(\boldsymbol{x})(\delta_s)_\sharp\mathcal{L}^m(\mathrm{d}\boldsymbol{x}) = \int_{\delta_{s^{-1}}(A)} \mathfrak{p}_{s^2 t}(\boldsymbol{x})\mathcal{L}^m(\mathrm{d}\boldsymbol{x}) = (\delta_{s^{-1}})_\sharp \mu_{s^2 t}(A).\end{aligned}$$

This proves the claim. Now fix $t \in (0,1]$. Setting $s = t^{-\frac{1}{2}}$ in (6.1), we have $\boldsymbol{\mu}_t = (\delta_{\sqrt{t}})_\sharp \boldsymbol{\mu}_1$. The map $\delta_{\sqrt{t}}\colon \mathbb{G} \to \mathbb{G}$ is L-Lipschitz with Lipschitz constant $L = \sqrt{t}$. Thus, since $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$, Lemma 2.2 implies that $\boldsymbol{\mu}_t \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha t^{-1})$. □

We now consider the product space $\mathbb{G}^{2^n} = \mathbb{G} \times \cdots \times \mathbb{G}$, for some $n \in \mathbb{N}$, and apply the dimension-free tensorisation property of the $\mathcal{T}_2$ inequality. Define $d_{\mathrm{CC},n}\colon \mathbb{G}^{2^n} \times \mathbb{G}^{2^n} \to [0,\infty)$ by

$$d^2_{\mathrm{CC},n}(\boldsymbol{x}, \overline{\boldsymbol{x}}) \coloneqq 2^n \sum_{i=1}^{2^n} d^2_{\mathrm{CC}}(\boldsymbol{x}_i, \overline{\boldsymbol{x}}_i),$$

for $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}_1, \dots, \overline{\boldsymbol{x}}_n) \in \mathbb{G}^{2^n}$. We have the following tensorisation result.

Proposition 6.2. *Suppose that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$, for some $\alpha \in (0,\infty)$. Then, for any $n \in \mathbb{N}$, $\boldsymbol{\mu}_{2^{-n}}^{\otimes 2^n} \in \mathcal{T}_2(\mathbb{G}^{2^n}, d_{\mathrm{CC},n}, \alpha)$.*

Proof. Fix $n \in \mathbb{N}$. First note that $\boldsymbol{\mu}_{2^{-n}} \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha 2^n)$, by Lemma 6.1. Define $\widetilde{d}_{\mathrm{CC},n}\colon \mathbb{G}^{2^n} \to [0,\infty)$ by

$$\widetilde{d}^2_{\mathrm{CC},n}(\boldsymbol{x}, \overline{\boldsymbol{x}}) = \sum_{i=1}^{2^n} d^2_{\mathrm{CC}}(\boldsymbol{x}_i, \overline{\boldsymbol{x}}_i),$$

for $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}_1, \dots, \overline{\boldsymbol{x}}_n) \in \mathbb{G}^{2^n}$. Since $(\mathbb{G}, d_{\mathrm{CC}})$ is a Polish space, [GL07, Theorem 6] implies that $\mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha 2^n)$ has the dimension-free tensorisation property; i.e. $\boldsymbol{\mu}_{2^{-n}}^{\otimes 2^n} \in \mathcal{T}_2(\mathbb{G}^{2^n}, \widetilde{d}_{\mathrm{CC},n}, \alpha 2^n)$. Applying Lemma 2.2 with ψ equal to the identity, we conclude that $\boldsymbol{\mu}_{2^{-n}}^{\otimes 2^n} \in \mathcal{T}_2(\mathbb{G}^{2^n}, d_{\mathrm{CC},n}, \alpha)$. □

We next prove a relative entropy bound for measures on the path space. For $n \in \mathbb{N}$, set $t^n_k = k 2^{-n}$ for $k \in \{0, \dots, 2^n\}$, and define $\Gamma^n\colon \boldsymbol{\Omega}_{\mathbb{G}} \to \mathbb{G}^{2^n}$ to be the projection of paths to their dyadic increments; i.e. $\Gamma^n \boldsymbol{\omega} = (\boldsymbol{\omega}_{0,t^n_1}, \boldsymbol{\omega}_{t^n_1,t^n_2}, \dots, \boldsymbol{\omega}_{t^n_{2^n-1},1})$, for $\boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbb{G}}$. Then define a cost function $C_n\colon \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}} \to [0,\infty)$ by

$$C_n(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \coloneqq d_{\mathrm{CC},n}(\Gamma^n \boldsymbol{\omega}, \Gamma^n \overline{\boldsymbol{\omega}}), \quad \boldsymbol{\omega}, \overline{\boldsymbol{\omega}} \in \boldsymbol{\Omega}_{\mathbb{G}}. \tag{6.2}$$

Lemma 6.3. *Suppose that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$, for some $\alpha \in (0, \infty)$, and let $\boldsymbol{\nu}$ be a Borel probability measure on $\boldsymbol{\Omega}_{\mathbb{G}}$. Define $\boldsymbol{\mu}^n = \Gamma^n_\sharp \boldsymbol{\mu}$, $\boldsymbol{\nu}^n = \Gamma^n_\sharp \boldsymbol{\nu} \in \mathcal{P}(\mathbb{G}^{2^n})$. Then $H(\boldsymbol{\nu}^n \| \boldsymbol{\mu}^n) \nearrow H(\boldsymbol{\nu} \| \boldsymbol{\mu})$ as $n \to \infty$ and, for any $n \in \mathbb{N}$,*

$$\mathrm{T}_{C_n,2}(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \sqrt{\frac{2}{\alpha} H(\boldsymbol{\nu} \| \boldsymbol{\mu})};$$

i.e. $\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{G}}, C_n, \alpha)$.

Proof. By independence and stationarity of the increments of $\boldsymbol{B}$, we have that $\boldsymbol{\mu}^n = \Gamma^n_\sharp \boldsymbol{\mu} = \boldsymbol{\mu}_{2^{-n}}^{\otimes 2^n}$ and thus, by Proposition 6.2, $\boldsymbol{\mu}^n \in \mathcal{T}_2(\mathbb{G}^{2^n}, d_{\mathrm{CC},n}, \alpha)$.

As shown in [DGW04, Lemma 2.1],

$$H(\boldsymbol{\nu}^n \| \boldsymbol{\mu}^n) = \inf \Big\{ H(\boldsymbol{\eta} \| \boldsymbol{\mu}) : \boldsymbol{\eta} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}}),\, \Gamma^n_\sharp \boldsymbol{\eta} = \boldsymbol{\nu}^n \Big\}, \tag{6.3}$$

and we see that the right-hand side is increasing in n and bounded above by $H(\boldsymbol{\nu} \| \boldsymbol{\mu})$. To see that the limit is equal to $H(\boldsymbol{\nu} \| \boldsymbol{\mu})$, we introduce the piecewise linear interpolation map $R^n \colon \mathbb{G}^{2^n} \to \boldsymbol{\Omega}_{\mathbb{G}}$, which is defined such that $\Gamma^n \circ R^n = \mathrm{id}$, and the image $\mathrm{Im}(R^n) \subset \boldsymbol{\Omega}_{\mathbb{G}}$ is the set of paths that are linear except at the dyadics t^n_k, $k \in \{0, \ldots, 2^n\}$. Define $\widetilde{\boldsymbol{\mu}}^n = R^n_\sharp \boldsymbol{\mu}^n$, $\widetilde{\boldsymbol{\nu}}^n = R^n_\sharp \boldsymbol{\nu}^n \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$. Since $R^n \colon \mathbb{G}^{2^n} \to \mathrm{Im}(R^n)$ is a bijection, applying the representation given in (6.3) for both R^n and its inverse gives the equality $H(\boldsymbol{\nu}^n \| \boldsymbol{\mu}^n) = H(\widetilde{\boldsymbol{\nu}}^n \| \widetilde{\boldsymbol{\mu}}^n)$. We conclude similarly to [AGS08, Corollary 9.4.6], as follows. For any $\boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbb{G}}$, we have that $R^n \circ \Gamma^n(\boldsymbol{\omega}) \to \boldsymbol{\omega}$ as $n \to \infty$ and so, by dominated convergence, $\widetilde{\boldsymbol{\mu}}^n \rightharpoonup \boldsymbol{\mu}$ and $\widetilde{\boldsymbol{\nu}}^n \rightharpoonup \boldsymbol{\nu}$. Then, using the joint lower semicontinuity of the relative entropy (see, e.g. [AGS08, Lemma 9.4.3]) together with the upper bound implied by (6.3), we conclude that

$$\lim_{n \to \infty} H(\boldsymbol{\nu}^n \| \boldsymbol{\mu}^n) = \lim_{n \to \infty} H(\widetilde{\boldsymbol{\nu}}^n \| \widetilde{\boldsymbol{\mu}}^n) = H(\boldsymbol{\nu} \| \boldsymbol{\mu}).$$

Finally, for any $n \in \mathbb{N}$, $\boldsymbol{\mu}^n \in \mathcal{T}_2(\mathbb{G}^{2^n}, d_{\mathrm{CC},n}, \alpha)$ implies that

$$\mathrm{T}^2_{C_n,2}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \mathrm{T}^2_{d_{\mathrm{CC},n},2}(\boldsymbol{\mu}^n, \boldsymbol{\nu}^n) \leq \frac{2}{\alpha} H(\boldsymbol{\nu}^n \| \boldsymbol{\mu}^n) \leq \frac{2}{\alpha} H(\boldsymbol{\nu} \| \boldsymbol{\mu}). \qquad \square$$

Before turning to the main result of this section, we prove an auxiliary lemma on the Euclidean cost on $\mathbb{R}^d$ and the associated Cameron–Martin cost $c_{\mathcal{H}}$ defined in (2.1). For $n \in \mathbb{N}$, define $c_n \colon \Omega \times \Omega \to [0, \infty)$ by

$$c_n^2(\omega, \overline{\omega}) = 2^n \sum_{k=1}^{2^n} |\overline{\omega}_{t^n_{k-1}, t^n_k} - \omega_{t^n_{k-1}, t^n_k}|^2, \tag{6.4}$$

for $\omega, \overline{\omega} \in \Omega$. Part (ii) of the following lemma is a standard stability result from optimal transport and is a consequence, for example, of [Rie17, Lemma 1.1]. Part (iii) will also be used in Proposition 6.21.

Lemma 6.4.

(i) For each $(\omega, \overline{\omega}) \in \Omega \times \Omega$, $(c_n(\omega, \overline{\omega}))_{n \in \mathbb{N}}$ is an increasing sequence, and

$$\lim_{n \to \infty} c_n(\omega, \overline{\omega}) = c_{\mathcal{H}}(\omega, \overline{\omega}). \tag{6.5}$$

(ii) For any $\nu \in \mathcal{P}(\Omega)$, the following convergence holds along a subsequence:

$$\lim_{n \to \infty} \mathrm{T}_{c_n,2}(\mu, \nu) = \mathrm{T}_{c_{\mathcal{H}},2}(\mu, \nu). \tag{6.6}$$

(iii) For any $(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \in \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ *and* $n \in \mathbb{N}$,

$$C_n(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \geq c_n(\pi_1 \boldsymbol{\omega}, \pi_1 \overline{\boldsymbol{\omega}}). \tag{6.7}$$

Remark 6.5. We will show in Theorem 6.6, that the convergence in (6.6) and the lower bound in (6.7) imply the lower estimate $\liminf_{n\to\infty} \mathrm{T}_{C_n,2}(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\mu}, \boldsymbol{\nu})$. An even stronger result will be derived in Section 6.4, namely, we prove the Γ-convergence of the cost functions C_n to the cost $C_{\mathcal{H}}$. This establishes that $C_{\mathcal{H}}$ is indeed the natural limiting cost. Moreover, we prove that the Γ-convergence of C_n implies the convergence of the associated transport problems $\mathrm{T}_{C_n,2}(\boldsymbol{\mu}, \widetilde{\boldsymbol{\nu}}^n)$ along a suitable sequence of probability measures $\widetilde{\boldsymbol{\nu}}^n \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$.

Proof. Let $\omega, \overline{\omega} \in \Omega$ and write $h = \overline{\omega} - \omega$. The sequence $(c_n(\omega, \overline{\omega}))_{n\in\mathbb{N}}$ is increasing by definition. Suppose that $h \in \mathcal{H}$. Then

$$\lim_{n\to\infty} c_n^2(\omega, \overline{\omega}) = \lim_{n\to\infty} 2^n \sum_{k=1}^{2^n} |h_{t_{k-1}^n, t_k^n}|^2 = \int_0^1 |\dot{h}_t|^2 \, \mathrm{d}t = \|h\|_{\mathcal{H}}^2.$$

For $h \notin \mathcal{H}$, the above limit is $+\infty$. This proves part (i).

