

NONEXPONENTIAL SANOV AND SCHILDER THEOREMS ON WIENER SPACE: BSDES, SCHRÖDINGER PROBLEMS AND CONTROL

BY JULIO BACKHOFF-VERAGUAS¹, DANIEL LACKER² AND LUDOVIC TANGPI³

¹*Stochastics and Financial Mathematics Group, University of Vienna, julio.backhoff@univie.ac.at*

²*Industrial Engineering and Operations Research, Columbia University, daniel.lacker@columbia.edu*

³*Operations Research and Financial Engineering, Princeton University, ludovic.tangpi@princeton.edu*

We derive new limit theorems for Brownian motion, which can be seen as nonexponential analogues of the large deviation theorems of Sanov and Schilder in their Laplace principle forms. As a first application, we obtain novel scaling limits of backward stochastic differential equations and their related partial differential equations. As a second application, we extend prior results on the small-noise limit of the Schrödinger problem as an optimal transport cost, unifying the control-theoretic and probabilistic approaches initiated respectively by T. Mikami and C. Léonard. Lastly, our results suggest a new scheme for the computation of mean field optimal control problems, distinct from the conventional particle approximation. A key ingredient in our analysis is an extension of the classical variational formula (often attributed to Borell or Boué–Dupuis) for the Laplace transform of Wiener measure.

1. Introduction. In this work we develop two new limit theorems for the Wiener process along with several applications. These can be seen as nonexponential extensions of the classical large deviation principles of Schilder and Sanov in their Laplace principle forms, in the spirit of recent limit theorems obtained in [23] by the second named author in an abstract setting. Along the way, we derive a variational principle for the Wiener process which can be seen as a reformulation of Gibbs variational principle as initiated by [5, 17]; see also [4, 26] for further developments. Our two limit theorems turn out to be a common ground for three domains of application, as we now describe.

Our first application concerns the theory of backward stochastic differential equations (BSDE), and their related convex dual and PDE representations. Our two main limit theorems lead to two new kinds of scaling limits for BSDEs. The first of these scaling limits, coming from our Schilder-type theorem, can be seen as a non-Markovian vanishing-viscosity limit. Indeed, by exploiting the well-known link between BSDEs and semilinear PDEs (see, e.g., [35, 36]), our result recovers as a special case the well-known convergence of a viscous Hamilton–Jacobi equation to its inviscid counterpart as the viscosity coefficient vanishes. Our Sanov-type theorem leads to a second and more unusual BSDE scaling limit, in which the terminal condition depends on the empirical distribution of n rescaled sub-paths of the Brownian motion. Although non-Markovian in nature, in special cases this translates to a limit theorem for “concatenated” semilinear PDEs.

Our second application concerns the convergence of Schrödinger-type problems (also called stochastic optimal transport) to classical optimal transport in the small noise limit. The Schrödinger problem is a well documented topic in probability theory and mechanics (see, e.g., [28] and the references therein), and its link to optimal transportation was developed by Föllmer [18] in his Saint Flour lecture notes. The study of small-noise limits

Received October 2018; revised June 2019.

MSC2010 subject classifications. 60F10, 60H30, 93E03.

Key words and phrases. Large deviations, Wiener space, Sanov theorem, Schilder theorem, nonexponential, BSDE, Schrödinger problem, stochastic control.

of Schrödinger problems was pioneered by Mikami in the works [32, 33], the second joint with Thieullen. The main tools in these articles were stochastic control and partial differential equations (PDEs). Subsequently, an elegant large deviations viewpoint was developed by Léonard in [27, 29]. We draw inspiration from both approaches, to a certain extent unifying them, as we exploit our Schilder-type theorem in order to obtain new small-noise results for Schrödinger-type problems.

The third application is a surprising connection with a particular type of optimal control problem, known as *mean field* or *McKean–Vlasov optimal control*, which have seen a surge of interest in recent years; see [6, 24, 37] and references therein. The limiting quantity in our Sanov-type theorems can be seen as the value of an optimal control problem in which the dependence of the optimization criterion on the law of the state process is nonlinear. Our limit theorem provides a new approximation scheme for such problems, markedly different from the natural particle approximation worked out in [24].

We now proceed to present the setting and main results in detail.

2. Setting and main results. Let $\mathcal{C} = C([0, 1]; \mathbb{R}^d)$ denote the continuous path space, equipped with the supremum norm $\|\cdot\|_\infty$ and its Borel σ -field. Let P denote the standard Wiener measure on \mathcal{C} . With $W = (W(t))_{t \in [0,1]}$ we denote the canonical (coordinate) process on \mathcal{C} , defined by setting $W(t)(\omega) = \omega(t)$, so that W is a standard d -dimensional Brownian motion under P . Let $\mathbb{F} = (\mathcal{F}_t)_{t \in [0,1]}$ denote the P -complete filtration generated by W . As usual, we denote by $L^0(P)$ the space of (real-valued) random variables quotiented with the P -a.s. identification, and by $L^\infty(P)$ the essentially bounded elements of $L^0(P)$. We will likewise identify processes that are $dt \otimes dP$ -almost surely equal.

We are given a function $g : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ and write $\text{dom}(g(t, \cdot)) := \{q \in \mathbb{R}^d : g(t, q) < \infty\}$ for its effective domain. We impose the following standing assumption:

(TI) The function g is measurable and bounded from below, and it is coercive in the sense that $\lim_{|q| \rightarrow \infty} \inf_{t \in [0,1]} \frac{g(t,q)}{|q|} = \infty$. For each $t \in [0, 1]$ the function $g(t, \cdot)$ is convex, proper, and lower semicontinuous. Finally, the following technical conditions hold:

$$(1) \quad 0 \in \text{ri}(\text{dom}(g(t, \cdot))) = \text{ri}(\text{dom}(g(s, \cdot))) =: \mathcal{R} \quad \text{for all } s, t \in [0, 1],$$

where $\text{ri}(\text{dom}(g(t, \cdot)))$ denotes the relative interior of $\text{dom}(g(t, \cdot))$, and

$$(2) \quad \int_0^1 |\sup\{g(t, q) : q \in \mathcal{R}, |q| \leq r\}| dt < \infty \quad \text{for all } r \geq 0.$$

The final technical conditions (1) and (2) always holds if g is finite-valued and jointly continuous. A typical example which takes the value $+\infty$ and which satisfies (TI) is $g(t, q) = +\infty 1_K(q)$, the convex indicator of a convex compact set $K \subset \mathbb{R}^d$. The assumption that $0 \in \mathcal{R}$ is unnecessary, but it is convenient and not terribly restrictive.

Define \mathcal{L} to be the set of progressively measurable \mathbb{R}^d -valued processes $q : [0, 1] \times \mathcal{C} \rightarrow \mathbb{R}^d$ satisfying $P(\int_0^1 |q(t)|^2 dt < \infty) = 1$. We often write $q(t) = q(t, \cdot)$, suppressing the dependence on $\omega \in \mathcal{C}$. We denote by $\int_0^1 q^Q(t) dW(t)$ the stochastic integral $\int_0^1 q^Q(t) \cdot dW(t)$. Let \mathcal{Q} be the set of probability measures absolutely continuous with respect to P . It is well known that for every $Q \in \mathcal{Q}$, there is a unique process $q^Q \in \mathcal{L}$ such that Q -a.s.

$$\frac{dQ}{dP} = \exp\left(\int_0^1 q^Q(t) dW(t) - \int_0^1 \frac{1}{2} |q^Q(t)|^2 dt\right).$$

A partial converse which we will often use is as follows: Letting $\mathcal{L}_b \subset \mathcal{L}$ denote the subset of *bounded* processes, for each $q \in \mathcal{L}_b$ there is a unique $Q^q \in \mathcal{Q}$ such that $q^{Q^q} = q$. By Girsanov’s theorem, we may express this measure as

$$(3) \quad Q^q = P \circ \left(W + \int_0^\cdot q(t) dt\right)^{-1}.$$

In the following we write \mathbb{E} for expectation under P and \mathbb{E}^Q for expectation under any other measure Q .

The main objects we study are the conjugate functionals

$$\alpha^g : \mathcal{Q} \rightarrow \mathbb{R} \cup \{+\infty\}, \quad \rho^g : L^\infty(P) \rightarrow \mathbb{R},$$

respectively given by

$$(4) \quad \alpha^g(Q) := \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] \quad \text{and} \quad \rho^g(X) := \sup_{Q \in \mathcal{Q}} (\mathbb{E}^Q[X] - \alpha^g(Q)).$$

Because g is bounded from below, note that $\alpha^g(Q)$ is well defined and bounded from below, taking values in $\mathbb{R} \cup \{+\infty\}$. Note also that α^g is not identically $+\infty$ since $\int_0^1 |g(t, 0)| dt < \infty$, and in particular $-\infty < \rho^g(X) < \infty$ for all $X \in L^\infty(P)$.

The classical example to keep in mind is the quadratic case, $g(t, q) = \frac{1}{2}|q|^2$. In this case, α^g is nothing but the relative entropy,

$$(5) \quad \alpha^g(Q) = H(Q|P) := \mathbb{E} \left[\frac{dQ}{dP} \log \frac{dQ}{dP} \right], \quad Q \in \mathcal{Q}$$

and, by the Gibbs variational principle, ρ^g is the cumulant generating functional, $\rho^g(X) = \log \mathbb{E}[e^X]$. A more trivial example is given by $g(t, q) = +\infty 1_{\{0\}^c}(q)$, the convex indicator of 0, in which case $\rho^g(X) = \mathbb{E}[X]$. More generally, if $g(t, q) = +\infty 1_{K^c}(q)$ for some compact convex set $K \subset \mathbb{R}^d$, then $\rho^g(X) = \sup\{\mathbb{E}^Q[X] : Q \in \mathcal{Q}, q_t^Q \in K \text{ dt } \otimes dP\text{-a.e.}\}$. When we turn to Schrödinger problems in Section 2.3, a general time-independent function $g(q)$ will lead us to the optimal transport problem with cost function $(x, y) \mapsto g(x - y)$.

We will derive in Theorem 3.1 yet another representation of ρ^g , in the spirit of stochastic optimal control. For $F \in L^\infty(P)$ we show that

$$(BBD) \quad \rho^g(F) = \sup_{q \in \mathcal{L}_b} \mathbb{E} \left[F \left(W + \int_0^1 q(t) dt \right) - \int_0^1 g(t, q(t)) dt \right].$$

The fact that F is path-dependent here means that the representation (BBD) does not follow as quickly from the definition of $\rho^g(F)$ as it may seem at first sight. In the case $g(t, q) = \frac{1}{2}|q|^2$, the representation (BBD) was a key technical result of Boué and Dupuis [5] and Lehec [26].

We first summarize in Section 2.1 our main limit theorems for the functionals ρ^g , and the remaining parts of this section explain the various applications: Section 2.2 explains the implications for BSDEs. This is followed by Section 2.3 where we present some new insights into the study of convergence of stochastic transport problems (i.e., Schrödinger-type problems) to optimal transport problems. We close this overview section with discussions of connections with PDEs in Section 2.4 and (mean field) optimal control in Section 2.5.