For part (ii), note that $\mathrm{T}_{c_{\mathcal{H}},2}(\mu, \nu)$ and $\mathrm{T}_{c_n,2}(\mu, \nu)$ admit minimisers, for each $n \in \mathbb{N}$, by e.g. [Vil09, Theorem 4.1], since the cost functions are lower semicontinuous and non-negative. Let $\lambda^* \in \Pi(\mu, \nu)$ attain the infimum in $\mathrm{T}_{c_{\mathcal{H}},2}(\mu, \nu)$. Then, by the monotone convergence theorem,

$$\limsup_{n\to\infty} \mathrm{T}_{c_n,2}^2(\mu, \nu) \leq \limsup_{n\to\infty} \int_{\Omega\times\Omega} c_n^2(\omega, \overline{\omega}) \, \mathrm{d}\lambda^*(\omega, \overline{\omega}) = \int_{\Omega\times\Omega} c_{\mathcal{H}}^2(\omega, \overline{\omega}) \, \mathrm{d}\lambda^*(\omega, \overline{\omega}) = \mathrm{T}_{c_{\mathcal{H}},2}^2(\mu, \nu).$$

On the other hand, for each $n \in \mathbb{N}$, let $\lambda^n \in \Pi(\mu, \nu)$ attain the infimum in $\mathrm{T}_{c_n,2}(\mu, \nu)$. Since $\Pi(\mu, \nu)$ is tight, Prohorov's theorem implies that $(\lambda^n)_{n\in\mathbb{N}}$ converges weakly along a subsequence $(n_k)_{k\in\mathbb{N}}$ to some $\widetilde{\lambda} \in \Pi(\mu, \nu)$. By monotonicity, for any $m \in \mathbb{N}$,

$$\begin{aligned}
\liminf_{k\to\infty} \mathrm{T}_{c_{n_k},2}^2(\mu, \nu) = \liminf_{k\to\infty} \int_{\Omega\times\Omega} c_{n_k}^2(\omega, \overline{\omega}) \, \mathrm{d}\lambda^{n_k}(\omega, \overline{\omega}) &\geq \liminf_{k\to\infty} \int_{\Omega\times\Omega} c_m^2(\omega, \overline{\omega}) \, \mathrm{d}\lambda^{n_k}(\omega, \overline{\omega}) \\
&= \int_{\Omega\times\Omega} c_m^2(\omega, \overline{\omega}) \, \mathrm{d}\widetilde{\lambda}(\omega, \overline{\omega}).
\end{aligned}$$

Applying monotone convergence once more,

$$\begin{aligned}
\liminf_{k\to\infty} \mathrm{T}_{c_{n_k},2}^2(\mu, \nu) \geq \lim_{m\to\infty} \int_{\Omega\times\Omega} c_m^2(\omega, \overline{\omega}) \, \mathrm{d}\widetilde{\lambda}(\omega, \overline{\omega}) &= \int_{\Omega\times\Omega} c_{\mathcal{H}}^2(\omega, \overline{\omega}) \, \mathrm{d}\widetilde{\lambda}(\omega, \overline{\omega}) \\
&\geq \mathrm{T}_{c_{\mathcal{H}},2}^2(\mu, \nu).
\end{aligned}$$

Now observe that, for $\boldsymbol{x}, \overline{\boldsymbol{x}} \in \mathbb{G}$, $d_{\mathrm{CC}}(\boldsymbol{x}, \overline{\boldsymbol{x}}) \geq |\pi_1 \boldsymbol{x} - \pi_1 \overline{\boldsymbol{x}}|$. Indeed, by definition of the Carnot–Carathéodory metric,

$$\begin{aligned}
d_{\mathrm{CC}}(\boldsymbol{x}, \overline{\boldsymbol{x}}) &= \min\Big\{ \int_0^1 |\dot{\gamma}_t| \, \mathrm{d}t : \gamma\colon [0,1] \to \mathbb{G} \text{ horizontal}, \gamma_0 = \boldsymbol{x}, \gamma_1 = \overline{\boldsymbol{x}} \Big\} \\
&\geq \inf\Big\{ \int_0^1 |\dot{\gamma}_t| \, \mathrm{d}t : \gamma\colon [0,1] \to \mathbb{G} \text{ horizontal}, \pi_1\gamma_0 = \pi_1\boldsymbol{x}, \pi_1\gamma_1 = \pi_1\overline{\boldsymbol{x}} \Big\} \\
&= \min\Big\{ \int_0^1 |\dot{g}_t| \, \mathrm{d}t : g\colon [0,1] \to \mathbb{R}^d \text{ absolutely continuous}, \ g_0 = \pi_1\boldsymbol{x}, g_1 = \pi_1\overline{\boldsymbol{x}} \Big\} \\
&= |\pi_1 \boldsymbol{x} - \pi_1 \overline{\boldsymbol{x}}|.
\end{aligned}$$

Hence, for $\boldsymbol{\omega}, \overline{\boldsymbol{\omega}} \in \boldsymbol{\Omega}_{\mathbb{G}}$,

$$C_n^2(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) = 2^n \sum_{k=1}^{2^n} d_{\mathrm{CC}}^2(\boldsymbol{\omega}_{t_{k-1}^n, t_k^n}, \overline{\boldsymbol{\omega}}_{t_{k-1}^n, t_k^n}) \geq 2^n \sum_{k=1}^{2^n} |\pi_1 \overline{\boldsymbol{\omega}}_{t_{k-1}^n, t_k^n} - \pi_1 \boldsymbol{\omega}_{t_{k-1}^n, t_k^n}|^2 = c_n^2(\pi_1 \boldsymbol{\omega}, \pi_1 \overline{\boldsymbol{\omega}}).$$

This concludes part (iii). □

We are now in position to prove the main result of this section.

Theorem 6.6. *Let $\mathbb{G}$ be a step-2 Carnot group and suppose that $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$, for some $\alpha \in (0, \infty)$. Then $\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{G}}, C_{\mathcal{H}}, \alpha)$.*

Proof. If $\boldsymbol{\nu}$ is not absolutely continuous with respect to $\boldsymbol{\mu}$, then $H(\boldsymbol{\nu} \| \boldsymbol{\mu}) = +\infty$ and the cost-information inequality holds trivially.

Now suppose that $\boldsymbol{\nu} \ll \boldsymbol{\mu}$. Recall the lift map $\Psi \colon C([0,1], \mathbb{R}^d) \to C([0,1], \mathbb{G})$ given by (2.15). By Proposition 2.13, $\boldsymbol{\mu} = \Psi_\sharp \mu$, and by Lemma 2.14, there exists $\nu \ll \mu$ such that $\boldsymbol{\nu} = \Psi_\sharp \nu$. Now let $\lambda^* \in \Pi(\mu, \nu)$ be such that

$$\mathrm{T}_{c_{\mathcal{H}},2}^2(\mu, \nu) = \int_{\Omega \times \Omega} c_{\mathcal{H}}^2(\omega, \overline{\omega}) \, \mathrm{d}\lambda^*(\omega, \overline{\omega}) = \int_{\Omega \times \Omega} \|\overline{\omega} - \omega\|_{\mathcal{H}}^2 \, \mathrm{d}\lambda^*(\omega, \overline{\omega}).$$

We have that $\lambda^*(\{ (\omega, \overline{\omega}) \in \Omega \times \Omega : \overline{\omega} - \omega \in \mathcal{H} \}) = 1$ and that $\widetilde{\boldsymbol{\lambda}} = (\Psi \times \Psi)_\sharp \lambda^* \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$ is an admissible coupling. Using the property (2.17) of the lift and shift from Proposition 2.16, we find that

$$\begin{aligned}
\int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C_{\mathcal{H}}^2(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \, \mathrm{d}\widetilde{\boldsymbol{\lambda}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) &= \int_{\Omega \times \Omega} C_{\mathcal{H}}^2(\Psi(\omega), \Psi(\overline{\omega})) \, \mathrm{d}\lambda^*(\omega, \overline{\omega}) \\
&= \int_{\Omega \times \Omega} C_{\mathcal{H}}^2(\Psi(\omega), T_{\overline{\omega} - \omega} \Psi(\omega)) \, \mathrm{d}\lambda^*(\omega, \overline{\omega}) \\
&= \int_{\Omega \times \Omega} \|\overline{\omega} - \omega\|_{\mathcal{H}}^2 \, \mathrm{d}\lambda^*(\omega, \overline{\omega}) = \mathrm{T}_{c_{\mathcal{H}},2}^2(\mu, \nu).
\end{aligned}$$

Hence $\mathrm{T}_{C_{\mathcal{H}},2}^2(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \mathrm{T}_{c_{\mathcal{H}},2}^2(\mu, \nu)$.

Combining (6.6) and (6.7) from Lemma 6.4, we have

$$\limsup_{n \to \infty} \mathrm{T}_{C_n,2}^2(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \limsup_{n \to \infty} \mathrm{T}_{c_n,2}^2(\mu, \nu) \geq \mathrm{T}_{c_{\mathcal{H}},2}^2(\mu, \nu) \geq \mathrm{T}_{C_{\mathcal{H}},2}^2(\boldsymbol{\mu}, \boldsymbol{\nu}).$$

By Lemma 6.3, we conclude that $\mathrm{T}_{C_{\mathcal{H}},2}^2(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \frac{2}{\alpha} H(\boldsymbol{\nu} \| \boldsymbol{\mu})$. □

Theorem 6.7. *Let $\mathbb{G}$ be an H-type group. Then the measure $\boldsymbol{\mu}$ on the space $\boldsymbol{\Omega}_{\mathbb{G}} = C_0([0,1], \mathbb{G})$ satisfies the cost-information inequality*

$$\boldsymbol{\mu} \in \mathcal{T}_2(\boldsymbol{\Omega}_{\mathbb{G}}, C_{\mathcal{H}}, \alpha),$$

for $\alpha > 0$ as in Theorem 5.3.

Proof. By Theorem 5.3, $\boldsymbol{\mu}_1 \in \mathcal{T}_2(\mathbb{G}, d_{\mathrm{CC}}, \alpha)$. We conclude by Theorem 6.6. □

6.2. **Failure of top-down projection and blow-up of cost functions.** In this section, we point out two major differences between the classical Euclidean and the Carnot group settings. In contrast to the Euclidean case, we cannot project the Talagrand inequality from Theorem 6.6 down to a Talagrand inequality for $\boldsymbol{B}_1$. Moreover, the cost functions C_n do not converge pointwise to the cost $C_{\mathcal{H}}$.

We start by giving the corresponding projection result in the Euclidean setting, which we prove via the contraction principle from Lemma 2.2. Let $\widetilde{P}_1\colon \Omega \to \mathbb{R}^{d_1}$, $\omega \mapsto \omega_1$ denote the map that evaluates a path at time $t=1$, and recall the Euclidean Cameron–Martin cost $c_{\mathcal{H}}$ defined in (2.1).

Proposition 6.8. *Let η be a Borel probability measure on Ω and suppose that $\eta \in \mathcal{T}_2(\Omega, c_{\mathcal{H}}, \alpha)$, for some $\alpha \in (0,\infty)$. Then $(\widetilde{P}_1)_\sharp \eta \in \mathcal{T}_2(\mathbb{R}^{d_1}, |.-.|, \alpha)$, where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^{d_1}$.*

Proof. For any $\omega, \overline{\omega} \in \Omega$ such that $\omega - \overline{\omega} \in \mathcal{H}$, Jensen's inequality implies that

$$\|\omega - \overline{\omega}\|_\infty^2 = \sup_{t\in[0,1]} \Big| \int_0^t (\dot{\omega}_s - \dot{\overline{\omega}}_s)\,\mathrm{d}s \Big|^2 \le \sup_{t\in[0,1]} t \int_0^t |\dot{\omega}_s - \dot{\overline{\omega}}_s|^2\,\mathrm{d}s \le \int_0^1 |\dot{\omega}_s - \dot{\overline{\omega}}_s|^2\,\mathrm{d}s = \|\omega - \overline{\omega}\|_{\mathcal{H}}^2$$

Thus, for any $\omega, \overline{\omega} \in \Omega$,

$$|\widetilde{P}_1\omega - \widetilde{P}_1\overline{\omega}|^2 \le \|\omega - \overline{\omega}\|_\infty^2 \le c_{\mathcal{H}}^2(\omega, \overline{\omega}).$$

Hence the contraction principle from Lemma 2.2 yields the claim. ☐

Now consider the law $\boldsymbol{\mu}$ of the Brownian motion $\boldsymbol{B}$ on $\mathbb{G}$. Let $P_1\colon \boldsymbol{\Omega}_{\mathbb{G}} \to \mathbb{G}$ denote the projection of a $\mathbb{G}$-valued path onto its final time evaluation; i.e. $P_1\boldsymbol{\omega} = \boldsymbol{\omega}_1$ for any $\boldsymbol{\omega} \in \mathbb{G}$. Suppose that there exists $\widetilde{\boldsymbol{\Omega}}_{\mathbb{G}} \subseteq \boldsymbol{\Omega}_{\mathbb{G}}$ with $\boldsymbol{\mu}(\widetilde{\boldsymbol{\Omega}}_{\mathbb{G}}) = 1$ and some measurable function $L\colon \boldsymbol{\Omega}_{\mathbb{G}} \to [0,\infty]$ such that

$$d_{\mathrm{CC}}(P_1\boldsymbol{\omega}, P_1\widetilde{\boldsymbol{\omega}}) \le L(\boldsymbol{\omega}) C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}), \tag{6.8}$$

for all $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \widetilde{\boldsymbol{\Omega}}_{\mathbb{G}}$. If $L \in L^\infty(\boldsymbol{\mu})$, then, by Lemma 2.2, the $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}$ implies a $\mathcal{T}_2$ inequality for $\boldsymbol{\mu}_1$. If we only have $L \in L^q(\boldsymbol{\mu})$ for some $q \in [2,\infty)$, then Lemma 2.2 still implies a $\mathcal{T}_p$ inequality for $p = \frac{2q}{2+q} \in [1,2)$. The following result shows that any such L cannot belong to L^q for any $q \in [2,\infty]$, and thus the contraction principle from Lemma 2.2 is not applicable.

Proposition 6.9. *Let $L\colon \boldsymbol{\Omega}_{\mathbb{G}} \to [0,\infty]$ be as in (6.8). Then $\boldsymbol{\mu}(L=\infty) > 0$. In particular, $L \notin L^q(\boldsymbol{\mu})$ for any $q \in (0,\infty]$.*

We make use of the following example in the proof of Proposition 6.9 and again below in the proof of Proposition 6.11.

Lemma 6.10. *Let $a > 0$ and define $h \in \mathcal{H}$ by $h_t = (at, 0, \dots, 0) \in \mathbb{R}^{d_1}$, for all $t \in [0,1]$. Then, for any $s,t \in [0,1]$ with $s \le t$, there exists a standard normal random variable $Z_{s,t}$ such that*

$$d_{\mathrm{CC}}^2(\boldsymbol{B}_{s,t}, (T_h\boldsymbol{B})_{s,t}) \ge aC(t-s)^{\frac{3}{2}} |Z_{s,t}|,$$

for some constant $C > 0$ independent of a, s and t. Moreover, for $u,v,s,t \in [0,1]$ with $u \le v \le s \le t$, the random variables $Z_{u,v}$ and $Z_{s,t}$ are independent.

Proof. For $\boldsymbol{B} = (\boldsymbol{B}^{(1)}, \boldsymbol{B}^{(2)})$, write $\boldsymbol{B}^{(1)} = (B^1, \dots, B^{d_1})$. Let $s,t \in [0,1]$ with $s \le t$. Since h is only non-zero in its first component, Remark 2.18 implies that

$$\boldsymbol{B}_{s,t}^{-1}(T_h\boldsymbol{B})_{s,t} = \bigg(h_{s,t}, a \sum_{j=2}^{d_1} w_{1,j} \int_s^t (s-r)\,\mathrm{d}B_r^j \bigg).$$

Choose $k \in \{1, \dots, m\}$ such that $s_k^2 := \sum_{j=2}^{d_1} |w_{1,j}^k|^2 > 0$. By the left-invariance of d_{CC} and the estimate (2.11), there exists a constant $\kappa > 0$ such that

$$d_{\mathrm{CC}}^2(\boldsymbol{B}_{s,t}, (T_h\boldsymbol{B})_{s,t}) \geq a\kappa^{-1} \left| \sum_{j=2}^{d_1} w_{1,j} \int_s^t (s-r)\,\mathrm{d}B_r^j \right| \geq a\kappa^{-1} \left| \sum_{j=2}^{d_1} w_{1,j}^k \int_s^t (s-r)\,\mathrm{d}B_r^j \right|.$$

By Itô's isometry, we can define a standard normal random variable

$$Z_{s,t} := s_k^{-1}(t-s)^{-\frac{3}{2}}\sqrt{3} \sum_{j=2}^{d_1} w_{1,j}^k \int_s^t (s-r)\,\mathrm{d}B_r^j.$$

Thus, setting $C = 3^{-\frac{1}{2}}\kappa^{-1}s_k$, we have

$$d_{\mathrm{CC}}^2(\boldsymbol{B}_{s,t}, (T_h\boldsymbol{B})_{s,t}) \geq aC(t-s)^{\frac{3}{2}}|Z_{s,t}|.$$

The independence property follows from the independence of Brownian increments. ☐

Proof of Proposition 6.9. Suppose for contradiction that $\boldsymbol{\mu}(L < \infty) = 1$. Let $\boldsymbol{B}$ be a Brownian motion on $\mathbb{G}$, let $\delta > 0$, and define $h \in \mathcal{H}$ by $h_t = (\delta t, 0 \dots, 0) \in \mathbb{R}^{d_1}$, for all $t \in [0,1]$. By Lemma 6.10, there exists a constant $C > 0$ independent of δ and a standard normal random variable Z such that we have the lower bound

$$d_{\mathrm{CC}}^2(P_1\boldsymbol{B}, P_1(T_h\boldsymbol{B})) \geq \delta C|Z|.$$

On the other hand, by definition of the cost $C_{\mathcal{H}}$,

$$C_{\mathcal{H}}^2(\boldsymbol{B}, T_h\boldsymbol{B}) = \|h\|_{\mathcal{H}}^2 = \delta^2.$$

Therefore, (6.8) implies that

$$C\delta|Z| \leq L(\boldsymbol{B})\delta^2.$$

Since both $|Z|$ and $L(\boldsymbol{B})$ are almost surely finite, taking the limit as $\delta \to 0$ gives a contradiction. ☐

We now show that, contrary to the Euclidean case, the cost functions C_n defined in (6.2) may not converge pointwise to $C_{\mathcal{H}}$. Again, we use the example from Lemma 6.10.

Proposition 6.11. *Let $\boldsymbol{B}$ be a Brownian motion on $\mathbb{G}$, and define $h \in \mathcal{H}$ by $h_t = (t, 0, \dots, 0) \in \mathbb{R}^{d_1}$, for all $t \in [0,1]$. Then*

$$\lim_{n\to\infty} C_n(\boldsymbol{B}, T_h\boldsymbol{B}) = \infty$$

almost surely.

Proof. Fix $n \in \mathbb{N}$. By Lemma 6.10, there exist independent standard normal random variables $Z_{n,k}$, $k \in \{1, \dots, 2^n\}$, such that

$$C_n^2(\boldsymbol{B}, T_h\boldsymbol{B}) = 2^n \sum_{k=1}^{2^n} d_{\mathrm{CC}}^2(\boldsymbol{B}_{t_{k-1}^n, t_k^n}, (T_h\boldsymbol{B})_{t_{k-1}^n, t_k^n}) \geq C2^{-n/2} \sum_{k=1}^{2^n} |Z_{n,k}|.$$

Note that $(|Z_{n,k}|)_{k=1,\dots,2^n}$ are independent half-normal random variables with mean $\sqrt{2/\pi}$ and variance $1-2/\pi$. Thus, by Chebyshev's inequality, for any $\delta>0$,

$$\mathbb{P}\left(\left|\sum_{k=1}^{2^n}\frac{|Z_{n,k}|-\sqrt{2/\pi}}{2^n}\right|>\delta\right)\le\delta^{-2}(1-2/\pi)2^{-n}.$$

The right-hand side is summable in n and so, by the first Borel–Cantelli lemma,

$$\lim_{n\to\infty}2^{-n}\sum_{k=1}^{2^n}|Z_{n,k}|=\sqrt{2/\pi}\quad\text{almost surely.}$$

Hence, we have

$$\liminf_{n\to\infty}C_n^2(\boldsymbol{B},T_h\boldsymbol{B})\ge C\liminf_{n\to\infty}2^{n/2}\cdot 2^{-n}\sum_{k=1}^{2^n}|Z_{n,k}|=C\sqrt{2/\pi}\lim_{n\to\infty}2^{n/2}=+\infty,$$

almost surely. □

When the marginals are related by a deterministic shift, we can identify the optimal coupling for the cost $C_{\mathcal{H}}$, and we have the following equality.

Lemma 6.12. *Let $h\in\mathcal{H}$ be deterministic and $\boldsymbol{\nu}=\mathrm{Law}(T_h\boldsymbol{B})$. Then $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\mu},\boldsymbol{\nu})$ admits a unique optimal coupling, this coupling is induced by a Monge map, and*

$$\mathrm{T}^2_{C_{\mathcal{H}},2}(\boldsymbol{\mu},\boldsymbol{\nu})=\mathbb{E}[C^2_{\mathcal{H}}(\boldsymbol{B},T_h\boldsymbol{B})]=2H(\boldsymbol{\nu}\|\boldsymbol{\mu})<\infty.$$

Proof. The second equality follows from Theorem 3.3. Indeed, if $\boldsymbol{\nu}=\mathrm{Law}(T_h\boldsymbol{B})$, then $h=b^{\nu}$ for the Föllmer drift b^{ν} from Theorem 3.3. Since $\boldsymbol{\nu}=\mathrm{Law}(T_h\boldsymbol{B})\ll\boldsymbol{\mu}$, we have that $H(\boldsymbol{\nu}\|\boldsymbol{\mu})<\infty$. Now, from the definition of $C_{\mathcal{H}}$, and the fact that $h\in\mathcal{H}$ is deterministic, we have optimality of $\boldsymbol{\lambda}^*=\mathrm{Law}(\boldsymbol{B},T_h\boldsymbol{B})$. To see this, let $\boldsymbol{\lambda}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ and consider the event $E:=\{\,(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\in\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}:\overline{\boldsymbol{\omega}}=T_{\pi_1\overline{\boldsymbol{\omega}}-\pi_1\boldsymbol{\omega}}\boldsymbol{\omega}\,\}$. If $\boldsymbol{\lambda}(E)<1$, then by definition of $C_{\mathcal{H}}$, we have $\int C^2_{\mathcal{H}}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})=+\infty$. Suppose now that $\boldsymbol{\lambda}(E)=1$. Then, by Jensen's inequality,

$$\begin{aligned}\int C^2_{\mathcal{H}}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})&=\int\|\pi_1\overline{\boldsymbol{\omega}}-\pi_1\boldsymbol{\omega}\|^2_{\mathcal{H}}\,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\\&\ge\left\|\int\pi_1\overline{\boldsymbol{\omega}}\,\mathrm{d}\boldsymbol{\nu}(\overline{\boldsymbol{\omega}})-\int\pi_1\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{\mu}(\boldsymbol{\omega})\right\|^2_{\mathcal{H}}=\|h\|^2_{\mathcal{H}}.\end{aligned}$$

Equality holds if and only if $\boldsymbol{\lambda}(\{\,(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\in\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}:\pi_1\overline{\boldsymbol{\omega}}-\pi_1\boldsymbol{\omega}=h\,\})=1$. Combined with the condition that $\boldsymbol{\lambda}(E)=1$, we see that any optimal coupling is concentrated on the graph of the function $T_h\colon\boldsymbol{\Omega}_{\mathbb{G}}\to\boldsymbol{\Omega}_{\mathbb{G}}$. Thus, there is a unique optimal coupling of Monge form given by $\boldsymbol{\lambda}^*=(\mathrm{id}\times T_h)_\sharp\boldsymbol{\mu}=\mathrm{Law}(\boldsymbol{B},T_h\boldsymbol{B})$, and $\mathrm{T}^2_{C_{\mathcal{H}},2}(\boldsymbol{\mu},\boldsymbol{\nu})=\mathbb{E}[C^2_{\mathcal{H}}(\boldsymbol{B},T_h\boldsymbol{B})]$. □

Remark 6.13. Let $h\in\mathcal{H}$ be as in Proposition 6.11. Since $\boldsymbol{\nu}=\mathrm{Law}(T_h\boldsymbol{B})\ll\boldsymbol{\mu}$, we have that $H(\boldsymbol{\nu}\|\boldsymbol{\mu})<\infty$. By Lemma 6.3, we thus observe that $\mathrm{T}^2_{C_n,2}(\boldsymbol{\mu},\boldsymbol{\nu})\le H(\boldsymbol{\nu}\|\boldsymbol{\mu})<\infty$. However, Proposition 6.11 shows that $\lim_{n\to\infty}C_n(\boldsymbol{B},T_h\boldsymbol{B})=+\infty$ almost surely. Thus $\mathrm{Law}(\boldsymbol{B},T_h\boldsymbol{B})$ is suboptimal for some C_n, $n\in\mathbb{N}$.

In the case of the Heisenberg group $\mathbb{G}=\mathbb{H}^n$, this suboptimality can already be seen for $\mathrm{T}_{d_{\mathrm{CC}},2}(\boldsymbol{\mu}_1,\boldsymbol{\nu}_1)$. Indeed, [AR04, Theorem 5.1] shows that there is a unique optimal coupling and that this coupling is concentrated on the graph of some function $\phi\colon\mathbb{H}^n\to\mathbb{H}^n$. Taking, for example, h as in Lemma 6.10, it is clear that $\mathrm{Law}(\boldsymbol{B}_1,(T_h\boldsymbol{B})_1)$ is not concentrated on any such graph, since $(T_h\boldsymbol{B})_1$ is not measurable with respect to $\sigma(\boldsymbol{B}_1)$.

6.3. **Riemannian approximation of the Heisenberg group.** In the case of classical Wiener space, we consider paths taking values in $\mathbb{R}^d$ with the Euclidean geometry. In the present Carnot group setting, we note the following two distinctions:

(i) $(\mathbb{G}, d_{\mathrm{CC}})$ is a sub-Riemannian metric space (the sub-Laplacian is hypoelliptic);
(ii) the group operation on $\mathbb{G}$ is non-commutative.

It is shown in [CDPT07, Theorem 2.12] that any Carnot group can be approximated by Riemannian manifolds in the sense of pointed Gromov–Hausdorff convergence; see also [AS20, Section 2.5] and, for the Heisenberg group, [AR04, Section 6]. In making this approximation, we move out of the sub-Riemannian setting but retain non-commutativity. We observe that, in this case, a $\mathcal{T}_2$ inequality on path space also holds (Proposition 6.14) and that the blow-up of discretised cost functions shown in Proposition 6.11 does not occur (see Proposition 6.17). The failure to recover the $\mathcal{T}_2$ inequality on the underlying space via projection that was shown in Proposition 6.9 is still observed (Proposition 6.16). However, in contrast to the sub-Riemannian case, we can use the contraction principle to obtain a $\mathcal{T}_p$ inequality on the underlying space for any $p \in [1, 2)$ (Proposition 6.15).