2.1. *Limit theorems.* To state our first main limit theorem, a nonexponential version of Sanov’s theorem in its Laplace principle form, we introduce the following notation: For a Polish space E , we denote by $\mathcal{P}(E)$ the set of Borel probability measures on E equipped with the topology of weak convergence and by $C_b(E)$ the space of bounded continuous functions on E . For $n \in \mathbb{N}$, $k = 1, \dots, n$ and a path $\omega \in \mathcal{C}$, we define the *chopped and rescaled path* $\omega_{(n,k)} \in \mathcal{C}$ by

$$(6) \quad \omega_{(n,k)}(t) := \sqrt{n} \left(\omega \left(\frac{k-1+t}{n} \right) - \omega \left(\frac{k-1}{n} \right) \right), \quad t \in [0, 1].$$

Note that $(W_{(n,k)})_{k=1}^n$ are n independent Brownian motions (under P). In the following, recall that we always work with a given function g satisfying the standing assumption (TI).

THEOREM 2.1. Define $G_n : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ by

$$G_n(t, q) := g\left(nt - \lfloor nt \rfloor, \frac{q}{\sqrt{n}}\right).$$

Then G_n satisfies (TI) for each n , and for every $F \in C_b(\mathcal{P}(\mathcal{C}))$ we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \rho^{G_n} \left(F \left(\frac{1}{n} \sum_{k=1}^n \delta_{W_{(n,k)}} \right) \right) &= \sup_{Q \in \mathcal{Q}} (F(Q) - \alpha^g(Q)) \\ &= \sup_{q \in \mathcal{L}_b} \left(F(Q^q) - \mathbb{E} \left[\int_0^1 g(t, q(t)) dt \right] \right), \end{aligned}$$

where Q^q was defined in (3) for $q \in \mathcal{L}_b$.

The proof is given at the end of Section 4. For the second main result, we adopt the convention that

$$(7) \quad \int_0^1 g(t, \dot{\omega}(t)) dt = +\infty$$

whenever $\omega \in \mathcal{C}$ is not absolutely continuous. Define $\mathcal{C}_0 := \{\omega \in \mathcal{C} : \omega(0) = 0\}$. Our second main limit theorem is a nonexponential version of Schilder’s theorem in Laplace principle form:

THEOREM 2.2. Denote $g_n(t, q) := g(t, q/\sqrt{n})$. Then g_n satisfies (TI) for each n , and for every $F \in C_b(\mathcal{C})$, we have

$$(8) \quad \lim_{n \rightarrow \infty} \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) = \sup_{\omega \in \mathcal{C}_0} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right).$$

Moreover, if $g(t, q) = g(q)$ does not depend on t , and if $h \in C_b(\mathbb{R}^d)$, we have

$$\lim_{n \rightarrow \infty} \rho^{g_n} \left(h \left(\frac{W(1)}{\sqrt{n}} \right) \right) = \sup_{x \in \mathbb{R}^d} (h(x) - g(x)).$$

The proof is given at the end of Section 3. Returning to the quadratic case $g(t, q) := \frac{1}{2}|q|^2$ reveals how Theorems 2.1 and 2.2 relate to the classical theorems of Sanov and Schilder. In this case, $G_n(t, q) = g_n(t, q) = \frac{1}{2n}|q|^2$ for every n , and as mentioned above we get

$$\rho^{G_n}(X) = \rho^{g_n}(X) = \frac{1}{n} \log \mathbb{E}[e^{nX}],$$

and $\alpha^g(Q) = H(Q|P)$ is the relative entropy, as defined in (5). Thus in the quadratic case Theorems 2.1 and 2.2 respectively reduce to Sanov and Schilder theorem for Brownian motion in their Laplace principle forms; see [9], Theorems 6.2.10 and 5.2.3, respectively for classical statements of Sanov and Schilder’s theorems and [14], Theorems 1.2.1 and 1.2.3, for the equivalence with Laplace principles. For another explicit but trivial example, if $g(t, q) = +\infty 1_{\{0\}}(q)$, then $\rho^g(X) = \mathbb{E}[X]$ as above, and Theorem 2.1 reduces to the law of large numbers, stating that the random measures $\frac{1}{n} \sum_{k=1}^n \delta_{W_{(n,k)}}$ converge weakly to P . Similarly, if $g(t, q) = +\infty 1_{\{z\}}(q)$ for some $z \in \mathbb{R}^d$ then $\rho^g(X) = \mathbb{E}^{Q_z}[X]$ with Q_z denoting the law of the Brownian motion with drift z , although strictly speaking this example does not fit assumption (TI) because $g(t, 0) = \infty$. Explicit formulas for $\rho^g(X)$ are unfortunately few and far between, and we do not know of any outside of the classical case and the trivial cases just discussed; see Remark 2.5 for more on this.

It is important to note that the chopped paths $(W_{(n,k)})_{k=1}^n$ appearing in Theorem 2.1 cannot be replaced with an arbitrary sequence of n independent Brownian motions, because the functional ρ^{G_n} is not necessarily law-invariant!¹ For this reason, Theorem 2.2 cannot be deduced from Theorem 2.1, contrary to the classical case in which Schilder’s theorem can be deduced from Sanov’s theorem and continuous mapping. Nevertheless, in Corollary 7.3 we derive from Theorem 2.1 a result more in the spirit of Cramér’s theorem, which notably shares the same limiting expression as Theorem 2.2 despite involving a quite distinct pre-limit quantity.

The key to proving these limit theorems is the stochastic control representation (BBD), which we establish in Theorem 3.1. Theorem 2.1 is ultimately a specialization of the abstract Sanov-type theorem of the second named author in [23], though computing the iterates denoted ρ_n therein is a highly nontrivial step here. The proof of Theorem 2.2 is direct and does not use the results of [23]. A major difficulty in our analysis is the lack of lower-semicontinuity of ρ^g and weak compactness of the sublevel sets of α^g when g is subquadratic, which necessitates the study of a better-behaved functional (see $\tilde{\alpha}^g$ in Section 4.2). In Section 7, we extend Theorem 2.2 to random initial conditions and Theorem 2.1 to stronger topologies.

REMARK 2.3. It is possible to leverage our results, which concern Brownian motion solely, in order to obtain analogous limit results for those stochastic differential equations which can be solved path-by-path in a continuous fashion. To exemplify, let $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be bounded, jointly continuous and Lipschitz in the second argument (uniformly in the first one). Denote by X^q the weakly-unique solution of the controlled SDE $dX(t) = [b(t, X(t)) + q(t)] dt + dW(t)$ with $X^q(0) = a$. Recalling $W_{(n,k)}$ as given in (6), we further denote by $X_{n,k}$ the unique strong solution of the SDE $dX_{n,k}(t) = b(t, X_{n,k}(t)) dt + dW_{(n,k)}(t)$ with $X_{n,k}(0) = a$. Then the analogue of Theorem 2.1 is

$$\lim_{n \rightarrow \infty} \rho^{G_n} \left(F \left(\frac{1}{n} \sum_{k=1}^n \delta_{X_{n,k}} \right) \right) = \sup_{q \in \mathcal{L}_b} \left(F(\text{Law}(X^q)) - E \left[\int_0^T g(t, q(t)) dt \right] \right).$$

A similar analogue of Theorem 2.2 is possible. The key point is that the assumptions on b ensure that the integral equation $x(t) = \int_0^t b(s, x(s)) ds + \omega(t)$ for each $t \in [0, 1]$ has a unique solution $x = S(\omega)$, for each $\omega \in \mathcal{C}$, where the map $S : \mathcal{C} \rightarrow \mathcal{C}$ is continuous. We leave it to the interested reader to complete the remaining straightforward arguments. We do not elaborate further in this direction, since the scope of this line of arguments is rather limited. To wit, if we wanted to cover the case of SDEs with a variable diffusion coefficient, then new techniques and arguments seem to be indispensable.

2.2. *Scaling limits of BSDE.* In this section, we develop the first application, stating two new results on scaling limits for BSDEs. One consequence of the assumption (TI) is a stochastic representation of ρ^g in terms of BSDEs: Let g^* stand for the convex conjugate of g in the spatial variable, namely

$$(9) \quad g^*(t, z) := \sup_{q \in \mathbb{R}^d} (q \cdot z - g(t, q)).$$

Following [10], we say that a pair (Y, Z) , where Y is a càdlàg and adapted process and with $Z \in \mathcal{L}$, is a supersolution to the BSDE (driven by W , with terminal condition $X \in L^0(P)$, and generator g^*)

$$(10) \quad dY(t) = -g^*(t, Z(t)) dt + Z(t) dW(t), \quad Y(1) = X,$$

¹A functional $\rho : L^0 \rightarrow \mathbb{R} \cup \{+\infty\}$ is law-invariant if $\rho(X) = \rho(X')$ whenever X and X' have the same law.

if it satisfies

$$(11) \quad \begin{cases} Y(s) - \int_s^t g^*(u, Z(u)) du + \int_s^t Z(u) dW(u) \geq Y(t) & \text{for every } 0 \leq s \leq t \leq 1, \\ Y(1) \geq X, \end{cases}$$

and $\int Z dW$ is a supermartingale. A supersolution (\bar{Y}, \bar{Z}) of (10) is said to be minimal if $\bar{Y}(t) \leq Y(t)$ a.s. for each $t \in [0, 1]$ and for every supersolution (Y, Z) . By [10], Theorem 4.17, under the condition (TI), the BSDE (10) admits a unique minimal supersolution for every terminal condition X bounded from below.

The crucial link is given in [11], Theorems 3.4/3.10, where it was shown that

$$(12) \quad \rho^g(X) = \bar{Y}(0),$$

where (\bar{Y}, \bar{Z}) is the minimal supersolution of (11), provided that X is for example, bounded. This is the aforementioned representation of ρ^g in terms of a BSDE. Additionally, it is well known that a nonlinear Feynman–Kac formula connects BSDEs with semilinear parabolic PDEs, and we will briefly elaborate on this perspective in Section 2.4 below.

REMARK 2.4. If $X \in L^\infty(P)$ and g has at least quadratic growth, then g^* has subquadratic growth and the BSDE (10) admits a unique solution (Y, Z) such that Y is bounded (see, e.g., [8, 21]). Thus, it follows by [11], Theorem 4.6, that $\rho^g(X) = Y_0$. That is, the minimal supersolution and the unique (true) solution coincide. Consequently, all results stated in this paper for minimal supersolutions transfer to true solutions when g is of superquadratic growth. When this is not the case, a solution to a BSDE need not exist or be unique (see, e.g., Delbaen et al. [8]), and the weaker concept of minimal supersolution becomes essential.

REMARK 2.5. The BSDE formulation sheds light on why we do not expect to find explicit formulas for $\rho^g(X)$ outside of the classical case $g(t, q) = c|q|^2$ for $c > 0$ and the trivial cases $g(t, q) = +\infty 1_{\{z\}}(q)$ for $z \in \mathbb{R}^d$. Indeed, nonlinear BSDEs (and their semilinear PDE counterparts discussed in Section 2.4) are rarely explicitly solvable.

Using the representation (12), Theorems 2.1 and 2.2 immediately translate into limit theorems for the time-zero values of suitable BSDEs. With some additional effort, we are able to bootstrap Theorems 2.1 and 2.2 in order to obtain limits at every time, and not just at time zero. We begin with the BSDE analogue of Theorem 2.1, with proofs deferred to Section 5.

THEOREM 2.6. *Let $F \in C_b(\mathcal{P}(C))$, and let (Y_n, Z_n) be the minimal supersolution of the BSDE*

$$(13) \quad dY(t) = -g^*(nt - \lfloor nt \rfloor, \sqrt{n}Z(t)) dt + Z(t) dW(t), \quad Y(1) = F\left(\frac{1}{n} \sum_{k=1}^n \delta_{W(n,k)}\right).$$

Then, for each $t \in [0, 1]$, we have the a.s. limit

$$\begin{aligned} \lim_{n \rightarrow \infty} Y_n\left(\frac{\lfloor nt \rfloor}{n}\right) &= \sup_{Q \in \mathcal{Q}} (F(tP + (1-t)Q) - (1-t)\alpha^g(Q)) \\ &= \sup_{q \in \mathcal{L}_b} \left(F(tP + (1-t)Q^q) - (1-t)\mathbb{E}\left[\int_0^1 g(s, q(s)) ds\right] \right). \end{aligned}$$

We do not know of any results similar to Theorem 2.6 in the literature. The limiting expression exhibits a structure remarkably parallel to the Hopf–Lax–Oleinik solution of a Hamilton–Jacobi equation, reviewed in Section 2.4.1 below. Another interesting feature is

that it is a decidedly non-Markovian result; even if F depends only on the time-1 marginal of the measure, the terminal conditional in (13) still depends on the value of W at n different times.

Next, in order to state a BSDE analogue of Theorem 2.2, we make use of the following notation: Define $\mathcal{C}_0[t, 1]$ to be the set of continuous paths $\omega : [t, 1] \rightarrow \mathbb{R}$ with $\omega(t) = 0$. Also, for $\omega \in \mathcal{C}$ and $\bar{\omega} \in \mathcal{C}_0[t, 1]$, define $\omega \oplus_t \bar{\omega} \in \mathcal{C}$ by

$$\omega \oplus_t \bar{\omega}(s) := \omega(s \wedge t) + \bar{\omega}(s)1_{[t,1]}(s).$$

Recall as in (7) that we set $\int_t^1 g(s, \dot{\omega}(s)) ds := \infty$ when ω is not absolutely continuous.

THEOREM 2.7. *Let $F \in C_b(\mathcal{C})$ and let (Y_n, Z_n) be the minimal supersolution of the BSDE*

$$(14) \quad dY(t) = -g^*(t, \sqrt{n}Z(t)) dt + Z(t) dW(t), \quad Y(1) = F\left(\frac{W}{\sqrt{n}}\right).$$

Then there exist progressively measurable functions $u_n : [0, 1] \times \mathcal{C} \rightarrow \mathbb{R}$ such that $Y_n(t) = u_n(t, W/\sqrt{n})$ a.s. for each n and $u_n \rightarrow u$ pointwise, where

$$u(t, \omega) := \sup_{\bar{\omega} \in \mathcal{C}_0[t,1]} \left(F(\omega \oplus_t \bar{\omega}) - \int_t^1 g(s, \dot{\bar{\omega}}(s)) ds \right).$$

Moreover, for each $t \in [0, 1]$, we have the a.s. limit

$$\lim_{n \rightarrow \infty} Y_n(t) = u(t, 0).$$

The previous theorem is noteworthy, as it shows that making the generator of the BSDE explode and its terminal condition trivialize at the same rate gives a nontrivial deterministic limit. Alternatively, we may move the rescaling to the Brownian motion itself. Letting $W^\epsilon = \sqrt{\epsilon}W$ denote Brownian motion with volatility $\epsilon = 1/n$, we can rewrite (14) as

$$(15) \quad dY(t) = -g^*(t, Z(t)) dt + Z(t) dW^\epsilon(t), \quad Y(1) = F(W^\epsilon),$$

and so Theorem 2.7 also shows a nontrivial effect of “cooling-down” the driving Brownian motion in such a BSDE. The closest related results seem to be those of the form of [38], Theorem 2.1, on (F)BSDEs with vanishing noise, though the factor \sqrt{n} in g^* in (14) is absent in [38].

The $\epsilon \downarrow 0$ limit of the BSDE (15) is intriguing from the perspective of BSDE stability theory. It has been known for some time that if the generator and terminal condition of a BSDE converge in a suitable sense, then so does the solution (Y, Z) . Modern BSDE theory has explored similar stability theorems in much more generality, when the driving martingale (in our case, W^ϵ) itself can vary (see [34] and the thesis [41] for thorough discussions and references). However, existing results in this direction require that the limiting BSDE admit a *unique* solution, and it is far from clear how to properly formulate a *uniquely* solvable BSDE driven by the zero martingale. Similarly, one may interpret Theorem 2.7 as a BSDE form of the regularization-by-noise phenomenon: The $\epsilon = 0$ equation is ill-posed, but for each $\epsilon > 0$ the equation is well-posed, and the $\epsilon \downarrow 0$ limit “selects” a particular solution. In Section 2.4.1, in the Markovian case, we will re-interpret this as the vanishing viscosity limit of Hamilton–Jacobi–Bellman equations.

The factor \sqrt{n} appears in the identity $Y_n(t) = u_n(t, W/\sqrt{n})$ in Theorem 2.7 for two reasons. On a purely mathematical level, this provides the scaling that results in a random (ω -dependent) limit for u_n . The second and more practical reason is that one can interpret $u(t, \omega)$ as the value function of a stochastic control problem in which the state process, W/\sqrt{n} , is

observed up to time t to agree with the path $(\omega(s))_{s \leq t}$. This will be perhaps more clear when we reinterpret Theorem 2.7 in terms of PDEs in Section 2.4.1 below.

In the quadratic case $g(t, q) = \frac{1}{2}|q|^2$, Theorem 2.7 reads as a “conditional” version of Schilder’s theorem for Brownian motion. Indeed, the solution of BSDE (14) and its a.s. limit in Theorem 2.7 are given by

$$\begin{aligned}
 Y_n(t) &= \frac{1}{n} \log \mathbb{E}[\exp(nF(W/\sqrt{n})) \mid \mathcal{F}_t] \\
 &\rightarrow \sup_{\omega \in \mathcal{C}_0[t,1]} \left(F(0 \oplus_t \omega) - \frac{1}{2} \int_t^1 |\dot{\omega}(s)|^2 ds \right).
 \end{aligned}$$

Of course, it is straightforward to derive this directly from the usual form of Schilder’s theorem. Similarly, in the quadratic case, Theorem 2.6 can be rewritten as the a.s. limit

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[\exp \left(nF \left(\frac{1}{n} \sum_{k=1}^n \delta_{W(n,k)} \right) \right) \mid \mathcal{F}_{\frac{[nt]}{n}} \right] \\
 = \sup_{Q \in \mathcal{Q}} (F(tP + (1-t)Q) + (1-t)H(Q|P)),
 \end{aligned}$$

for each $t \in [0, 1]$. It is likely that a direct argument would yield in this case that the same holds even if we replace $\frac{[nt]}{n}$ with t on the left-hand side. More generally, we conjecture that $Y_n(t)$ converges to the same limit as $Y_n(\frac{[nt]}{n})$ in the setting of Theorem 2.6.

2.3. *Small noise limit of Schrödinger-type problems.* In this part we aim to deepen the study of optimal transport as a small-noise limit of stochastic optimal transport. We first present the setting and main result, before discussing the connection with prior literature.

For $\epsilon > 0$ we introduce the set $\mathcal{P}_\epsilon^*(\mathcal{C})$ of $Q \in \mathcal{P}(\mathcal{C})$ for which there exists a progressively measurable \mathbb{R}^d -valued processes q^Q such that the process

$$\frac{1}{\sqrt{\epsilon}} \left(W(t) - W(0) - \int_0^t q^Q(s) ds \right)$$

is a standard d -dimensional Brownian motion under Q . We stress that for $Q \in \mathcal{P}_\epsilon^*(\mathcal{C})$ the process q^Q is uniquely determined (in the $dt \otimes dQ$ -a.s. sense), and that it is understood in the above definition that q^Q is dt -integrable Q -a.s. Note that \mathcal{Q} is a proper subset of $\mathcal{P}_\epsilon^*(\mathcal{C})$, and membership in \mathcal{Q} requires some integrability of q^Q which is not required in $\mathcal{P}_\epsilon^*(\mathcal{C})$.

We now introduce the problems of interest in this part of the work. Let Z be a separable Banach space, which we endow with its Borel sigma-algebra. We are given an *observable* H , which is nothing more than a continuous linear operator

$$H : \mathcal{C} \rightarrow Z.$$

We think of H as an *observable random quantity* whose distribution ν we know, and impose in advance into the problem. For instance, H could give the value of a path at different time points, as well as the value of successive integrals of the path. We denote by Q^t the marginal at time t of a path measure $Q \in \mathcal{P}(\mathcal{C})$. For μ and ν Borel probability measures respectively on \mathbb{R}^d and Z , we examine here the problems

$$(16) \quad \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \\ Q^0 = \mu, Q \circ H^{-1} = \nu}} \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right],$$

and their limits when $\epsilon \downarrow 0$.

The *classical Schrödinger problem* (see, e.g., the survey of Léonard [28]) arises from the specification

$$(17) \quad Z = \mathbb{R}^d, \quad H(\omega) = \omega(1), \quad g(t, q) = \frac{1}{2}|q|^2, \quad \nu \in \mathcal{P}(\mathbb{R}^d).$$

In this setting, the quantity (16) can be written as the problem of minimizing relative entropy with respect to P_ϵ subject to marginal constraints,

$$(18) \quad \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \\ Q^0 = \mu, Q^1 = \nu}} \epsilon H(Q|P_\epsilon),$$

where P_ϵ denotes Wiener measure with volatility ϵ , that is, $P_\epsilon = P \circ (\sqrt{\epsilon}W)^{-1}$. Moreover, if one replaces the target measure ν by its convolution ν_ϵ with a centered Gaussian of variance ϵ (see the discussion following Theorem 2.9 for why this is necessary), then the $\epsilon \rightarrow 0$ limit of (18) is precisely the quadratic Wasserstein distance:

$$\lim_{\epsilon \rightarrow 0} \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \\ Q^0 = \mu, Q^1 = \nu_\epsilon}} \epsilon H(Q|P_\epsilon) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}|x - y|^2 \pi(dx, dy),$$

where $\Pi(\mu, \nu)$ is the set of couplings of (μ, ν) , that is, the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν .

We now state our first main result, which relies fundamentally on our Schilder-type result Theorem 2.2, or rather the extension given in Section 7.2 to the case of random initial positions. Proofs of the results announced in this section are given in Section 8.

THEOREM 2.8. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(Z)$, and let $H : \mathcal{C} \rightarrow Z$ be linear and continuous. Let $\nu_\epsilon := \nu * (P_\epsilon \circ H^{-1})$ denote the convolution of ν with the push-forward by H of P_ϵ . Then²*

$$(19) \quad \lim_{\epsilon \downarrow 0} \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \\ Q^0 = \mu, Q \circ H^{-1} = \nu_\epsilon}} \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] = \inf_{\substack{Q \in \mathcal{P}(\mathcal{C}) \\ Q^0 = \mu, Q \circ H^{-1} = \nu}} \mathbb{E}^Q \left[\int_0^1 g(t, \dot{W}(t)) dt \right].$$

Furthermore, we have:

- *The problem*

$$(20) \quad \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \\ Q^0 = \mu, Q \circ H^{-1} = \nu_\epsilon}} \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right],$$

has an optimizer as soon as $\{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) : Q^0 = \mu, Q \circ H^{-1} = \nu_\epsilon\} \neq \emptyset$. Analogously,

$$(21) \quad \inf_{\substack{Q \in \mathcal{P}(\mathcal{C}) \\ Q^0 = \mu, Q \circ H^{-1} = \nu}} \mathbb{E}^Q \left[\int_0^1 g(t, \dot{W}(t)) dt \right],$$

has an optimizer as soon as $\{Q \in \mathcal{P}(\mathcal{C}) : Q^0 = \mu, Q \circ H^{-1} = \nu\} \neq \emptyset$.

- *If for all $\epsilon > 0$ small, an optimizer Q_ϵ of (20) exists, then any cluster point of $\{Q_\epsilon\}_\epsilon$ is an optimizer of (21). In particular, if the latter problem has a unique optimizer, then any cluster point of $\{Q_\epsilon\}_\epsilon$ is equal to it.*

²We adopt the convention that infimum over an empty set equals $+\infty$.

The connection with optimal transport arises from working with the classical observable, $H(\omega) = \omega(1)$ with $Z = \mathbb{R}^d$, and a time-independent function $g = g(q)$. Then, from Jensen’s inequality (as g is always convex under assumption (TI)) it is clear that the right-hand side of (19) becomes

$$\inf_{\substack{Q \in \mathcal{P}(C) \\ Q^0 = \mu, Q^1 = \nu}} \mathbb{E}^Q \left[\int_0^1 g(\dot{W}(t)) dt \right] = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} g(y - x) \pi(dx, dy).$$

In other words, by modifying the notion of entropy used in the definition of the Schrödinger problem, we obtain a different optimal transport cost in the limit, with the cost function given by precisely the function g governing the entropy. Interestingly, Léonard [28] shows that one can also obtain a different transport cost by changing the Brownian motion to another Markov process, while sticking with the usual entropy.

We stress that introducing the *mollified measures* ν_ϵ is in general unavoidable in Theorem 2.8. For instance, in the classical case (17), and when μ and ν are discrete, the value in the left-hand side of (19) is $+\infty$ whereas the right-hand side could very well be finite. Our second result on the matter shows that the mollification can be avoided when g is strictly subquadratic.

THEOREM 2.9. *Let us assume that $g(t, q) = g(q)$ and consider the classical observable*

$$Z = \mathbb{R}^d \quad \text{and} \quad H(\omega) = \omega(1).$$

Assume that g satisfies

$$(22) \quad \limsup_{|q| \rightarrow \infty} \frac{g(q)}{|q|^r} < \infty \quad \text{for some } r \in (1, 2),$$

and also that g has the Δ_2 doubling property, that is, there exist $R_0, C_0 > 0$ such that $g(2q) \leq C_0 g(q)$ for all $|q| \geq R_0$. Then all conclusions of Theorem 2.8 are valid when we take $\nu_\epsilon \equiv \nu$ for all ϵ .

In the important special case $g(q) = |q|^p$ for $1 < p < 2$, which indeed satisfies (TI) and the assumption of Theorem 2.9, we find in particular that, for $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\lim_{\epsilon \downarrow 0} \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(C) \\ Q^0 = \mu, Q^1 = \nu}} \mathbb{E}^Q \left[\int_0^1 |q^Q(t)|^p dt \right] = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy),$$

where we recognize the right-hand side as the p th power of the p -Wasserstein distance. It is worth noting that, assuming (22), a sufficient condition for the Δ_2 doubling property is that $\liminf_{|q| \rightarrow \infty} \frac{g(q)}{|q|^r} > 0$.

Let us explain how our results relate to prior literature. It was established by Mikami [32] that classical quadratic optimal transport is the small-noise limit of the so-called Schrödinger problem. This was then extended by Mikami and Thieullen [33] to nonquadratic situations. In this case, optimal transport is obtained as a small-noise limit of a *stochastic transport problem*. This latter stochastic variant can be interpreted as a nonexponential, Schrödinger-type problem. The method employed by the authors relies on PDE techniques and the given Brownian setting that they propose. On the other hand, Léonard [27] extended these considerations to a non-Brownian setting by employing large deviations arguments instead of PDEs; his is therefore a fully probabilistic approach, which was further developed in [28, 29]. However, the approach of Léonard, when applied in the aforementioned Brownian setting of Mikami–Thieullen, can only cover the quadratic case.

We draw inspiration in Léonard’s fully probabilistic approach of [27], crucially applying our generalized Schilder-type result (Theorem 2.2). By working probabilistically and not with PDEs, we avoid the regularity assumptions imposed in [33]. In particular, our Theorem 2.8 extends [33], Theorem 3.2, in the sense that we allow for a time-dependent function g with nearly no regularity assumptions as well as a general observable H .

Regarding our second result, Theorem 2.9, the closest counterpart in the literature is [32], Proposition 2.1, which shows that in the quadratic case (17) one can still take $\nu_\epsilon = \nu$ under the additional assumption that ν has finite relative entropy with respect to Lebesgue measure, resulting in

$$\lim_{\epsilon \downarrow 0} \inf_{\substack{Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \\ Q^0 = \mu, Q^1 = \nu}} \epsilon H(Q|P_\epsilon) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2} |x - y|^2 \pi(dx, dy).$$

In fact, this fails without an additional assumption on ν . For instance, if ν has finite support, then the right-hand side is finite, whereas for any $Q \in \mathcal{P}_\epsilon^*(\mathcal{C})$ with $Q^1 = \nu$ we have $H(Q|P_\epsilon) = \infty$ for each $\epsilon > 0$ because Q must be singular with respect to P_ϵ . Interestingly, Theorem 2.9 states that the mollification can be completely avoided under no additional assumptions on ν , by working with a strictly subquadratic entropy. Our proof makes crucial use of Brownian bridges, which give rise to drifts which are not square-integrable but are r -integrable for $r < 2$.

The reader who is mostly interested in these results on the Schrödinger-type problem may skip directly to Section 8, or first to Section 7.2 for some technical preparations.

2.4. Connections with PDEs. In this subsection we specialize the limit theorems to functions F on $\mathcal{P}(\mathcal{C})$ (resp. \mathcal{C}) which depend only on the time-1 marginal of the measure (resp. the time-1 value of the path). In this case, the so-called *nonlinear Feynman–Kac formula* (see the recent book [45], Section 5.1.3, for a typical case) allows to reinterpret the BSDE results of Section 2.2 in terms of vanishing viscosity limits for semilinear parabolic partial differential equations.

2.4.1. A PDE form of Theorem 2.7. As a first special case, suppose the function F in Theorem 2.7 depends only on the final value of the path; that is, $F(w) = f(w(1))$ for all $w \in \mathcal{C}$, for some $f \in C_b(\mathbb{R}^d)$. Then, according to [12], Theorem 5.2, we can write $Y_n(t) = v_n(t, W(t))$, where $v_n : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$ solves (i.e., is the minimal viscosity supersolution of) the Hamilton–Jacobi–Bellman PDE

$$(23) \quad \begin{cases} \partial_t v_n(t, x) + \frac{1}{2} \Delta v_n(t, x) + g^*\left(t, \frac{1}{\sqrt{n}} \nabla v_n(t, x)\right) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ v_n(1, x) = f\left(\frac{x}{\sqrt{n}}\right) & \text{for } x \in \mathbb{R}^d, \end{cases}$$

where the gradient and Laplacian operators act on the x variable. Alternatively, defining $u_n(t, x) = v_n(t, \sqrt{n}x)$, we find that u_n should solve the PDE

$$\begin{cases} \partial_t u_n(t, x) + \frac{1}{2n} \Delta u_n(t, x) + g^*(t, \nabla u_n(t, x)) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ u_n(1, x) = f(x) & \text{for } x \in \mathbb{R}^d. \end{cases}$$

In this PDE, the factor n appears only in the denominator of the diffusion coefficient, and as $n \rightarrow \infty$ we expect u_n to converge to the solution u of the first-order PDE

$$(24) \quad \begin{cases} \partial_t u(t, x) + g^*(t, \nabla u(t, x)) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ u(1, x) = f(x) & \text{for } x \in \mathbb{R}^d. \end{cases}$$

If $g(t, x) = g(x)$ is time-independent, the solution should be given by the Hopf–Lax–Oleinik formula,

$$(25) \quad u(t, x) = \sup_{y \in \mathbb{R}^d} \left(f(y) - (1 - t)g\left(\frac{y - x}{1 - t}\right) \right).$$

We then obtain

$$\lim_{n \rightarrow \infty} v_n(0, 0) = \lim_{n \rightarrow \infty} u_n(0, 0) = u(0, 0) = \sup_{x \in \mathbb{R}^d} (f(x) - g(x)),$$

which agrees with the limiting expressions Theorems 2.2 and 2.7. We will expand and formalize these heuristics in Proposition 6.4 below. Noting that $Y_n(t, \omega) = u_n(t, \omega(t)/\sqrt{n})$, this explains the choice of scaling in the first claimed limit of Theorem 2.7.

2.4.2. *Path-dependent PDEs.* It is tempting to search for a PDE formulation of Theorem 2.7, analogous to the discussion in Section 2.4.1. Indeed, the quantity $u_n(t, \omega)$ in Theorem 2.7 can be viewed as the value function of a stochastic control problem with a path-dependent objective functional, and Theorem 2.7 identifies the limiting function $u(t, \omega)$ as itself the value of a deterministic control problem. In analogy with Section 2.4.1, we speculate that Theorem 2.7 could be rewritten as a vanishing viscosity limit of path-dependent Hamilton–Jacobi–Bellman equations, but this is beyond the scope of this paper. Refer to [2, 15, 30] and the references therein for relevant literature on path-dependent PDEs and particularly to [31] where a connection with large deviations appears.