In order to ease the presentation of this section, we specialise to the Heisenberg group $\mathbb{H} = \mathbb{H}^1 \cong \mathbb{R}^2 \oplus \mathbb{R}$. Recall the left-invariant vector fields $(V_1, V_2, V_3) = (X, Y, Z)$, where

$$X = \partial_x + \tfrac{1}{2} y \partial_z, \quad Y = \partial_y - \tfrac{1}{2} x \partial_z, \quad Z = [X, Y] = \partial_z,$$

and the group operation

$$\boldsymbol{x}\boldsymbol{x}' = \big(x + x', y + y', z + z' + \tfrac{1}{2}(xy' - x'y)\big), \quad \boldsymbol{x} = (x, y, z),\ \boldsymbol{x}' = (x', y', z') \in \mathbb{H}.$$

For $\varepsilon > 0$, define the manifold M_ε to be $\mathbb{R}^2 \oplus \mathbb{R}$ equipped with the Euclidean topology and orthonormal basis (X, Y, Z_ε), where $Z_\varepsilon = \varepsilon Z$. Let d_ε denote the induced Riemannian distance, which is again left invariant. By [CDPT07, Theorem 2.12], $(\mathbb{H}, d_{\mathrm{CC}})$ is the limit of the Riemannian manifolds $(M_\varepsilon, d_\varepsilon)$ as $\varepsilon \to 0$, in the sense of pointed Gromov–Hausdorff convergence. As in [AR04, Section 6], we see that, for any $\boldsymbol{x}, \boldsymbol{y} \in M_\varepsilon$,

$$d_\varepsilon(\boldsymbol{x}, \boldsymbol{y}) = \inf\Big\{ \int_0^1 \sqrt{|\dot\gamma_t^1|^2 + |\dot\gamma_t^2|^2 + \varepsilon^{-2} |\dot\gamma_t^3 - \tfrac{1}{2}(\gamma_t^1 \dot\gamma_t^2 - \dot\gamma_t^1 \gamma_t^2)|^2}\, \mathrm{d}t : \gamma \in \mathrm{AC}([0,1], M_\varepsilon),\ \gamma_0 = \boldsymbol{x}, \gamma_1 = \boldsymbol{y} \Big\},$$

and, for $\varepsilon_0, \varepsilon_1 > 0$ with $\varepsilon_1 \le \varepsilon_0$,

$$d_{\varepsilon_0}(\boldsymbol{x}, \boldsymbol{y}) \le d_{\varepsilon_1}(\boldsymbol{x}, \boldsymbol{y}) \le d_{\mathrm{CC}}(\boldsymbol{x}, \boldsymbol{y}) = \sup_{\varepsilon > 0} d_\varepsilon(\boldsymbol{x}, \boldsymbol{y}).$$

Moreover, by [Jui14, Lemma 1.1], there exists a constant $c > 0$ such that, for any $\varepsilon > 0$, $\boldsymbol{x}, \boldsymbol{y} \in M_\varepsilon$,

$$d_{\mathrm{CC}}(\boldsymbol{x}, \boldsymbol{y}) \le d_\varepsilon(\boldsymbol{x}, \boldsymbol{y}) + c\varepsilon. \tag{6.9}$$

We will also make use of the following bounds. There exists a constant $\overline{\kappa} > 0$ such that, for any $\varepsilon > 0$ and $\boldsymbol{x} = (0, 0, z) \in M_\varepsilon$,

$$\overline{\kappa}(|z|^{\frac{1}{2}} - \varepsilon) \le d_\varepsilon(0, \boldsymbol{x}) \le \varepsilon^{-1} |z|, \tag{6.10}$$

where the lower bound follows from (2.11) combined with (6.9), and the upper bound from considering the length of a purely vertical path.

On the space $(M_\varepsilon, d_\varepsilon)$, we consider the same non-commutative group law as on $\mathbb{H}$, but now the distance d_ε is Riemannian.

Consider a Brownian motion B on $\mathbb{R}^3$ with law μ and Cameron–Martin space $\mathcal{H} = W_0^{1,2}([0,1], \mathbb{R}^3)$. We can define a Brownian motion $\boldsymbol{B}^\varepsilon$ on $(M_\varepsilon, d_\varepsilon)$ by

$$\mathrm{d}\boldsymbol{B}^\varepsilon_t = X(\boldsymbol{B}^\varepsilon)\,\mathrm{d}B^1_t + Y(\boldsymbol{B}^\varepsilon)\,\mathrm{d}B^2_t + Z_\varepsilon\,\mathrm{d}B^3_t,$$

and let $\boldsymbol{\mu}^\varepsilon = \mathrm{Law}(\boldsymbol{B}^\varepsilon)$. Explicitly, $\boldsymbol{B}^\varepsilon = (\boldsymbol{B}^{\varepsilon,(1)}, \boldsymbol{B}^{\varepsilon,(2)})$ with

$$\mathrm{d}\boldsymbol{B}^{\varepsilon,(1)}_t = \mathrm{d}(B^1, B^2)_t, \quad \mathrm{d}\boldsymbol{B}^{\varepsilon,(2)}_t = \frac{1}{2}(B^2_t\,\mathrm{d}B^1_t - B^1_t\,\mathrm{d}B^2_t) + \varepsilon\,\mathrm{d}B^3_t.$$

Let $\Omega := C_0([0,1], \mathbb{R}^3)$ and $\boldsymbol{\Omega}^\varepsilon := C_0([0,1], M_\varepsilon)$, and define a map $\Psi^\varepsilon \colon \Omega \to \boldsymbol{\Omega}^\varepsilon$ by

$$\Psi^\varepsilon(\omega) = (0, \varepsilon\omega^3)\Psi((\omega^1, \omega^2)), \quad \omega = (\omega^1, \omega^2, \omega^3) \in \Omega,$$

where Ψ is the lift map defined in Definition 2.12. Define its domain as $\mathrm{Dom}(\Psi^\varepsilon) := \{\, \omega = (\omega^1, \omega^2, \omega^3) \in \Omega : (\omega^1, \omega^2) \in \mathrm{Dom}(\Psi) \,\}$. For ω absolutely continuous, Ψ^ε takes the explicit form

$$\Psi^\varepsilon(\omega)_t = \Big((\omega^1_t, \omega^2_t),\, \frac{1}{2}\int_0^t (\omega^1_r\,\mathrm{d}\omega^2_r - \omega^2_r\,\mathrm{d}\omega^1_r) + \varepsilon\omega^3_t\Big), \quad t \in [0,1].$$

Similarly to Proposition 2.13, we have that $\boldsymbol{B}^\varepsilon = \Psi^\varepsilon(B)$ almost surely. We can also define a shift map $T^\varepsilon_h \colon \boldsymbol{\Omega}^\varepsilon \to \boldsymbol{\Omega}^\varepsilon$, for any $h = (h^1, h^2, h^3) \in \mathcal{H}$, by

$$T^\varepsilon_h \boldsymbol{\omega} = (0, \varepsilon h^3) T_{(h^1,h^2)}\boldsymbol{\omega}, \quad \boldsymbol{\omega} \in \boldsymbol{\Omega}^\varepsilon.$$

Then, similarly to Proposition 2.16, for any $\omega \in \mathrm{Dom}(\Psi^\varepsilon)$ and $h \in \mathcal{H}$, we have

$$T^\varepsilon_h \Psi^\varepsilon(\omega) = \Psi^\varepsilon(\omega + h).$$

Now define a cost function $C^\varepsilon_{\mathcal{H}} \colon \boldsymbol{\Omega}^\varepsilon \times \boldsymbol{\Omega}^\varepsilon \to [0, \infty]$ by

$$C^\varepsilon_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) = \begin{cases} \|h\|_{\mathcal{H}}, & \text{if } \widetilde{\boldsymbol{\omega}} = T^\varepsilon_h \boldsymbol{\omega}, \quad \text{for some } h \in \mathcal{H}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Taking the same approach as in Theorem 4.4, we see that $\boldsymbol{\mu}^\varepsilon$ satisfies a $\mathcal{T}_2$ inequality with this cost.

Proposition 6.14. *We have the cost-information inequality $\boldsymbol{\mu}^\varepsilon \in \mathcal{T}_2(\boldsymbol{\Omega}^\varepsilon, C^\varepsilon_{\mathcal{H}}, 1)$.*

Proof. As noted in Section 2.1, $\mu \in \mathcal{T}_2(\Omega, c_{\mathcal{H}}, 1)$, where the cost $c_{\mathcal{H}}$ is defined in (2.2). We also have that $\boldsymbol{\mu}^\varepsilon = \Psi^\varepsilon_\sharp \mu$ and $\mu(\mathrm{Dom}(\Psi^\varepsilon)) = 1$. Moreover, for any $x, y \in \mathrm{Dom}(\Psi^\varepsilon)$ with $h := y - x \in \mathcal{H}$,

$$C^\varepsilon_{\mathcal{H}}(\Psi^\varepsilon(x), \Psi^\varepsilon(y)) = C^\varepsilon_{\mathcal{H}}(\Psi^\varepsilon(x), \Psi^\varepsilon(x+h)) = C^\varepsilon_{\mathcal{H}}(\Psi^\varepsilon(x), T^\varepsilon_h\Psi^\varepsilon(x)) = \|h\|^2_{\mathcal{H}} = c_{\mathcal{H}}(x,y).$$

In case $y - x \notin \mathcal{H}$, then both sides are infinite. Thus, applying the contraction principle from Lemma 2.2, we have that $\boldsymbol{\mu}^\varepsilon \in \mathcal{T}_2(\boldsymbol{\Omega}^\varepsilon, C^\varepsilon_{\mathcal{H}}, 1)$. ☐

In contrast to the sub-Riemannian setting, $\boldsymbol{\mu}^\varepsilon \in \mathcal{T}_2(\boldsymbol{\Omega}^\varepsilon, C^\varepsilon_{\mathcal{H}}, 1)$ implies a $\mathcal{T}_p$ inequality for $\boldsymbol{\mu}^\varepsilon_1 := \mathrm{Law}(\boldsymbol{B}^\varepsilon_1)$, for $p \in [1,2)$, with a constant depending on ε.

Let $P_1 \colon \boldsymbol{\Omega}^\varepsilon \to M_\varepsilon$ denote the projection $P_1(\boldsymbol{\omega}) = \boldsymbol{\omega}_1$, for $\boldsymbol{\omega} \in \boldsymbol{\Omega}^\varepsilon$, so that $\boldsymbol{\mu}^\varepsilon_1 = (P_1)_\sharp \boldsymbol{\mu}^\varepsilon$.

Proposition 6.15. *Suppose that $\boldsymbol{\mu}^\varepsilon \in \mathcal{T}_2(\boldsymbol{\Omega}^\varepsilon, C^\varepsilon_{\mathcal{H}}, 1)$. Then, for any $p \in [1,2)$, there exists $\alpha(\varepsilon, p) > 0$ such that $\lim_{\varepsilon\to 0}\alpha(\varepsilon,p) = 0$ and $\boldsymbol{\mu}^\varepsilon_1 \in \mathcal{T}_p(M_\varepsilon, d_\varepsilon, \alpha(\varepsilon,p))$.*

Proof. Suppose that $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \boldsymbol{\Omega}^\varepsilon$ with $\widetilde{\boldsymbol{\omega}} = T_h^\varepsilon \boldsymbol{\omega}$, for some $h \in \mathcal{H}$. Define $\boldsymbol{h} = \Psi^\varepsilon(h)$ and $\boldsymbol{\gamma}_1 = \boldsymbol{\omega}_1^{-1}(T_h^\varepsilon \boldsymbol{\omega})_1$. Then

$$d_\varepsilon(P_1(\boldsymbol{\omega}), P_1(\widetilde{\boldsymbol{\omega}})) \leq d_\varepsilon(\boldsymbol{h}_1, \boldsymbol{\gamma}_1) + d_\varepsilon(0, \boldsymbol{h}_1).$$

We bound $d_\varepsilon(0, \boldsymbol{h}_1)$ by the length of the curve $t \mapsto \boldsymbol{h}_t$ in $\boldsymbol{\Omega}^\varepsilon$ to get

$$\begin{aligned} d_\varepsilon(0, \boldsymbol{h}_1) &\leq \int_0^1 \sqrt{|\dot{h}_r^1|^2 + |\dot{h}_r^2|^2 + \varepsilon^{-2}|\varepsilon \dot{h}_r^3 + \tfrac{1}{2}(h_r^1 \dot{h}_r^2 - h_r^2 \dot{h}_r^1) - \tfrac{1}{2}(h_r^1 \dot{h}_r^2 - h_r^2 \dot{h}_r^1)|^2} \, \mathrm{d}r \\ &= \int_0^1 \sqrt{|\dot{h}_r^1|^2 + |\dot{h}_r^2|^2 + |\dot{h}_r^3|^2} \, \mathrm{d}r \leq \|h\|_{\mathcal{H}}. \end{aligned} \tag{6.11}$$

Similarly to Remark 2.18, by integration by parts,

$$\boldsymbol{h}_1^{-1}\boldsymbol{\gamma}_1 = \Big(0, \int_0^1 \Big(h_r^1 \, \mathrm{d}\omega_r^2 - h_r^2 \, \mathrm{d}\omega_r^1\Big)\Big), \tag{6.12}$$

and so, by (6.10),

$$\begin{aligned} d_\varepsilon(\boldsymbol{h}_1, \boldsymbol{\gamma}_1) &\leq \varepsilon^{-1} \Big| \int_0^1 \Big(h_r^1 \, \mathrm{d}\omega_r^2 - h_r^2 \, \mathrm{d}\omega_r^1\Big) \Big| \\ &\leq 2\varepsilon^{-1} \|\omega\|_\infty \|h\|_{\mathcal{H}}, \end{aligned}$$

where $\|\omega\|_\infty \coloneqq \sup_{t \in [0,1]} |(\omega_t^1, \omega_t^2)|$. Hence

$$d_\varepsilon(P_1(\boldsymbol{\omega}), P_1(\widetilde{\boldsymbol{\omega}})) \leq (1 + 2\varepsilon^{-1}\|\omega\|_\infty)\|h\|_{\mathcal{H}} = (1 + 2\varepsilon^{-1}\|\omega\|_\infty) C_{\mathcal{H}}^\varepsilon(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}).$$

In the case that there does not exist $h \in \mathcal{H}$ such that $\widetilde{\boldsymbol{\omega}} = T_h \boldsymbol{\omega}$, then the same inequality holds trivially.

Next note that, for any $q \in [1, \infty)$, $\boldsymbol{\omega} \mapsto \|\omega\|_\infty \in L^q(\boldsymbol{\mu})$. Let $p \in [1, 2)$ and set $q = \frac{2p}{2-p} \in [2, \infty)$. By the contraction principle from Lemma 2.2, we conclude that

$$\boldsymbol{\mu}_1 \in \mathcal{T}_p(M_\varepsilon, d_\varepsilon, \alpha(\varepsilon, p)), \quad \text{where} \quad \alpha(\varepsilon, p) = \big(1 + 2\varepsilon^{-1}\mathbb{E}[\|B\|_\infty^q]^{\frac{1}{q}}\big)^{-2},$$

and we see that $\lim_{\varepsilon \to 0} \alpha(\varepsilon, p) = 0$. □

Analogously to the sub-Riemannian setting, however, it is not possible to recover a $\mathcal{T}_2$ inequality via the contraction principle, as the next result shows.

Proposition 6.16. *Let $\widetilde{\boldsymbol{\Omega}} \subseteq \boldsymbol{\Omega}^\varepsilon$ such that $\boldsymbol{\mu}(\widetilde{\boldsymbol{\Omega}}) = 1$, and let $L \colon \boldsymbol{\Omega}^\varepsilon \to [0, \infty]$ be a measurable function such that*

$$d_\varepsilon(P_1\boldsymbol{\omega}, P_1\widetilde{\boldsymbol{\omega}}) \leq L(\boldsymbol{\omega}) C_{\mathcal{H}}^\varepsilon(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}),$$

for all $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \widetilde{\boldsymbol{\Omega}}$. Then $L \notin L^\infty(\boldsymbol{\mu}^\varepsilon)$.

Proof. Suppose for a contradiction that $L \in L^\infty(\boldsymbol{\mu}^\varepsilon)$. Define $h \in \mathcal{H}$ by $h_t = (t, 0, 0) \in \mathbb{R}^3$, for all $t \in [0, 1]$. Then, similarly to Lemma 6.10, we can apply (6.12) and (6.10) to see that there exists a standard normal random variable Z and a constant $C(\varepsilon) > 0$ such that

$$d_\varepsilon^2(\boldsymbol{B}_1^\varepsilon, (T_h^\varepsilon \boldsymbol{B}^\varepsilon)_1) \geq C(\varepsilon)|Z|.$$

We also have $C_{\mathcal{H}}^\varepsilon(\boldsymbol{B}^\varepsilon, T_h^\varepsilon \boldsymbol{B}^\varepsilon) = \|h\|_{\mathcal{H}} = 1$. Thus $C(\varepsilon)|Z| \leq L(\boldsymbol{B}^\varepsilon)$. Since $|Z|$ is not essentially bounded, we arrive at a contradiction. □

Finally, in the Riemannian setting, we do not observe the blow-up shown in Proposition 6.11. For $n \in \mathbb{N}$, define a cost $C_n^\varepsilon \colon \boldsymbol{\Omega}^\varepsilon \times \boldsymbol{\Omega}^\varepsilon \to [0,\infty)$ by

$$C_n^\varepsilon(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})^2 = 2^n \sum_{k=1}^{2^n} d_\varepsilon^2(\boldsymbol{\omega}_{t_{k-1}^n, t_k^n}, \widetilde{\boldsymbol{\omega}}_{t_{k-1}^n, t_k^n}), \quad \boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \boldsymbol{\Omega}^\varepsilon.$$

Proposition 6.17. *For any $h \in \mathcal{H}$, we have*

$$\limsup_{n\to\infty} C_n^\varepsilon(\boldsymbol{B}^\varepsilon, T_h^\varepsilon \boldsymbol{B}^\varepsilon) \le C_{\mathcal{H}}^\varepsilon(\boldsymbol{B}^\varepsilon, T_h^\varepsilon \boldsymbol{B}^\varepsilon) = \|h\|_{\mathcal{H}},$$

almost surely.

Proof. Define $\boldsymbol{Z}$ by $\boldsymbol{Z}_{s,t} = \boldsymbol{B}_{s,t}^{-1}(T_h^\varepsilon \boldsymbol{B})_{s,t}$, for $s,t \in [0,1]$ with $s \le t$, and $\boldsymbol{h} = \Psi^\varepsilon(h)$. Let $n \in \mathbb{N}$. By Young's inequality, we bound

$$C_n^\varepsilon(\boldsymbol{B}^\varepsilon, T_h^\varepsilon \boldsymbol{B}^\varepsilon)^2 \le (1+n) C_n^\varepsilon(\boldsymbol{h}, \boldsymbol{Z})^2 + (1+\tfrac{1}{n}) C_n^\varepsilon(0,\boldsymbol{h})^2. \tag{6.13}$$

As in (6.11), we bound $d_\varepsilon(\boldsymbol{h}_{t_{k-1}^n}, \boldsymbol{h}_{t_k^n})$ by the length of the curve $\boldsymbol{h}$; i.e.

$$d_\varepsilon(\boldsymbol{h}_{t_{k-1}^n}, \boldsymbol{h}_{t_k^n}) \le \int_{t_{k-1}^n}^{t_k^n} \sqrt{|\dot{h}_r^1|^2 + |\dot{h}_r^2|^2 + |\dot{h}_r^3|^2}\, \mathrm{d}r.$$

Applying the Cauchy–Schwarz inequality, we have

$$C_n^\varepsilon(0,\boldsymbol{h})^2 = 2^n \sum_{k=1}^{2^n} d_\varepsilon^2(\boldsymbol{h}_{t_{k-1}^n}, \boldsymbol{h}_{t_k^n}) \le \sum_{k=1}^{2^n} \int_{t_{k-1}^n}^{t_k^n} \Big(|\dot{h}_r^1|^2 + |\dot{h}_r^2|^2 + |\dot{h}_r^3|^2\Big) \mathrm{d}r = \|h\|_{\mathcal{H}}^2,$$

and so

$$\limsup_{n\to\infty} (1+\tfrac{1}{n}) C_n^\varepsilon(0,\boldsymbol{h})^2 \le \|h\|_{\mathcal{H}}^2.$$

Next, similarly to (6.12) and Remark 2.18, we have

$$\boldsymbol{h}_{s,t}^{-1} \boldsymbol{Z}_{s,t} = \Big(0, \int_s^t \Big(h_{s,r}^1 \,\mathrm{d}B_r^2 - h_{s,r}^2 \,\mathrm{d}B_r^1\Big)\Big).$$

Therefore, using the estimate (6.10) and the fact that B is almost surely β-Hölder continuous for any $\beta \in (0,1/2)$, there exists a constant $c > 0$ such that we have the almost sure bound

$$\begin{aligned} d_\varepsilon^2(0, \boldsymbol{h}_{s,t}^{-1}\boldsymbol{Z}_{s,t}) &\le \varepsilon^{-2} \Big| \int_s^t h_{s,r}^1 \,\mathrm{d}B_r^2 \Big|^2 + \varepsilon^{-2} \Big| \int_s^t h_{s,r}^2 \,\mathrm{d}B_r^1 \Big|^2 \\ &\le c\varepsilon^{-2} \|B\|_\beta^2 |t-s|^{1+2\beta} \int_s^t |\dot{h}_r|^2 \,\mathrm{d}r, \end{aligned}$$

where $\|B\|_\beta$ is the β-Hölder norm of B. Hence

$$C_n^\varepsilon(\boldsymbol{h}, \boldsymbol{Z})^2 = 2^n \sum_{k=1}^{2^n} d_\varepsilon^2(0, \boldsymbol{h}_{t_{k-1}^n, t_k^n}^{-1} \boldsymbol{Z}_{t_{k-1}^n, t_k^n}) \le c\varepsilon^{-2} 2^n 2^{-(1+2\beta)n} \|h\|_{\mathcal{H}}^2 \|B\|_\beta^2 = c\varepsilon^{-2} \|h\|_{\mathcal{H}}^2 \|B\|_\beta^2 2^{-2\beta n},$$

and so $\lim_{n\to\infty} (1+n) C_n^\varepsilon(\boldsymbol{h}, \boldsymbol{Z})^2 = 0$ almost surely. We conclude by (6.13). ☐

6.4. **Γ-convergence of the cost functions.** Despite the pointwise blow-up of the cost functions C_n that we demonstrated in Proposition 6.11, we now show that C_n does converge to $C_{\mathcal{H}}$ in a variational sense. More precisely, the sequence C_n converges to $C_{\mathcal{H}}$ in the sense of Γ-convergence, a notion of convergence for families of minimisation problems that is formulated in terms of asymptotic lower and upper bounds. On a metric space (E, d_E), we say that a sequence of functionals $F_n\colon E \to \mathbb{R} \cup \{\infty\}$ Γ-converges to a limit $F_\infty\colon E \to \mathbb{R} \cup \{\infty\}$ if

(i) for every sequence $x_n \to x$ in E, we have $F_\infty(x) \le \liminf_{n\to\infty} F_n(x_n)$; and
(ii) for every $x \in E$, there exists a sequence $\widetilde{x}_n \to x$ in E such that $\limsup_{n\to\infty} F_n(\widetilde{x}_n) \le F_\infty(x)$.

The sequence $(\widetilde{x}_n)$ in condition (ii) is called a *recovery sequence* as it "recovers" the correct energy level $F_\infty(x)$ from the approximating energies $F_n(\widetilde{x}_n)$ by adding suitable oscillations to x. One may view Γ-convergence as describing the asymptotic behavior of energy landscapes, in close analogy with large deviation principles, which characterise the asymptotics of probability measures via their rate functions. Indeed, the interplay between Γ-convergence and large deviation principles has been studied in several publications; see e.g. [Mar18, Ber18]. A central advantage of Γ-convergence is its stability property: convergence of functionals implies convergence of minimal values and, under mild compactness assumptions, convergence of (almost) minimisers. For a comprehensive treatment we refer to the monographs [DM93, Bra02, Rin18].

We will see that the Γ-convergence of the cost functions C_n implies the convergence of the associated optimal transport problems; i.e. for every $\boldsymbol{\nu} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ there exists a sequence of probability measures $\widetilde{\boldsymbol{\nu}}^n \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ such that the associated transport costs $\mathrm{T}_{C_n,2}(\widetilde{\boldsymbol{\nu}}^n, \boldsymbol{\mu})$ converges to $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\nu}, \boldsymbol{\mu})$ as $n \to \infty$.

In the Euclidean case with cost c_n defined as in (6.4), we have for $\omega, \widetilde{\omega} \in C_0([0,1];\mathbb{R}^d)$ with $h = \omega - \widetilde{\omega}$ the formula $c_n(\omega, \widetilde{\omega}) = \int_0^1 |\partial_s \widehat{h}^n|^2 \,\mathrm{d}s$, where $\widehat{h}^n$ is the piecewise affine interpolant for h. One readily verifies that $c_n(\omega, \widetilde{\omega})$ converges to $\|h\|_{\mathcal{H}}$ if $h \in \mathcal{H}$ and to $+\infty$ otherwise; cf. Lemma 6.4. In particular, the Γ-convergence of c_n also holds in this setting with the same limiting cost.

Recall that we consider the metric space $(\boldsymbol{\Omega}_{\mathbb{G}}, d_\infty)$, where d_∞ is the uniform metric defined in (2.13).

We start with the following lemma, which gives the pointwise convergence of the cost to the Cameron–Martin norm for horizontal curves.

Lemma 6.18. *Let $h \in \mathcal{H}$, and let $\boldsymbol{h} = \Psi(h)$ denote its lift to $\boldsymbol{\Omega}_{\mathbb{G}}$. Then, for the family of cost functions $(C_n)_{n\in\mathbb{N}}$ defined in* (6.2)*, we have $C_n(0, \boldsymbol{h}) \le \|h\|_{\mathcal{H}}$ and $\lim_{n\to\infty} C_n(0, \boldsymbol{h}) = \|h\|_{\mathcal{H}}$.*

Proof. Let $h \in \mathcal{H}$ and let $\boldsymbol{h} = \Psi(h)$. The curve $t \mapsto \boldsymbol{h}_t \in \mathbb{G}$ is horizontal and therefore, for every $0 \le s < t \le 1$, we have

$$d_{\mathrm{CC}}(\boldsymbol{h}_s, \boldsymbol{h}_t) \le \int_s^t |\dot{h}_r| \,\mathrm{d}r.$$

Applying Hölder's inequality, we obtain the estimate

$$C_n^2(0, \boldsymbol{h}) = 2^n \sum_{i=1}^{2^n} d_{\mathrm{CC}}(\boldsymbol{h}_{t_i^n}, \boldsymbol{h}_{t_{i-1}^n})^2 \le \sum_{i=1}^{2^n} \int_{t_{i-1}^n}^{t_i^n} |\dot{h}_r|^2 \,\mathrm{d}r = \|h\|_{\mathcal{H}}^2.$$

Taking the lim sup on the left-hand side gives $\limsup_{n\to\infty} C_n(0, \boldsymbol{h}) \le \|h\|_{\mathcal{H}}$.