2.4.3. *A PDE form of Theorem 2.6.* In the general context of Theorem 2.6, when $F \in C_b(\mathcal{P}(\mathcal{C}))$ depends on the whole path, the BSDE of Theorem 2.6 cannot be expressed using PDEs. However, when F depends only on the marginal law at the final time, that is, $F = F(m(1))$ for some $F \in C_b(\mathcal{P}(\mathbb{R}^d))$, a different PDE representation is available. The terminal condition in the BSDE of Theorem 2.6 becomes

$$F\left(\frac{1}{n} \sum_{k=1}^n \delta_{W_{(n,k)}(1)}\right) = F\left(\frac{1}{n} \sum_{k=1}^n \delta_{\sqrt{n}(W(k/n) - W((k-1)/n))}\right).$$

This terminal condition depends on the path of W only through the values of $W(t)$ at the finitely many time points $t = 1/n, 2/n, \dots, 1$. Hence, the BSDE of Theorem 2.6 can be seen as a concatenation of n Markovian BSDEs, each of which can be represented by a PDE.

More details will be given in Section 6, specifically in Proposition 6.2, but let us briefly summarize the idea. Define an operator \mathbb{L}_n , taking lower semicontinuous lower bounded functions of $(\mathbb{R}^d)^n$ to lower semicontinuous lower bounded functions of $(\mathbb{R}^d)^{n-1}$, as follows: Given $f : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ and $(x_1, \dots, x_{n-1}) \in (\mathbb{R}^d)^{n-1}$, we define $\mathbb{L}_n f(x_1, \dots, x_{n-1}) := v(0, 0)$, where $v = v(t, x)$ is the the minimal viscosity supersolution of the PDE

$$\begin{cases} \partial_t v(t, x) + \frac{1}{2} \Delta v(t, x) + g^*(t, \nabla v(t, x)) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ v(1, x) = f(x_1, \dots, x_{n-1}, x) & \text{for } x \in \mathbb{R}^d. \end{cases}$$

For $n = 1$, we interpret \mathbb{L}_1 as mapping from functions of \mathbb{R}^d to real numbers. The composition $\mathbb{L}_1 \cdots \mathbb{L}_{n-1} \mathbb{L}_n$ then maps a function of $(\mathbb{R}^d)^n$ to a real number. For $F \in C_b(\mathcal{P}(\mathbb{R}^d))$, define $F^n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ by

$$F^n(x_1, \dots, x_n) := F\left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}\right).$$

Then we have $\frac{1}{n} \mathbb{L}_1 \cdots \mathbb{L}_{n-1} \mathbb{L}_n(nF^n) = Y_n(0)$, where Y_n is as in Theorem 2.6.

2.5. *Approximation schemes for (mean field) optimal control problems.* Interpreting the quantity $\rho^g(F)$ as well as the limiting expressions of Section 2.1 as the values of optimal control problems suggests certain numerical schemes for mean field stochastic control problems. We stress that by allowing the function g to be $+\infty$ -valued, we can induce pointwise control constraints in these problems.

The limiting quantity in Theorem 2.1, or in Theorem 2.6, is a stochastic optimal control problem of mean field type. Indeed, one may express this limit quantity as

$$(26) \quad \sup_{q \in \mathcal{L}_b} \left\{ F(P \circ (X^q)^{-1}) - \mathbb{E} \left[\int_0^1 g(t, q(t)) dt \right] \right\},$$

where we define

$$(27) \quad X^q(t) = \int_0^t q(s) ds + W(t).$$

This kind of optimization problem has been the subject of active research in recent years, with most of the literature focused on solution techniques, using either maximum principles [1, 6] or infinite-dimensional Hamilton–Jacobi–Bellman equations [25, 37]. In this literature, the function g or the coefficients of the SDE for X sometimes depend additionally on X and even its law. In this sense, we encounter in this paper only a special type of mean field control problem, but one which nonetheless includes many noteworthy examples, such as mean-variance optimization problems.

A mean field control problem such as (26) arises heuristically as an $n \rightarrow \infty$ (mean field) limit of an optimal control problem consisting of n state processes, described as follows:

$$(28) \quad \sup_{(q_1, \dots, q_n)} \mathbb{E} \left[F \left(\frac{1}{n} \sum_{k=1}^n \delta_{X_k} \right) - \frac{1}{n} \sum_{k=1}^n \int_0^1 g(t, q_k(t)) dt \right],$$

where the supremum is over progressively measurable square-integrable processes (q_1, \dots, q_n) , adapted to the filtration generated by n independent Brownian motions W_1, \dots, W_n , with the state processes X_k defined by

$$X_k(t) := \int_0^t q_k(s) ds + W_k(t).$$

The optimal value in (28) should converge to the optimal value in (26), as was rigorously justified only recently in [24], at least for certain functions F . The n -particle control problem (28) is arguably more amenable to numerical approximation than the mean field counterpart (26), as (finite-dimensional) dynamic programming and PDE methods are available for the former; the jury is still out on this question, but see [7] and references therein for direct perspectives on problems like (26).

Interestingly, our Theorem 2.1 provides an alternative approximation for (26) which could presumably be the basis for a numerical scheme. In particular, the pre-limit expression in Theorem 2.1 can be written as the value of a stochastic control problem:

$$(29) \quad \sup_{q \in \mathcal{L}_b} \mathbb{E} \left[F \left(\frac{1}{n} \sum_{k=1}^n \delta_{X_{(n,k)}^q} \right) - \int_0^1 g \left(nt - \lfloor nt \rfloor, \frac{q(t)}{\sqrt{n}} \right) dt \right],$$

with X^q as in (27). The potential advantage, compared to the n -particle approximation of the previous paragraph, is that here there is only one controlled process. The trade-off, however, is that the control problem (29) is inevitably highly path-dependent. If we assume $F \in C_b(\mathcal{P}(\mathcal{C}))$ depends only on the time-1 marginal of the measure, then the n -particle problem (28) becomes Markovian, whereas our approximation (29) remains path-dependent, as the cost function depends on the value of the state process at the n grid points

$(X^q(1/n), X^q(2/n), \dots, X^q(1))$. We discussed in Section 2.4.3 (with full details to come in Section 6) how one can essentially still apply dynamic programming and PDE methods to this kind of non-Markovian control problem. A proper exploration of the numerical feasibility of this approach, however, is beyond the scope of this already long paper, so we pursue this no further.

The two approximations (28) and (29) may appear more closely related than they truly are. On the one hand, in (29), we may interpret $X_{(n,k)}$ for $k = 1, \dots, n$ as playing the role of the n particles in (28). Indeed, these chopped paths are driven by the independent Brownian motions $W_{(n,k)}$. However, in (29), the control $q(t)$ in the time interval $t \in [k/n, (k + 1)/n]$ is allowed to depend on the entire past of the process $(X_s)_{s \leq t}$, which includes the entire paths $(X_{(n,1)}, \dots, X_{(n,k)})$ on the entire interval $[0, 1]$. On the other hand, in (28), the control $q_k(t)$ of particle k at time t depends on the paths of all particles up to time t , or $(X_1(s), \dots, X_n(s))_{s \leq t}$.

2.6. *Outline of the remainder of the paper.* The rest of the paper is devoted to proving the results stated above. First, Section 3 proves the variational formula (BBD) and then uses it to prove Theorem 2.2. Section 4 gives the more involved proof of Theorem 2.1. The remaining four sections address the applications, beginning with BSDEs and PDEs in Sections 5 and 6, respectively. Section 7 gives some modest extensions of our main results, in particular to allow for nonrandom initial states, which is crucial in proving our results on Schrödinger problems in the final Section 8.

3. The stochastic control representation. This section is devoted to the stochastic control representation of ρ^g , already hinted at in (BBD). In fact, we will establish a stronger result. In the following, the total variation metric on $\mathcal{P}(\mathcal{C})$ is defined by $(Q, Q') \mapsto \sup \int f d(Q - Q')$, where the supremum is over measurable functions $f : \mathcal{C} \rightarrow [-1, 1]$.

THEOREM 3.1. *Let $H : \mathcal{P}(\mathcal{C}) \rightarrow \mathbb{R}$ be bounded and continuous with respect to total variation, then*

$$(30) \quad \sup_{Q \in \mathcal{Q}} \{H(Q) - \alpha^g(Q)\} = \sup_{q \in \mathcal{L}_b} \left\{ H(Q^q) - \mathbb{E} \left[\int_0^1 g(t, q(t)) dt \right] \right\},$$

where

$$Q^q := P \circ \left(W + \int_0^\cdot q(t) dt \right)^{-1}.$$

In particular, if $F : \mathcal{C} \rightarrow \mathbb{R}$ is Borel measurable and bounded, then

$$(BBD) \quad \rho^g(F) = \sup_{q \in \mathcal{L}_b} \mathbb{E} \left[F \left(W + \int_0^\cdot q(t) dt \right) - \int_0^1 g(t, q(t)) dt \right].$$

Recall that in the quadratic case $g(t, q) := |q|^2/2$ we have $\rho^g(X) = \log \mathbb{E}[e^X]$, and equation (BBD) becomes the celebrated variational principle obtained in [4, 5, 17]. We stress that in such case, (BBD) has already proved to be a powerful tool in stochastic analysis, for example, in large deviations theory [5], in convex geometry (e.g., functional inequalities [26]) and in the study of convexity properties of Gaussian measure [4, 42]. For these reasons we employ the name *Borell–Boué–Dupuis formula* for the representation (BBD). On the other hand, for nonlinear H , the identity (30) seems to be novel even in the quadratic case and will be useful in the proofs of Theorem 2.1 and 2.2.

For the stochastic control connoisseur we stress that the formula (BBD) is a natural consequence of the definition of ρ^g (see 4) and the fact that optimizing over open-loop or closed-loop controls should yield the same optimal value. The difficulty lies mainly in the rather arbitrary path-dependence of F .

We prepare with a lemma which allows us to restrict the supremum in the definition of ρ^g to a more convenient class. In the following, recall that \mathcal{L}_b denotes the set of bounded progressively measurable functions $q : [0, 1] \times \mathcal{C} \rightarrow \mathbb{R}^d$. Let \mathcal{L}_b^s denote the set of $q \in \mathcal{L}_b$ such that the SDE

$$(31) \quad dX(t) = q(t, X) dt + dW(t), \quad X(0) = 0,$$

admits a unique strong solution. If Q denotes the law of X then $q = q^Q$. We find it useful, and intuitive, to overload the notation ρ^g in the following way: if $H : \mathcal{P}(\mathcal{C}) \rightarrow \mathbb{R}$ we write

$$\rho^g(H) := \sup_{Q \in \mathcal{Q}} \{H(Q) - \alpha^g(Q)\}.$$

This notation is only employed within this section of the article.

LEMMA 3.2. *Let $H : \mathcal{P}(\mathcal{C}) \rightarrow \mathbb{R}$ be as stated in Theorem 3.1. We have*

$$(32) \quad \rho^g(H) = \sup \left\{ H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] : Q \in \mathcal{Q}, q^Q \in \mathcal{L}_b^s \right\}.$$

PROOF. As g is bounded from below, we may assume without loss of generality that $g \geq 0$, by making an additive shift to both H and g . We make two intermediate approximations. First, define \mathcal{Q}_∞ to be the set of $Q \in \mathcal{Q}$ such that $\int_0^1 g(t, q^Q(t)) dt \in L^\infty(P)$. Let us show

$$(33) \quad \rho^g(H) = \sup_{Q \in \mathcal{Q}_\infty} \left\{ H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] \right\}.$$

To prove this, we first note that we may trivially restrict the supremum in the definition of $\rho^g(F)$ to those $Q \in \mathcal{Q}$ for which $\mathbb{E}^Q \int_0^1 g(t, q^Q(t)) dt < \infty$. Fix one such $Q \in \mathcal{Q}$. In the notation of (TI), we have $q^Q(t) \in \text{dom}(g(t, \cdot))$, $dt \otimes dP$ -a.e. Let $\tau_n = \inf\{t : \int_0^t g(s, q^Q(s)) ds > n\} \wedge 1$ and define $\frac{dQ_n}{dP} := \mathbb{E}[\frac{dQ}{dP} | \mathcal{F}_{\tau_n}]$, so that $q^{Q_n} = q_n$, where $q_n(t) := q^Q(t) 1_{\{t \leq \tau_n\}}$. We easily check that $dQ_n/dP \rightarrow dQ/dP$ in probability, and, by Scheffe’s lemma, in $L^1(P)$. This implies that $Q_n \rightarrow Q$ in total variation, and so $H(Q_n) \rightarrow H(Q)$. Moreover, $Q_n = Q$ on \mathcal{F}_{τ_n} , and we deduce

$$\begin{aligned} & H(Q_n) - \mathbb{E}^{Q_n} \left[\int_0^1 g(t, q^{Q_n}(t)) dt \right] \\ &= H(Q_n) - \mathbb{E}^Q \left[\int_0^{\tau_n} g(t, q(t)) dt \right] - \mathbb{E}^{Q_n} \left[\int_{\tau_n}^1 g(t, 0) dt \right] \\ &\rightarrow H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q(t)) dt \right]. \end{aligned}$$

With (33) established, we next show that in fact

$$(34) \quad \rho^g(H) = \sup \left\{ H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] : Q \in \mathcal{Q}_\infty, q^Q \in \mathcal{L}_b \right\}.$$

To prove this, fix $Q \in \mathcal{Q}_\infty$. We again have $q^Q(t) \in \text{dom}(g(t, \cdot))$, $dt \otimes dP$ -a.e. Define $q_n(t)$ as the projection of $q^Q(t)$ onto the centered ball of radius n , that is:

$$q_n(t) := q^Q(t) 1_{\{|q^Q(t)| \leq n\}} + \frac{n}{|q^Q(t)|} q^Q(t) 1_{\{|q^Q(t)| > n\}}.$$

Using convexity of $g(t, \cdot)$ and $g \geq 0$, we have

$$(35) \quad g(t, q_n(t)) \leq g(t, 0) + g(t, q^Q(t)).$$

For each (t, ω) it holds for all sufficiently large n that $q_n(t, \omega) = q^Q(t, \omega)$, and thus $g(t, q_n(t, \omega)) \rightarrow g(t, q^Q(t, \omega))$ pointwise. Find $Q_n \in \mathcal{Q}$ such that $q^{Q_n} = q_n$. Since $q_n \rightarrow q^Q$, we deduce, as in the previous step, that $dQ_n/dP \rightarrow dQ/dP$ in $L^1(P)$ and thus $Q_n \rightarrow Q$ in total variation. Thanks to (35) we may apply dominated convergence to get

$$H(Q_n) - \mathbb{E}^{Q_n} \left[\int_0^1 g(t, q^{Q_n}(t)) dt \right] \rightarrow H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right].$$

Now that we have proven (34), we show as a final approximation that

$$(36) \quad \rho^s(H) = \sup \left\{ H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] \right\},$$

where the supremum is taken over $Q \in \mathcal{Q}$ such that q^Q is a simple process. We say here that $q : [0, 1] \times \mathcal{C} \rightarrow \mathbb{R}^d$ is a *simple process* if there is a (deterministic) partition $t_0 < t_1 < \dots < t_N$ and bounded \mathcal{F}_{t_i} -measurable random variables ξ_i for which

$$q(t) = \xi_0 1_{\{0\}}(t) + \sum_{i=0}^{N-1} \xi_i 1_{(t_i, t_{i+1}]}(t).$$

We start from (34). Fix $Q \in \mathcal{Q}_\infty$ such that $q^Q \in \mathcal{L}_b$, noting that necessarily $q^Q(t) \in \text{dom}(g(t, \cdot))$ $dt \otimes dP$ -a.e. Suppose $|q^Q| \leq C$ pointwise, where $C < \infty$. Due to convexity and lower semicontinuity of $g(t, \cdot)$, upon making the further approximation $q_\epsilon(t) := \epsilon \bar{q}(t) + (1 - \epsilon)q^Q(t)$, with $\epsilon \in (0, 1)$ and for $\bar{q} \equiv 0 \in \text{ri}(\text{dom}(g(t, \cdot))) =: \mathcal{R}$, we can assume $q^Q(t) \in \mathcal{R}$. The convex set \mathcal{R} is, by assumption, independent of the time t . We now show that q^Q can be suitably approximated by measurable processes with continuous paths. First remark that q^Q can be identified with a measurable function on

$$E := \Psi([0, 1] \times \mathcal{C}),$$

where $\Psi(t, \omega) := (t, \omega(\cdot \wedge t))$. The space E is Polish, as a closed subset of the Polish space $[0, 1] \times \mathcal{C}$. By Lusin’s Theorem, there is for every k a closed set $E^k \subset E$ such that q^Q restricted to E^k is continuous and $dt \otimes dP(E^k) \geq 1 - 2^{-k}$. By the Tietze extension theorem [13], Theorem 4.1, we can find a continuous function q_k on E which coincides with q^Q when restricted to E^k and which takes values in the closed convex hull of $\{q^Q(t, \omega) : (t, \omega) \in E\}$. In particular, $q_k(t, \omega) \in \mathcal{R}$ and $|q_k(t, \omega)| \leq C$ for each (t, ω) . By Borel–Cantelli, q_k converges $dt \otimes dP$ -a.s. to q^Q . By further approximating each q_k , we may obtain the existence of a sequence of simple processes converging $dt \otimes dP$ -a.s. to q^Q , each of which still takes values in \mathcal{R} and is bounded uniformly by C . Let us re-brand by q_n this sequence of simple processes. It follows that $g(t, q_n(t)) \rightarrow g(t, q^Q(t))$, $dt \otimes dP$ -almost surely, since g is continuous in the relative interior of its domain. Since the sequence (q_n) is uniformly bounded, it follows from the assumption (2) that $\sup_n g(t, q_n(t)) \in L^1([0, 1], dt)$. By dominated convergence we then have

$$(37) \quad \int_0^1 g(t, q_n(t)) dt \rightarrow \int_0^1 g(t, q^Q(t)) dt \quad P\text{-a.s.}$$

Now find $Q_n \in \mathcal{Q}$ such that $q^{Q_n} = q_n$, and note as before that $dQ_n/dP \rightarrow dQ/dP$ in $L^1(P)$. The sequence $(\int_0^1 g(t, q_n(t)) dt)_n$ is essentially bounded thanks to (2). Hence $E^{Q_n}[\int_0^1 g(t, q_n(t)) dt] \rightarrow E^Q[\int_0^1 g(t, q^Q(t)) dt]$. Since q^{Q_n} is a simple process, this proves (36).

With (36) in hand, we complete the proof as follows. It is clear from the definition that $\rho^s(H)$ is larger than the right-hand side of (32). The reverse inequality follows from (36) and the fact that whenever q is a simple process in the sense described above, the SDE (31) admits a unique strong solution. \square

We can now provide the proof of Theorem 3.1. Our argument is reminiscent of [26].

PROOF OF THEOREM 3.1. We prove equation (30), establishing first the inequality “ \leq ”. By Lemma 3.2, we fix $Q \in \mathcal{Q}$ such that $q^Q \in \mathcal{L}_b^s$. Note that the completed filtrations of W and W^Q coincide, where $W^Q := W - \int_0^\cdot q^Q(t) dt$ is a Q -Brownian motion by Girsanov’s theorem. Hence, there exists $\tilde{q} \in \mathcal{L}_b$ such that $q^Q(t) = \tilde{q}(t, W^Q)$ and so $Q = P \circ (W + \int_0^\cdot \tilde{q}(t, W) dt)^{-1}$. Thus

$$\begin{aligned} H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t, W)) dt \right] &= H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, \tilde{q}(t, W^Q)) dt \right] \\ &= H \left(P \circ \left(W + \int_0^\cdot \tilde{q}(t, W) dt \right)^{-1} \right) - \mathbb{E} \left[\int_0^1 g(t, \tilde{q}(t, W)) dt \right] \\ &\leq \sup_{q \in \mathcal{L}_b} \left\{ H(Q^q) - \mathbb{E} \left[\int_0^1 g(t, q(t)) dt \right] \right\}. \end{aligned}$$

To prove the opposite inequality, let $q \in \mathcal{L}_b$, and set

$$X(t) = W(t) + \int_0^t q(s) ds = W(t) + \int_0^t q(s, W) ds.$$

Letting $\mathbb{F}^X = (\mathcal{F}_t^X)_{t \in [0,1]}$ denote the complete filtration generated by X , let us choose $\tilde{q} : [0, 1] \times \mathcal{C} \rightarrow \mathbb{R}^d$ to be any bounded progressively measurable function satisfying

$$\tilde{q}(t, X) = \mathbb{E}[q(t, W) \mid \mathcal{F}_t^X] \quad \text{a.s., for each } t \in [0, 1].$$

In particular, \tilde{q} may be defined via optional projection. It is well known [39], Exercise (5.15), that the *innovation process*

$$\tilde{W}(t) := X(t) - \int_0^t \tilde{q}(s, X) ds$$

is an \mathbb{F}^X -Brownian motion. Hence, if $Q := P \circ X^{-1}$, then $q^Q = \tilde{q}$ by Girsanov’s theorem. Using convexity of g and Jensen’s inequality, we conclude

$$\begin{aligned} H \left(P \circ \left(W + \int_0^\cdot q(t) dt \right)^{-1} \right) - \mathbb{E} \left[\int_0^1 g(t, q(t)) dt \right] &= H(Q) - \mathbb{E} \left[\int_0^1 g(t, q(t, W)) dt \right] \\ &\leq H(Q) - \mathbb{E} \left[\int_0^1 g(t, \tilde{q}(t, X)) dt \right] \\ &= H(Q) - \mathbb{E}^Q \left[\int_0^1 g(t, \tilde{q}(t, W)) dt \right] \\ &\leq \rho^g(H), \end{aligned}$$

where the last inequality follows from the identity $q^Q = \tilde{q}$ and the (overloaded) definition of ρ^g . As this inequality is valid for any $q \in \mathcal{L}_b$, the proof of equation (30) is complete. Finally, equation (BBD) follows since $Q \mapsto H(Q) := \mathbb{E}^Q[F(W)]$ is sequentially continuous in the desired way (if F is bounded and Borel) and $\rho^g(H) = \rho^g(F)$ of course. \square

The functional ρ^g can be extended to random variables $X \in L^0(P)$ that are bounded from below by setting $\rho^g(X) := \lim_{n \rightarrow \infty} \rho^g(X \wedge n)$. It is easily checked that this extension also satisfies (BBD), though we will make no use of this.