To show the lower bound, define a piecewise constant function $g^n\colon [0,1] \to [0,\infty)$, for each $n \in \mathbb{N}$, by $g_t^n = 2^n d_{\mathrm{CC}}(\boldsymbol{h}_{t_i^n}, \boldsymbol{h}_{t_{i-1}^n})$ for $t \in [t_{i-1}^n, t_i^n)$, $i \in \{1, \ldots, n\}$. Note that the sequence (g^n) is uniformly bounded in $L^2([0,1])$. Hence, we can extract a weakly converging subsequence such that $g^{n_k} \to g$ in $L^2([0,1])$. For given $0 \le r < s \le 1$, we can find indices i_n, $j_n \in \{1, \ldots, 2^n\}$ such that, for $r_n = t_{i_n}^n$ and $s_n = t_{j_n}^n$,

$$0 \le r_n \le r < s \le s_n \le 1, \quad \text{and} \quad \lim_{n\to\infty} r_n = r, \quad \lim_{n\to\infty} s_n = s.$$

By the triangle inequality and the continuity of $t \mapsto \boldsymbol{h}_t$, there is a sequence $(\varepsilon_n) \subset \mathbb{R}$ such that $\varepsilon_n \to 0$ and

$$d_{\mathrm{CC}}(\boldsymbol{h}_r, \boldsymbol{h}_s) \le \sum_{i=i_n}^{j_n} d_{\mathrm{CC}}(\boldsymbol{h}_{t_i^n}, \boldsymbol{h}_{t_{i-1}^n}) + d_{\mathrm{CC}}(\boldsymbol{h}_r, \boldsymbol{h}_{r_n}) + d_{\mathrm{CC}}(\boldsymbol{h}_s, \boldsymbol{h}_{s_n}) = \int_{r_n}^{s_n} g_t^n \,\mathrm{d}t + \varepsilon_n.$$

Passing to the limit as $n \to \infty$, we infer that $d_{\mathrm{CC}}(\boldsymbol{h}_r, \boldsymbol{h}_s) \le \int_r^s g_t \,\mathrm{d}t$. Now, by the minimality of the metric derivative (see Remark 2.8), we obtain $g_t \ge |\dot{h}_t|$ for almost every $t \in (0,1)$. Finally, since $C_n^2(0, \boldsymbol{h}) = 2^n \sum_{i=1}^{2^n} d_{\mathrm{CC}}^2(\boldsymbol{h}_{t_i^n}, \boldsymbol{h}_{t_{i-1}^n})$, lower-semicontinuity of the norm implies that

$$\liminf_{n\to\infty} C_n^2(0, \boldsymbol{h}) = \liminf_{n\to\infty} \int_0^1 (g_t^n)^2 \,\mathrm{d}t \ge \int_0^1 g_t^2 \,\mathrm{d}t \ge \|h\|_{\mathcal{H}}^2,$$

which finishes the proof. □

The next lemma shows that the cost C_n blows up along sequences $\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n$ that converge to limits $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}$ whose difference $\boldsymbol{\omega}^{-1}\widetilde{\boldsymbol{\omega}}$ is a purely vertical process; i.e. $t \mapsto (\boldsymbol{\omega}^{-1}\widetilde{\boldsymbol{\omega}})_t = (0, \theta_t)$.

Lemma 6.19. *Let $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \boldsymbol{\Omega}_{\mathbb{G}}$, and let $(\boldsymbol{\omega}^n), (\widetilde{\boldsymbol{\omega}}^n) \subset \boldsymbol{\Omega}_{\mathbb{G}}$ be sequences such that $\lim_{n\to\infty}(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) = (\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$. Suppose that there exists a non-zero $\boldsymbol{\theta} \in \boldsymbol{\Omega}_{\mathbb{G}}$ such that $\boldsymbol{\theta}_t = (0, \theta_t)$ and $\widetilde{\boldsymbol{\omega}}_t = \boldsymbol{\omega}_t \boldsymbol{\theta}_t$ for all $t \in [0, T]$. Then $\lim_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) = \infty$.*

Proof. Suppose for contradiction that $C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n)$ is bounded uniformly in n. We use the equivalence of the gauge distance d_g and the Carnot–Carathéodory distance d_{CC} from (2.11) to obtain the lower bound

$$C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n)^2 = 2^n \sum_{k=1}^{2^n} d_{\mathrm{CC}}\big(\boldsymbol{\omega}^n_{t^n_{k-1}, t^n_k}, \widetilde{\boldsymbol{\omega}}^n_{t^n_{k-1}, t^n_k}\big)^2 \ge \frac{2^n}{\kappa} \sum_{k=1}^{2^n} d_g\big(\boldsymbol{\omega}^n_{t^n_{k-1}, t^n_k}, \widetilde{\boldsymbol{\omega}}^n_{t^n_{k-1}, t^n_k}\big)^2 \ge \frac{2^n}{\kappa} \sum_{k=1}^{2^n} \big|\theta^n_{t^n_{k-1}, t^n_k}\big|,$$

where $\theta^n_{s,t} = \pi_2((\boldsymbol{\omega}^n_{s,t})^{-1} \widetilde{\boldsymbol{\omega}}^n_{s,t})$ are the increments of the vertical process. Let $\widehat{\theta}^n_t$ denote the piecewise affine interpolant associated with the increments $\theta^n_{t^n_{k-1}, t^n_k}$ with $\widehat{\theta}^n_0 = 0$. The above estimate gives the bound $C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n)^2 \ge \frac{2^n}{\kappa} \|\widehat{\theta}^n\|_{W^{1,1}}$, where the left-hand side is uniformly bounded with respect to n by assumption. Thus $\widehat{\theta}^n \to 0$ in $W_0^{1,1}([0,1], \mathbb{R}^{d_2})$ and so $\theta \equiv 0$ as $W_0^{1,1}([0,1], \mathbb{R}^{d_2}) \hookrightarrow C_0([0,1], \mathbb{R}^{d_2})$, giving a contradiction. We conclude that $\lim_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) = \infty$. □

Now we prove condition (ii) in the definition of Γ-convergence for C_n, i.e. the existence of a recovery sequence. In fact, we show a stronger version; see Remark 6.23.

Proposition 6.20. *Let $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \boldsymbol{\Omega}_{\mathbb{G}}$ and let $(\boldsymbol{\omega}^n) \subset \boldsymbol{\Omega}_{\mathbb{G}}$ be a sequence such that $\boldsymbol{\omega}^n \to \boldsymbol{\omega}$. Then there exists a sequence $(\widetilde{\boldsymbol{\omega}}^n) \subset \boldsymbol{\Omega}_{\mathbb{G}}$ such that $\widetilde{\boldsymbol{\omega}}^n \to \widetilde{\boldsymbol{\omega}}$ and*

$$\limsup_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) \le C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}). \tag{6.14}$$

Proof. We only have to consider the case $\widetilde{\boldsymbol{\omega}} = T_h \boldsymbol{\omega}$ for $h \in \mathcal{H}$, since the right-hand side in (6.14) is otherwise infinite by definition of $C_{\mathcal{H}}$ in (2.18), and the inequality holds trivially. Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbb{G}}$, $h \in \mathcal{H}$, and $\widetilde{\boldsymbol{\omega}} = T_h \boldsymbol{\omega}$. Consider a sequence $(\boldsymbol{\omega}^n) \subset \boldsymbol{\Omega}_{\mathbb{G}}$ such that $\boldsymbol{\omega}^n \to \boldsymbol{\omega}$ in $\boldsymbol{\Omega}_{\mathbb{G}}$. For each $n \in \mathbb{N}$, define $\overline{\boldsymbol{\omega}}^n := T_h \boldsymbol{\omega}^n \in \boldsymbol{\Omega}_{\mathbb{G}}$. By the continuity of the shift map from Proposition 2.16, $\overline{\boldsymbol{\omega}}^n \to \widetilde{\boldsymbol{\omega}} = T_h \boldsymbol{\omega}$ in $\boldsymbol{\Omega}_{\mathbb{G}}$. As in Remark 2.18, we introduce the non-commutativity error

$$\boldsymbol{\theta}^n_{s,t} := \boldsymbol{h}^{-1}_{s,t} (\boldsymbol{\omega}^n_{s,t})^{-1} \overline{\boldsymbol{\omega}}^n_{s,t} = \big(0, \theta^n_{s,t}\big),$$

$$\text{where} \quad \theta^n_{s,t} = \Big(\int_s^t \mathsf{W} h_{s,r} \otimes \mathrm{d}\omega^n_r \Big).$$

We define $\boldsymbol{\vartheta}^n \in \boldsymbol{\Omega}_{\mathbb{G}}$ such that its increments satisfy $(\boldsymbol{\vartheta}^n_{t^n_{k-1}})^{-1}\boldsymbol{\vartheta}^n_{t^n_k} = \boldsymbol{\theta}^n_{t^n_{k-1},t^n_k}$. Indeed, we set $\boldsymbol{\vartheta}^n_t = (0, \vartheta^n_t)$, where

$$\vartheta^n_0 = 0 \quad \text{and} \quad \vartheta^n_t = \vartheta^n_{t^n_{k-1}} + \theta^n_{t^n_{k-1},t}, \quad \text{for} \quad t \in (t^n_{k-1}, t^n_k],\ k \in \{1, \dots, 2^n\}.$$

We emphasise that, for any $t \in [0,1]$, $\boldsymbol{\vartheta}^n_t$ is an element in the centre of the group $\mathbb{G}$. In particular, it commutes with every element in $\mathbb{G}$. Therefore, defining the curve $t \mapsto \widetilde{\boldsymbol{\omega}}^n_t = \overline{\boldsymbol{\omega}}^n_t(\boldsymbol{\vartheta}^n_t)^{-1} \in \mathbb{G}$, we obtain that its increments satisfy

$$\widetilde{\boldsymbol{\omega}}^n_{s,t} = \overline{\boldsymbol{\omega}}^n_{s,t}(\boldsymbol{\vartheta}^n_{s,t})^{-1} = (\boldsymbol{\vartheta}^n_{s,t})^{-1}\overline{\boldsymbol{\omega}}^n_{s,t}, \quad 0 \le s \le t \le 1.$$

Using the left-invariance of d_{CC} and the definition of $\boldsymbol{\vartheta}^n$, we find that, for $k \in \{1, \dots, 2^n\}$,

$$\begin{aligned} d_{\mathrm{CC}}(\boldsymbol{\omega}^n_{t^n_{k-1},t^n_k}, \widetilde{\boldsymbol{\omega}}^n_{t^n_{k-1},t^n_k}) &= d_{\mathrm{CC}}\big(0, (\boldsymbol{\omega}^n_{t^n_{k-1},t^n_k})^{-1}\overline{\boldsymbol{\omega}}^n_{t^n_{k-1},t^n_k}(\boldsymbol{\vartheta}^n_{t^n_{k-1},t^n_k})^{-1}\big) \\ &= d_{\mathrm{CC}}\big(0, \boldsymbol{h}_{t^n_{k-1},t^n_k}\boldsymbol{\theta}^n_{t^n_{k-1},t^n_k}(\boldsymbol{\theta}^n_{t^n_{k-1},t^n_k})^{-1}\big) = d_{\mathrm{CC}}(0, \boldsymbol{h}_{t^n_{k-1},t^n_k}). \end{aligned}$$

Lemma 6.18 now allows us to pass to the lim sup. More precisely, we have that

$$\limsup_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) = \limsup_{n\to\infty} C_n(0, \boldsymbol{h}) \le C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}).$$

It remains to show that $\widetilde{\boldsymbol{\omega}}^n \to \widetilde{\boldsymbol{\omega}}$. Due to the convergence $\overline{\boldsymbol{\omega}}^n \to \widetilde{\boldsymbol{\omega}}$ it suffices to show that $\boldsymbol{\vartheta}^n \to 0$ in $\boldsymbol{\Omega}_{\mathbb{G}}$ or, equivalently, $\vartheta^n \to 0$ in $C([0,1], \mathbb{R}^{d_2})$. Using the definition of ϑ^n, we find a constant $C > 0$ such that

$$\|\vartheta^n\|_\infty \le C \sup_{k=1,\dots,2^n} \max_{i<j} \sup_{t\in[t^n_{k-1},t^n_k]} |\omega^{n,i}_t - \omega^{n,i}_{t^n_k}| \big\|\dot h^j\big\|_{L^1}.$$

Since $\boldsymbol{\omega}^n$ converges uniformly to $\boldsymbol{\omega}$, the right-hand side vanishes as $n \to \infty$. □

We now prove the lower estimate that is required in condition (i) of the definition of Γ-convergence.

Proposition 6.21. *Let* $\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}} \in \boldsymbol{\Omega}_{\mathbb{G}}$*, and let* $(\boldsymbol{\omega}^n), (\widetilde{\boldsymbol{\omega}}^n) \subset \boldsymbol{\Omega}_{\mathbb{G}}$ *be sequences such that* $\lim_{n\to\infty}(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) = (\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$*. Then* $\liminf_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) \ge C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$.

Proof. Consider a pair of curves $(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) \in \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ and a pair of sequences $(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) \subset \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ such that $\lim_{n\to\infty}(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) = (\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$. We will consider three cases.