Using Theorem 3.1, we now prove the Schilder-type result of Theorem 2.2. The argument is reminiscent of the weak convergence proof of the Freidlin–Wentzell theorem [5], Theorem 4.3.

PROOF OF THEOREM 2.2. By (BBD) we have

$$\begin{aligned} \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) &= \sup_{q \in \mathcal{L}_b} \mathbb{E} \left[F \left(\frac{W + \int_0^1 q(t) dt}{\sqrt{n}} \right) - \int_0^1 g \left(t, \frac{q(t)}{\sqrt{n}} \right) dt \right] \\ &= \sup_{q \in \mathcal{L}_b} \mathbb{E} \left[F \left(\frac{W}{\sqrt{n}} + \int_0^1 q(t) dt \right) - \int_0^1 g(t, q(t)) dt \right]. \end{aligned}$$

We first bound the $\liminf_{n \rightarrow \infty}$ of the above expression. For each absolutely continuous $\omega \in \mathcal{C}_0$ such that $\int_0^1 g(t, \dot{\omega}(t)) dt < \infty$, define the absolutely continuous path $w_k \in \mathcal{C}_0$ by setting

$$\dot{w}_k(t) := \dot{\omega}(t) 1_{\{|\dot{\omega}(t)| \leq k\}} + \frac{k}{|\dot{\omega}(t)|} \dot{\omega}(t) 1_{\{|\dot{\omega}(t)| > k\}}, \quad k \geq 1.$$

Note that $w_k \in \mathcal{L}_b$. For every $k \in \mathbb{N}$ we have

$$\begin{aligned} (38) \quad \liminf_{n \rightarrow \infty} \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) &\geq \liminf_{n \rightarrow \infty} \mathbb{E} \left[F \left(\frac{W}{\sqrt{n}} + w_k \right) - \int_0^1 g(t, \dot{w}_k(t)) dt \right] \\ &\geq F(w_k) - \int_0^1 g(t, \dot{w}_k(t)) dt. \end{aligned}$$

By convexity of $g(t, \cdot)$, we have

$$g(t, \dot{w}_k(t)) \leq g(t, \dot{\omega}(t)) + g(t, 0) + 2b,$$

where $b \geq 0$ is a constant such that $g \geq -b$. Moreover, since $\dot{w}_k(t) = \dot{\omega}(t)$ for sufficiently large k , it holds that $w_k \rightarrow \omega$ and $g(t, \dot{w}_k(t)) \rightarrow g(t, \dot{\omega}(t))$ for every t . Thus, taking the limit as k goes to infinity in (38), it follows by dominated convergence (noting that $|\dot{w}_k| \leq |\dot{\omega}|$) that

$$\liminf_{n \rightarrow \infty} \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) \geq F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt.$$

Recalling the convention that $\int_0^1 g(t, \dot{\omega}(t)) dt := \infty$ whenever ω is not absolutely continuous, we may take the supremum over $\omega \in \mathcal{C}_0$ to get

$$\liminf_{n \rightarrow \infty} \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) \geq \sup_{\omega \in \mathcal{C}_0} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right).$$

For the opposite inequality, first notice that we may always choose a constant $q \equiv 0$ to get the lower bound

$$(39) \quad \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) \geq \mathbb{E} \left[F \left(\frac{W}{\sqrt{n}} \right) - \int_0^1 g(t, 0) dt \right] \geq -2C \quad \forall n \in \mathbb{N},$$

where $C < \infty$ is any constant such that $\inf_{\omega \in \mathcal{C}} F(\omega) \geq -C$ and $\int_0^1 g(t, 0) dt \leq C$ (see Assumption (TI)). Now, take q_n to be $1/n$ -optimal; that is, let $q_n \in \mathcal{L}_b$ be such that

$$(40) \quad \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) - \frac{1}{n} \leq \mathbb{E} \left[F \left(\frac{W}{\sqrt{n}} + \int_0^1 q_n(t) dt \right) - \int_0^1 g(t, q_n(t)) dt \right].$$

From (39), we have

$$(41) \quad \sup_n \mathbb{E} \int_0^1 g(t, q_n(t)) dt < \infty.$$

Letting $A_n(t) = \int_0^t q_n(s) ds$, it follows from Lemma A.1 that the sequence (A_n) of C_0 -valued random variables is tight. Moreover, if we fix a subsequence A_{n_k} which converges in law to some A , then we may write $A = \int_0^1 q(t) dt$ for some process q satisfying

$$\mathbb{E} \int_0^1 g(t, q(t)) dt \leq \liminf_{k \rightarrow \infty} \mathbb{E} \int_0^1 g(t, q_{n_k}(t)) dt.$$

Because $\lim_{n \rightarrow \infty} W/\sqrt{n} = 0$ in probability, we have $W/\sqrt{n_k} + A_{n_k} \rightarrow A$ in law. Recalling (40), we have (taking limits still along the same subsequence)

$$\begin{aligned} \limsup_{k \rightarrow \infty} \rho^{g_{n_k}} \left(F \left(\frac{W}{\sqrt{n_k}} \right) \right) &\leq \limsup_{k \rightarrow \infty} \mathbb{E} \left[F \left(\frac{W}{\sqrt{n_k}} + A_{n_k} \right) - \int_0^1 g(t, q_{n_k}(t)) dt \right] \\ &\leq \mathbb{E} \left[F(A) - \int_0^1 g(t, q(t)) dt \right] \\ &= \mathbb{E} \left[F \left(\int_0^1 q(t) dt \right) - \int_0^1 g(t, q(t)) dt \right] \\ &\leq \sup_{\omega \in C_0} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right). \end{aligned}$$

We have argued that for any subsequence we can extract a further subsequence along which the above limsup bound is valid, and we conclude that the same upper bound is valid without passing to a subsequence. This completes the proof. \square

4. The Sanov-type limit theorem. This section develops the necessary machinery for proving Theorem 2.1, some of which will be used again in later sections. The goal is to write our problem in a setting amenable to [23], Theorem 1.1. A first key step is to use Theorem 3.1 to derive an alternative expression for the pre-limit quantity in Theorem 2.1, relating it to the iterates denoted ρ_n in [23], and this will explain the precise form of the scaling limit. This is carried out in Section 4.1. A second key ingredient in applying [23] is to check that the sub-level sets of α^g are weakly compact, which turns out to fail in general. Section 4.2 provides a suitable work-around. Finally, Section 4.3 assembles these pieces into a complete proof.

4.1. The rescaled control problem. Let C^n be the n -fold product space, and denote by $(\omega_1, \dots, \omega_n)$ a typical element in C^n . Let $B_b(C^n)$ be the space of bounded measurable functions on C^n . We define inductively the iterates of $\rho_n^g : B_b(C^n) \rightarrow \mathbb{R} \cup \{+\infty\}$ as follows: We set $\rho_1^g \equiv \rho^g$, and for $n > 1$ define

$$(42) \quad \rho_n^g(f) := \rho_{n-1}^g((\omega_1, \dots, \omega_{n-1}) \mapsto \rho(f(\omega_1, \dots, \omega_{n-1}, \cdot))).$$

In other words, given $f \in B_b(C^n)$ for $n > 1$, we define³ $\tilde{f} \in B_b(C^{n-1})$ by $\tilde{f}(\omega_1, \dots, \omega_{n-1}) = \rho^g(f(\omega_1, \dots, \omega_{n-1}, \cdot))$, and then we set $\rho_n^g(f) = \rho_{n-1}^g(\tilde{f})$.

Recall from (6) the definition of the chopped paths $W_{(n,k)}$ for $k = 1, \dots, n$. The following representation for ρ_n^g underlies our proof of Theorem 2.1:

³Actually, the function \tilde{f} is merely upper-semianalytic in general. But this does not pose any problems, since upper-semianalytic functions are universally measurable.

PROPOSITION 4.1. For $q \in \mathcal{L}_b$ define $X^q = W + \int_0^\cdot q(t) dt$. For $f \in B_b(\mathbb{C}^n)$, we have

$$\begin{aligned} \rho_n^g(f) &= \sup_{q \in \mathcal{L}_b} \mathbb{E} \left[f(X_{(n,1)}^q, \dots, X_{(n,n)}^q) - n \int_0^1 g\left(nt - \lfloor nt \rfloor, \frac{q(t)}{\sqrt{n}}\right) dt \right] \\ &= n\rho^{G_n} \left(\frac{1}{n} f(W_{(n,1)}, \dots, W_{(n,n)}) \right). \end{aligned}$$

PROOF. The second claimed equality follows immediately from Theorem 3.1 and the definition of G_n , so we prove only the first.

For $n = 1$ this is Theorem 3.1. Fix $n > 1$. Define a process $B^n : [0, n] \times \mathbb{C}^n \rightarrow \mathbb{R}^d$ by setting

$$\begin{aligned} B^n(t, \omega_1, \dots, \omega_n) &= \begin{cases} \omega_1(t) - \omega_1(0) & \text{if } t \in [0, 1], \\ \omega_{k+1}(t - k) - \omega_{k+1}(0) + \sum_{i=1}^k [\omega_i(1) - \omega_i(0)] & \text{if } t \in [k, k + 1], k \leq n - 1. \end{cases} \end{aligned}$$

In other words, $B^n(t, \omega_1, \dots, \omega_n)$ follows the increments of ω_k on the interval $[k - 1, k]$. Define the filtration \mathbb{F}^n on \mathbb{C}^n by setting $\mathcal{F}_t^n = \sigma(B_s^n : s \leq t)$. Note that $B^n = (B^n(t))_{t \in [0, n]}$ is a Brownian motion on $(\mathbb{C}^n, \mathbb{F}^n, P^n)$ with P^n the n -fold product of P . In the following, the symbol \mathbb{E}^n will denote expectation on $(\mathbb{C}^n, \mathbb{F}^n, P^n)$, and we note that $\mathbb{E} = \mathbb{E}^1$.