Case 1a. First, we consider the case that $\widetilde{\boldsymbol{\omega}} = T_h\boldsymbol{\omega}$ for some $h \in \mathcal{H}$. We may assume that $I := \liminf_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) < \infty$, since otherwise the inequality holds trivially. Let $h^n = \pi_1\widetilde{\boldsymbol{\omega}}^n - \pi_1\boldsymbol{\omega}^n$ so that $h^n \to h$ in $C([0,1], \mathbb{R}^{d_1})$. We now apply Lemma 6.4 (iii), to see that

$$\liminf_{n\to\infty} C_n(\widetilde{\boldsymbol{\omega}}^n, \boldsymbol{\omega}^n)^2 \ge \liminf_{n\to\infty} 2^n \sum_{k=1}^{2^n} |h^n_{t^n_k} - h^n_{t^n_{k-1}}|^2 = \liminf_{n\to\infty} \|\widehat{h}^n\|^2_{\mathcal{H}}, \tag{6.15}$$

where $\widehat{h}^n$ is the piecewise affine interpolant of h^n. Since $I \in [0,\infty)$, we can assume that $\widehat{h}^n$ is bounded in $\mathcal{H}$ and is weakly converging to a limit $\overline{h} \in \mathcal{H}$, which we see is equal to h. By weak lower semicontinuity of the L^2 norm we obtain $\liminf_{n\to\infty} C_n(\widetilde{\boldsymbol{\omega}}^n, \boldsymbol{\omega}^n)^2 \ge \|h\|^2_{\mathcal{H}}$.

Case 1b. Now suppose that $\widetilde{\boldsymbol{\omega}} = T_h\boldsymbol{\omega}$, where $h = \pi_1\widetilde{\boldsymbol{\omega}} - \pi_1\boldsymbol{\omega} \notin \mathcal{H}$. Then $C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) = \infty$. Supposing again that $I := \liminf_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) < \infty$, following the same argument as above leads to $h \in \mathcal{H}$, which is a contradiction. Thus, $\liminf_{n\to\infty} C_n(\widetilde{\boldsymbol{\omega}}^n, \boldsymbol{\omega}^n)^2 = \infty$.

Case 2. We now assume that $h := \pi_1\widetilde{\boldsymbol{\omega}} - \pi_1\boldsymbol{\omega} \in \mathcal{H}$ but $\widetilde{\boldsymbol{\omega}} \ne T_h\boldsymbol{\omega}$; i.e. $\widetilde{\boldsymbol{\omega}}$ is not a shift of $\boldsymbol{\omega}$. We show that $C_n(\widetilde{\boldsymbol{\omega}}^n, \boldsymbol{\omega}^n) \to \infty$. Define $\widehat{\boldsymbol{\omega}} = T_h\boldsymbol{\omega}$ such that, by assumption, $\widehat{\boldsymbol{\omega}} \ne \widetilde{\boldsymbol{\omega}}$ but $\pi_1\widehat{\boldsymbol{\omega}} = \pi_1\widetilde{\boldsymbol{\omega}}$. Therefore, there exists $\theta \in C([0,1],\mathbb{R}^{d_2})$ such that $\theta \not\equiv 0$ and $\widetilde{\boldsymbol{\omega}} = \widehat{\boldsymbol{\omega}}\boldsymbol{\theta}$ with $\boldsymbol{\theta} = (0,\theta)$. By Proposition 6.20, we find a sequence $\widehat{\boldsymbol{\omega}}^n$ such that $\widehat{\boldsymbol{\omega}}^n \to \widehat{\boldsymbol{\omega}}$ and

$$\limsup_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widehat{\boldsymbol{\omega}}^n) \le C_{\mathcal{H}}(\boldsymbol{\omega}, \widehat{\boldsymbol{\omega}}) = \|h\|_{\mathcal{H}} < \infty. \tag{6.16}$$

Now, by the triangle inequality and the estimate $(a+b)^2 \le 2a^2 + 2b^2$, for any $a, b \in \mathbb{R}$, we see that

$$C_n(\widehat{\boldsymbol{\omega}}^n, \widetilde{\boldsymbol{\omega}}^n)^2 \le 2C_n(\boldsymbol{\omega}^n, \widehat{\boldsymbol{\omega}}^n)^2 + 2C_n(\widetilde{\boldsymbol{\omega}}^n, \boldsymbol{\omega}^n)^2.$$

By (6.16), the first term on the right-hand side is bounded by $2\|h\|_{\mathcal{H}}^2$. By Lemma 6.19, we also have $\lim_{n\to\infty} C_n(\widehat{\boldsymbol{\omega}}^n, \widetilde{\boldsymbol{\omega}}^n)^2 = \infty$. Thus we conclude that $\lim_{n\to\infty} C_n(\widetilde{\boldsymbol{\omega}}^n, \boldsymbol{\omega}^n) = \infty$. ☐

Combining Propositions 6.20 and 6.21, we deduce the following Γ-convergence.

Corollary 6.22. *On $\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ equipped with the uniform topology, we have the Γ-convergence $C_n \xrightarrow{\Gamma} C_{\mathcal{H}}$.*

Proof. The liminf inequality follows directly from Proposition 6.21. The limsup inequality in this setting reads: For every pair $(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) \in \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$, we can find a sequence $(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n)$ converging to $(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$ such that

$$\limsup_{n\to\infty} C_n(\boldsymbol{\omega}^n, \widetilde{\boldsymbol{\omega}}^n) \le C_n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}).$$

Proposition 6.20 tells us that in fact we can take any sequence $\boldsymbol{\omega}^n$ converging to $\boldsymbol{\omega}$ and the sequence $\widetilde{\boldsymbol{\omega}}^n$ constructed via adding a suitable perturbation. ☐

Remark 6.23. Let us note that Proposition 6.20 is stronger than the standard lim sup condition in Γ-convergence. In particular, we can choose the constant sequence $\boldsymbol{\omega}^n = \boldsymbol{\omega}$ such that the recovery sequence is obtained via a map $\Phi^n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) = (\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}^n)$. We will use this map Φ^n to construct sequences of transport plans $\widetilde{\boldsymbol{\lambda}}^n$ that are recovery sequences for the family of optimal transport problems associated with C_n; see Proposition 6.26 below.

Having shown the Γ-convergence of the cost functions C_n, we can now deduce the Γ-convergence of the associated transport problems. For $n \in \mathbb{N}$, define the family of transport functionals $\mathrm{I}_n \colon \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}) \to [0,\infty]$ and $\mathrm{I}_\infty \colon \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}) \to [0,\infty]$ via

$$\mathrm{I}_n(\boldsymbol{\lambda}) = \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C_n^2 \,\mathrm{d}\boldsymbol{\lambda} \quad \text{and} \quad \mathrm{I}_\infty(\boldsymbol{\lambda}) = \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C_{\mathcal{H}}^2 \,\mathrm{d}\boldsymbol{\lambda}, \quad \boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}).$$

Proposition 6.24. *Let $(\boldsymbol{\lambda}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ be a sequence of probability measures such that $\boldsymbol{\lambda}^n \rightharpoonup \boldsymbol{\lambda}$ in $\mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$. Then*

$$\liminf_{n\to\infty} \mathrm{I}_n(\boldsymbol{\lambda}^n) \ge \mathrm{I}_\infty(\boldsymbol{\lambda}).$$

Proof. By Skorokhod's representation theorem, there exists a probability space $(\Xi, \mathfrak{A}, \mathbb{P})$ and random variables $\boldsymbol{Y}^n \colon \Xi \to \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ and $\boldsymbol{Y} \colon \Xi \to \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ such that $\boldsymbol{\lambda}^n = \boldsymbol{Y}^n_{\#}\mathbb{P}$, $\boldsymbol{\lambda} = \boldsymbol{Y}_{\#}\mathbb{P}$, and $\boldsymbol{Y}^n \to \boldsymbol{Y}$ $\mathbb{P}$-almost surely. We conclude that

$$\begin{aligned}\liminf_{n\to\infty} \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C_n(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}})^2 \,\mathrm{d}\boldsymbol{\lambda}^n &= \liminf_{n\to\infty} \int_{\Xi} C_n(\boldsymbol{Y}^n)^2 \,\mathrm{d}\mathbb{P} \\ &\ge \int_{\Xi} \liminf_{n\to\infty} C_n(\boldsymbol{Y}^n)^2 \,\mathrm{d}\mathbb{P} \ge \int_{\Xi} C_{\mathcal{H}}(\boldsymbol{Y})^2 \,\mathrm{d}\mathbb{P} = \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C_{\mathcal{H}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}})^2 \,\mathrm{d}\boldsymbol{\lambda},\end{aligned}$$

by Fatou's lemma and Proposition 6.21. ☐

Remark 6.25. Given Proposition 6.24, we find a much more direct proof of Theorem 6.6. Indeed, combining Lemma 6.3 and Proposition 6.24 yields the result immediately.

Proposition 6.26. *Let $\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$. Then there exists a sequence $(\widetilde{\boldsymbol{\lambda}}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ such that $\widetilde{\boldsymbol{\lambda}}^n \rightharpoonup \boldsymbol{\lambda}$ and*

$$\limsup_{n\to\infty} \mathrm{I}_n(\widetilde{\boldsymbol{\lambda}}^n) \le \mathrm{I}_\infty(\boldsymbol{\lambda}). \tag{6.17}$$

Proof. We may assume that the right-hand side in (6.17) is finite as the inequality is trivially true otherwise. In particular, we have $(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \mapsto C^2_{\mathcal{H}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \in L^1(\boldsymbol{\lambda})$ and, for $\boldsymbol{\lambda}$-almost every $(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \in \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$, we have that $\overline{\boldsymbol{\omega}} = T_h \boldsymbol{\omega}$ for $h = \pi_1(\overline{\boldsymbol{\omega}}^{-1}\boldsymbol{\omega}) \in \mathcal{H}$.

Define $\widetilde{\boldsymbol{\lambda}}^n = \Phi^n_{\#}\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$, where $\Phi^n \colon \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}} \to \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$ maps $(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$ to $(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}^n)$ as in Remark 6.23. Then, for any $(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) \in \boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}$, we have $\Phi^n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) \to (\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$ and $\widetilde{\boldsymbol{\lambda}}^n \rightharpoonup \boldsymbol{\lambda}$ as $n \to \infty$. By Propositions 6.20 and 6.21,

$$\lim_{n\to\infty} C_n(\Phi^n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})) = C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}). \tag{6.18}$$

Moreover, by Lemma 6.18, we have $C_n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}^n) \le \|h\|_{\mathcal{H}} = C_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$. Using Fatou's lemma with integrable upper bound $C^2_{\mathcal{H}}(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})$ gives

$$\limsup_{n\to\infty} \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C^2_n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}) \,\mathrm{d}\widetilde{\boldsymbol{\lambda}}^n = \limsup_{n\to\infty} \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C^2_n(\Phi^n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})) \,\mathrm{d}\boldsymbol{\lambda} \le \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} \limsup_{n\to\infty} C^2_n(\Phi^n(\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}})) \,\mathrm{d}\boldsymbol{\lambda}.$$

The assertion now follows from (6.18). □

Corollary 6.27. *On $\mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ equipped with the weak topology, we have the Γ-convergence $\mathrm{I}_n \xrightarrow{\Gamma} \mathrm{I}_\infty$.*

The following theorem is a version of the fundamental theorem of Γ-convergence in the present case.

Theorem 6.28. *Let $\boldsymbol{\eta} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$. Then $\mathrm{T}_{C_n,2}(\boldsymbol{\eta}, \cdot) \xrightarrow{\Gamma} \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta}, \cdot)$ with respect to the weak topology on $\mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$. That is*

- *(i) For any $\boldsymbol{\nu} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ and any $(\boldsymbol{\nu}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ such that $\boldsymbol{\nu}^n \rightharpoonup \boldsymbol{\nu}$, $\liminf_{n\to\infty} \mathrm{T}_{C_n,2}(\boldsymbol{\eta}, \boldsymbol{\nu}^n) \ge \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta}, \boldsymbol{\nu})$; and*
- *(ii) For any $\boldsymbol{\nu} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$, there exists a sequence $(\widetilde{\boldsymbol{\nu}}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ such that $\lim_{n\to\infty} \mathrm{T}_{C_n,2}(\boldsymbol{\eta}, \widetilde{\boldsymbol{\nu}}^n) = \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta}, \boldsymbol{\nu})$.*

Proof. (i) Let $\boldsymbol{\nu} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ and $(\boldsymbol{\nu}^n) \subset \mathcal{P}(\Omega_{\mathbb{G}})$ such that $\boldsymbol{\nu}^n \rightharpoonup \boldsymbol{\nu}$, and let $(\boldsymbol{\lambda}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ be a sequence of optimal transport plans for $\mathrm{T}_{C_n,2}(\boldsymbol{\eta}, \widetilde{\boldsymbol{\nu}}^n)$. We can assume that $\boldsymbol{\lambda}^n$ converges weakly to a limit $\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ since its marginals are tight by Prokhorov's theorem; see [AGS08, Lemma 5.2.2]. The limit $\boldsymbol{\lambda}$ has marginals $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ and is hence an admissible transport plan for $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta}, \boldsymbol{\nu})$. Using Proposition 6.24, we get the chain of inequalities

$$\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta}, \boldsymbol{\nu}) \le \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C^2_{\mathcal{H}}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \,\mathrm{d}\boldsymbol{\lambda} \le \liminf_{n\to\infty} \int_{\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}}} C^2_n(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}) \,\mathrm{d}\boldsymbol{\lambda}^n = \liminf_{n\to\infty} \mathrm{T}_{C_n,2}(\boldsymbol{\eta}, \widetilde{\boldsymbol{\nu}}^n).$$

(ii) Now let $\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ be an optimal transport plan for $\mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta}, \boldsymbol{\nu})$ (note that $C_{\mathcal{H}}$ is lower semi-continuous; see Lemma 2.20). Let the sequence of transport plans $(\widetilde{\boldsymbol{\lambda}}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}} \times \boldsymbol{\Omega}_{\mathbb{G}})$ be given as in Proposition 6.26, and define $\widetilde{\boldsymbol{\nu}}^n$ as the second marginal of $\widetilde{\boldsymbol{\lambda}}^n$, for $n \in \mathbb{N}$. The first marginal of $\widetilde{\boldsymbol{\lambda}}^n$ is fixed to $\boldsymbol{\eta}$ for all $n \in \mathbb{N}$. Thus $\widetilde{\boldsymbol{\nu}}^n \rightharpoonup \boldsymbol{\nu}$. Moreover, by Proposition 6.26 and the optimality of $\boldsymbol{\lambda}$,

$$\limsup_{n\to\infty} \int_{\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}} C_n^2(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\,\mathrm{d}\widetilde{\boldsymbol{\lambda}}^n \le \mathrm{T}^2_{C_{\mathcal{H}},2}(\boldsymbol{\eta},\boldsymbol{\nu}).$$

On the other hand, if $(\overline{\boldsymbol{\lambda}}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}})$ is a sequence of optimal transport plans for $\mathrm{T}_{C_n,2}(\boldsymbol{\eta},\widetilde{\boldsymbol{\nu}}^n)$, we can assume that $\overline{\boldsymbol{\lambda}}^n \rightharpoonup \overline{\boldsymbol{\lambda}}$, where the limit $\overline{\boldsymbol{\lambda}}$ has marginals $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$. By Proposition 6.24, we get

$$\mathrm{T}^2_{C_{\mathcal{H}},2}(\boldsymbol{\eta},\boldsymbol{\nu}) \le \int_{\boldsymbol{\Omega}_{\mathbb{G}}\times\boldsymbol{\Omega}_{\mathbb{G}}} C_{\mathcal{H}}^2(\boldsymbol{\omega},\overline{\boldsymbol{\omega}})\,\mathrm{d}\overline{\boldsymbol{\lambda}} \le \liminf_{n\to\infty} \mathrm{T}^2_{C_n,2}(\boldsymbol{\eta},\widetilde{\boldsymbol{\nu}}^n).$$

Combining both estimates proves the claim. □

Remark 6.29. In general, we cannot rule out that there exists a sequence $\boldsymbol{\nu}^n$ converging to some limit $\boldsymbol{\nu}$ such that $\lim_{n\to\infty} \mathrm{T}_{C_n,2}(\boldsymbol{\eta},\boldsymbol{\nu}^n) > \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta},\boldsymbol{\nu})$. The crucial point in Theorem 6.28 is that the sequence $\widetilde{\boldsymbol{\nu}}^n$ is a special sequence constructed via the push-forward of the recovery map Φ^n. It is an interesting question whether the following stronger result holds: Let $\delta_{\Pi(\boldsymbol{\eta},\boldsymbol{\nu}^n)}$ denote the convex indicator function (taking values in $\{0,\infty\}$) for the set of admissible plans, i.e. $\delta_{\Pi(\boldsymbol{\eta},\boldsymbol{\nu}^n)}(\boldsymbol{\lambda}) = 0$ if and only if $\boldsymbol{\lambda} \in \Pi(\boldsymbol{\eta},\boldsymbol{\nu}^n)$, and let $\boldsymbol{\nu} \in \mathcal{P}(\boldsymbol{\Omega})$, $(\boldsymbol{\nu}^n) \subset \mathcal{P}(\boldsymbol{\Omega}_{\mathbb{G}})$ such that $\boldsymbol{\nu}^n \rightharpoonup \boldsymbol{\nu}$ and $\sup_{n\in\mathbb{N}} H(\boldsymbol{\nu}^n\|\boldsymbol{\eta}) < \infty$. Do we have the Γ-convergence $\mathrm{I}_n + \delta_{\Pi(\boldsymbol{\eta},\boldsymbol{\nu}^n)} \xrightarrow{\Gamma} \mathrm{I}_\infty + \delta_{\Pi(\boldsymbol{\eta},\boldsymbol{\nu})}$? This property would imply that $\mathrm{T}_{C_n,2}(\boldsymbol{\eta},\boldsymbol{\nu}^n) \to \mathrm{T}_{C_{\mathcal{H}},2}(\boldsymbol{\eta},\boldsymbol{\nu})$ for *every* converging sequence $\boldsymbol{\nu}^n$ with finite relative entropy.

7. Beyond step-2 Carnot groups

Parts of this work are valid in the generality of general Carnot groups (see, e.g. [BLU07]). However, Carnot groups for which the log-Sobolev inequality is known are the Heisenberg group and more general H-type groups, which are examples of step-2 Carnot groups, as discussed in Section 2.2. This explains our focus on step-2. Nevertheless, Theorem 5.1 holds for general Carnot groups with no restriction on the step of the group, and the proof remains unchanged, given the appropriate definitions. Similarly, Lemma 6.1, Proposition 6.2, and Lemma 6.3 carry over without change to the general Carnot group setting. Finally, Theorem 6.6 also holds for $\mathbb{G} = \mathbb{F}^{d_1,N}$, i.e. for step-$N$ free Carnot groups, under additional regularity assumptions for $N > 2$. On the space of p-variation paths, for any p such that the shift by an absolutely continuous path is well defined, the proof of Theorem 6.6 remains valid; see also Remark 2.17.

Acknowledgement. PF, HK, and ML acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CRC/TRR 388 "Rough Analysis, Stochastic Dynamics and Related Fields" – Project ID 516748464 within sub-project A02. This project also funded research stays of BR, during which part of this work was completed. The research of VL has been partially funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB1294/1-318763901.

References

[ABC+00] Cécile Ané, Sébastien Blachère, Djalil Chafai, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les Inégalités de Sobolev Logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000.

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[Ald81] David J. Aldous. Weak convergence and general theory of processes. Unpublished monograph: Department of Statistics, University of California, Berkeley, 1981.

[AR04] Luigi Ambrosio and Séverine Rigot. Optimal mass transportation in the Heisenberg group. *J. Funct. Anal.*, 208(2):261–301, 2004.

[AS20] Luigi Ambrosio and Giorgio Stefani. Heat and entropy flows in Carnot groups. *Rev. Mat. Iberoam.*, 36(1):257–290, 2020.

[BB15] Marco Bramanti and Luca Brandolini. A proof of Hörmander's theorem for sublaplacians on Carnot groups. *Nonlinear Anal.*, 126:170–200, 2015.

[BBBC08] Dominique Bakry, Fabrice Baudoin, Michel Bonnefont, and Djalil Chafaï. On gradient bounds for the heat kernel on the Heisenberg group. *J. Funct. Anal.*, 255(8):1905–1938, 2008.

[BBBE20a] Julio Backhoff-Veraguas, Daniel Bartl, Mathias Beiglböck, and Manu Eder. Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.*, 24(3):601–632, 2020.

[BBBE20b] Julio Backhoff-Veraguas, Daniel Bartl, Mathias Beiglböck, and Manu Eder. All adapted topologies are equal. *Probab. Theory Relat. Fields*, 178(3-4):1125–1172, 2020.

[BBLZ17] Julio Backhoff-Veraguas, Mathias Beiglböck, Yiqing Lin, and Anastasiia Zalashko. Causal transport in discrete time and applications. *SIAM J. Optim.*, 27(4):2528–2562, 2017.

[BBP+25] Daniel Bartl, Mathias Beiglböck, Gudmund Pammer, Stefan Schrott, and Xin Zhang. The Wasserstein space of stochastic processes in continuous time. *arXiv:2501.14135*, 2025.

[Ber18] Robert J. Berman. On large deviations for Gibbs measures, mean energy and Gamma-convergence. *Constr. Approx.*, 48(1):3–30, 2018.

[BLU07] Andrea Bonfiglioli, E. Lanconelli, and F. Uguzzoni. *Stratified Lie Groups and Potential Theory for their Sub-Laplacians*. Springer Monographs in Mathematics. Springer, Berlin, Berlin, 2007.

[Bra02] Andrea Braides. *Γ-convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[BV05] François Bolley and Cédric Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math. (6)*, 14(3):331–352, 2005.

[CDPT07] Luca Capogna, Donatella Daniellie, Scott D. Pauls, and Jeremy T. Tyson, editors. *An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem*, volume 259 of *Progress in Mathematics*. Birkhäuser, Basel, 2007.

[CF26] Gideon Chiusole and Peter K. Friz. Towards abstract wiener model spaces. *Probab. Theory Related Fields*, 2026. To appear.

[DGW04] Hacène Djellout, Arnaud Guillin, and Liming Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B), 2004.

[DM93] Gianni Dal Maso. *An introduction to Γ-convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, Boston, MA, 1993.

[DPMM+25] Guido De Philippis, Andrea Marchese, Andrea Merlo, Andrea Pinamonti, and Filip Rindler. On the converse of Pansu's theorem. *Arch. Ration. Mech. Anal.*, 249(1):3, 2025.

[DZ10] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[Eld10] Nathaniel Eldredge. Gradient estimates for the subelliptic heat kernel on H-type groups. *J. Funct. Anal.*, 258(2):504–533, 2010.

[FH20] Peter K. Friz and Martin Hairer. *A Course on Rough Paths*. Universitext. Springer, Cham, second edition, 2020.

[Föl86] Hans Föllmer. Time reversal on Wiener space. In *Stochastic Processes – Mathematics and Physics (Bielefeld 1984)*, volume 1158 of *Lecture Notes in Mathematics*, pages 119–129. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.

[Föl88] Hans Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Mathematics*, pages 101–203. Springer Berlin Heidelberg, Berlin, Heidelberg, 1988.

[Föl22] Hans Föllmer. Optimal couplings on Wiener space and an extension of Talagrand's transport inequality. In George Yin and Thaleia Zariphopoulou, editors, *Stochastic Analysis, Filtering, and Stochastic Optimization*, page 30. Springer, 2022.

[FÜ02] Denis Feyel and Ali S. Üstünel. Measure transport on Wiener space and the Girsanov theorem. *C. R. Math. Acad. Sci. Paris*, 334(11):1025–1028, 2002.

[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.

[FV05] Peter K. Friz and Nicolas B. Victoir. Approximations of the Brownian rough path with applications to stochastic analysis. *Ann. Inst. Henri Poincaré Probab. Stat.*, 41(4):703–724, 2005.

[FV10] 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.

[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.

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

[Hai11] Martin Hairer. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.

[Hör67] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[HZ10] Waldemar Hebisch and Boguslaw Zegarliński. Coercive inequalities on metric measure spaces. *J. Funct. Anal.*, 258(3):814–851, 2010.

[Jan97] Svante Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[Jui14] Nicolas Juillet. Diffusion by optimal transport in Heisenberg groups. *Calc. Var. Partial Differential Equations*, 50(3-4):693–721, 2014.

[Koh73] Joseph J. Kohn. Pseudo-differential operators and hypoellipticity. In *Partial differential equations (Proc. Sympos. Pure Math., Vol. XXIII, Univ. California, Berkeley, Calif., 1971)*, volume XXIII of *Proc. Sympos. Pure Math.*, pages 61–69. Amer. Math. Soc., Providence, RI, 1973.

[Las18] Rémi Lassalle. Causal transference plans and their Monge–Kantorovich problems. *Stoch. Anal. Appl.*, 36(3):452–484, 2018.

[LD25] Enrico Le Donne. *Metric Lie Groups*, volume 306 of *Graduate Texts in Mathematics*. Springer, Cham, 2025. Carnot-Carathéodory Spaces from the Homogeneous Viewpoint.

[Leh13] Joseph Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(3):885–899, 2013.

[Li06] Hong-Quan Li. Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg. *J. Funct. Anal.*, 236(2):369–394, 2006.

[LQZ02] Michel Ledoux, Zhongmin Qian, and Tusheng Zhang. Large deviations and support theorem for diffusion processes via rough paths. *Stochastic Process. Appl.*, 102(2):265–283, 2002.

[Lyo98] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[Lyo14] Terry J. Lyons. Rough paths, signatures and the modelling of functions on streams. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*, pages 163–184. Kyung Moon Sa, Seoul, 2014.

[Mar18] Mauro Mariani. A Γ-convergence approach to large deviations. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 18(3):951–976, 2018.

[Mon01] Roberto Monti. *Distances, Boundaries and Surface Measures in Carnot-Carathéodory Spaces*. PhD thesis, Università di Trento, 2001. Available at `https://cvgmt.sns.it/paper/3706/`.

[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.

[Pan89] Pierre Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math.*, 129(1):1–60, 1989.

[Par26] Jonghwa Park. On a T_1 transport inequality for the adapted Wasserstein distance. *Electron. Commun. Probab.*, 31:1–12, 2026.

[PS17] Andrea Pinamonti and Gareth Speight. Porosity, differentiability and Pansu's theorem. *J. Geom. Anal.*, 27(3):2055–2080, 2017.

[Rie17] Sebastian Riedel. Transportation-cost inequalities for diffusions driven by Gaussian processes. *Electron. J. Probab.*, 22(24), 2017.

[Rin18] Filip Rindler. *Calculus of Variations.* Universitext. Springer, Cham, 2018.

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

[Vil09] Cédric Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer, 2009.