Let \mathcal{A}_n denote the set of bounded \mathbb{F}^n -progressively measurable processes $q : [0, n] \times \mathbb{C}^n \rightarrow \mathbb{R}^d$. For $q \in \mathcal{A}_n$, define a continuous process $X^{n,q} = (X^{n,q}(t))_{t \in [0, n]}$ on $(\mathbb{C}^n, \mathbb{F}^n, P^n)$ by

$$X^{n,q}(t, \omega_1, \dots, \omega_n) := \int_0^t q(s, \omega_1, \dots, \omega_n) ds + B^n(t, \omega_1, \dots, \omega_n).$$

In the following, for a path $x \in C([0, n]; \mathbb{R}^d)$ and for $k = 1, \dots, n$, define the chopped (but not rescaled) path $x_{(c,n,k)} \in \mathcal{C} = C([0, 1]; \mathbb{R}^d)$ by

$$x_{(c,n,k)}(t) = x(k - 1 + t) - x(k - 1), \quad t \in [0, 1].$$

In other words, $x_{(c,n,k)}$ is simply the increment over the time interval $[k - 1, k]$.

Let us understand first the case $n = 2$. For a fixed $\omega \in \mathcal{C}$, by Theorem 3.1 we have

$$\rho^g(f(\omega, \cdot)) = \sup_{q \in \mathcal{A}_1} \mathbb{E}^1 \left[f(\omega, X^{1,q}) - \int_0^1 g(t, q(t)) dt \right].$$

Applying Theorem 3.1 once again, we have by definition

$$\begin{aligned} \rho_2^g(f) &= \sup_{\beta \in \mathcal{A}_1} \mathbb{E}^1 \left[\rho^g(f(X^{1,\beta}, \cdot)) - \int_0^1 g(t, \beta(t)) dt \right] \\ &= \sup_{\beta \in \mathcal{A}_1} \mathbb{E}^1 \left[\left(\sup_{q \in \mathcal{A}_1} \mathbb{E} \left[f(\omega, X^{1,q}) - \int_0^1 g(t, q(t)) dt \right] \right) \Big|_{\omega = X^{1,\beta}} - \int_0^1 g(t, \beta(t)) dt \right]. \end{aligned}$$

The key idea here is to apply a form of dynamic programming. In particular, let $\widehat{\mathcal{A}}_1$ denote the set of bounded functions $[0, 1] \times \mathcal{C} \times \mathcal{C} \ni (t, \omega_1, \omega_2) \mapsto \widehat{q}[\omega_1](t, \omega_2) \in \mathbb{R}^d$ which are jointly measurable, using the progressive σ -field on $[0, 1] \times \mathcal{C}$ for the argument (t, ω_2) and the Borel σ -field on \mathcal{C} for the argument ω_1 . A standard measurable selection argument [3], Proposition 7.50, lets us write the above as

$$\begin{aligned} (43) \quad \rho_2^g(f) &= \sup_{\substack{\beta \in \mathcal{A}_1 \\ \widehat{q} \in \widehat{\mathcal{A}}_1}} \mathbb{E}^1 \left[\mathbb{E}^1 \left[f(\omega_1, X^{1, \widehat{q}[\omega_1]}) - \int_0^1 g(t, \widehat{q}[\omega_1](t)) dt \right] \Big|_{\omega_1 = X^{1,\beta}} \right. \\ &\quad \left. - \int_0^1 g(t, \beta(t)) dt \right]. \end{aligned}$$

Now consider a fixed $\widehat{q} \in \widehat{\mathcal{A}}_1$ and $\beta \in \mathcal{A}_1$. We may define a process $q : [0, 2] \times \mathcal{C}^2 \rightarrow \mathbb{R}^d$ by setting

$$(44) \quad q(t, \omega_1, \omega_2) = \beta(t, \omega_1)1_{[0,1]}(t) + \widehat{q}[\omega_1](t - 1, \omega_2)1_{(1,2]}(t).$$

Then $q \in \mathcal{A}_2$, and unpacking the definitions reveals the identities

$$\begin{aligned} X^{1, \widehat{q}[\omega_1]}(t, \omega_2) &= X_{(c,2,2)}^{2,q}(t, \omega_1, \omega_2), \quad t \in [0, 1], \\ X^{1, \beta}(t, \omega_1) &= X_{(c,2,1)}^{2,q}(t, \omega_1, \omega_2), \quad t \in [0, 1], \\ \mathbb{E}^1 \int_0^1 g(t, \beta(t)) dt &= \mathbb{E}^2 \int_0^1 g(t, q(t)) dt, \end{aligned}$$

which in turn imply

$$\begin{aligned} &\mathbb{E}^1 \left[f(\omega_1, X^{1, \widehat{q}[\omega_1]}) - \int_0^1 g(t, \widehat{q}[\omega_1](t)) dt \right] \Big|_{\omega_1 = X^{1, \beta}} \\ &= \mathbb{E}^2 \left[f(\omega_1, X_{(c,2,2)}^{2,q}(\omega_1, \cdot)) - \int_1^2 g(t - 1, q(t, \omega_1, \cdot)) dt \right] \Big|_{\omega_1 = X^{1, \beta}} \\ &= \mathbb{E}^2 \left[f(X_{(c,2,1)}^{2,q}, X_{(c,2,2)}^{2,q}) - \int_1^2 g(t - 1, q(t)) dt \mid \mathcal{F}_1^2 \right]. \end{aligned}$$

Indeed, the last identity follows from the fact that the \mathcal{C} -valued random variable $(\omega_1, \omega_2) \mapsto \omega_2$ is independent of \mathcal{F}_1^2 . Finally, we plug this last expression into (43). Then, note that the map $(\beta, \widehat{q}) \mapsto q$ given by (44) defines a bijection between $\mathcal{A}_1 \times \widehat{\mathcal{A}}_1$ and \mathcal{A}_2 , and use the tower property of conditional expectation to get

$$\rho_2^g(f) = \sup_{q \in \mathcal{A}_2} \mathbb{E}^2 \left[f(X_{(c,2,1)}^{2,q}, X_{(c,2,2)}^{2,q}) - \int_0^2 g(t - \lfloor t \rfloor, q(t)) dt \right].$$

This argument adapts, mutatis mutandis, to the case of general $n > 1$, and we find

$$(45) \quad \rho_n^g(f) = \sup_{q \in \mathcal{A}_n} \mathbb{E}^n \left[f(X_{(c,n,1)}^{n,q}, \dots, X_{(c,n,n)}^{n,q}) - \int_0^n g(t - \lfloor t \rfloor, q(t)) dt \right].$$

To complete the proof, we rescale this control problem to live on the time interval $[0, 1]$ instead of $[0, n]$. Still working on the space $(\mathcal{C}^n, \mathbb{F}^n, P^n)$, define for each $q \in \mathcal{A}_n$ the process

$$\overline{X}^{n,q}(t) := \frac{1}{\sqrt{n}} X^{n,q}(nt) = \frac{1}{\sqrt{n}} \int_0^{nt} q(s) ds + \frac{1}{\sqrt{n}} B^n(nt) \quad \text{for } t \in [0, 1].$$

By a change of variables and Brownian scaling, we can write

$$\overline{X}^{n,q}(t) = \int_0^t \overline{q}(s) ds + \overline{B}^n(t),$$

where $\overline{q}(s) := \sqrt{n}q(ns)$, and $\overline{B}^n(t) := \frac{1}{\sqrt{n}} B^n(nt)$ is a Brownian motion. Another change of variables yields

$$\int_0^n g(t - \lfloor t \rfloor, q(t)) dt = n \int_0^1 g\left(nt - \lfloor nt \rfloor, \frac{\overline{q}(t)}{\sqrt{n}}\right) dt.$$

Lastly, it is straightforward to check that $X_{(c,n,k)}^{n,q} \equiv \overline{X}_{(n,k)}^{n,q}$. Putting it all together, (45) becomes

$$\rho_n^g(f) = \sup_{q \in \mathcal{A}_n} \mathbb{E}^n \left[f(\overline{X}_{(n,1)}^{n,q}, \dots, \overline{X}_{(n,n)}^{n,q}) - n \int_0^1 g\left(nt - \lfloor nt \rfloor, \frac{\overline{q}(t)}{\sqrt{n}}\right) dt \right].$$

Complete the proof by transferring everything from the probability space $(\mathcal{C}^n, \mathbb{F}^n, P^n)$ to the original space $(\mathcal{C}, \mathbb{F}, P)$, using the map $\mathcal{C}^n \ni (\omega_1, \dots, \omega_n) \mapsto \overline{B}^n(\omega_1, \dots, \omega_n) \in \mathcal{C}$. \square

4.2. *In search of compactness.* As mentioned above, the goal of this section is to overcome the technical impediment that the functional α^g does not necessarily have compact sub-level sets. We illustrate this with an example, but we stress that this is only an issue when we do not assume that g has at least quadratic growth.

EXAMPLE 4.2. Take $d = 1$ and $g(t, q) = |q|^{5/4}$. Set $q_n(t) := t^{-3/4} \mathbf{1}_{(1/n, 1]}(t)$ and $q_\infty(t) := t^{-3/4}$. Define Q_n as the measure with density

$$\frac{dQ_n}{dP} = \exp\left(\int_0^1 q_n(t) dW(t) - \frac{1}{2} \int_0^1 |q_n(t)|^2 dt\right),$$

and let $Q_\infty = \text{Law}(W + \int_0^1 q_\infty(t) dt)$. One can easily check the following:

- (1) Q_n converges to Q_∞ in the weak topology of measures.
- (2) $\alpha^g(Q_n) \leq 16$ for each n .
- (3) Q_∞ is singular to P , so in particular $\alpha^g(Q_\infty) = \infty$.

This shows that the sublevel set $\{\alpha^g \leq 16\}$ is not even closed in the weak topology of measures.

For this reason, we initially replace α^g and ρ^g by two new functionals better suited for our purposes. Let \mathcal{P}^* denote the set of those measures Q on \mathcal{C} for which there exists a progressive \mathbb{R}^d -valued process q^Q such that $\int_0^1 |q^Q(s)| ds < \infty$ Q -a.s. and

$$W(t) - \int_0^t q^Q(s) ds \quad \text{is a } Q\text{-Brownian motion.}$$

The process q^Q is then uniquely defined up to $dt \otimes dQ$ -almost everywhere equality. This does not reduce to Girsanov theory, since we are not asking that elements in \mathcal{P}^* be absolutely continuous with respect to Wiener measure (e.g., the set \mathcal{P}^* contains measures singular to P , such as the laws of Brownian bridges or Bessel processes). Note, however, that $\alpha^g(Q) = \tilde{\alpha}^g(Q)$ for $Q \in \mathcal{Q}$.

Consider the functional

$$(46) \quad \mathcal{P}^* \ni Q \mapsto \tilde{\alpha}^g(Q) := \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] \in \mathbb{R} \cup \{+\infty\},$$

where we define the functional as $+\infty$ outside of \mathcal{P}^* . Let $B_b(\mathcal{C})$ denote the set of bounded measurable functions on \mathcal{C} and define the functional

$$\tilde{\rho}^g(F) := \sup_{Q \in \mathcal{P}^*} (\mathbb{E}^Q[F] - \tilde{\alpha}^g(Q)), \quad F \in B_b(\mathcal{C}).$$

We now give some elementary facts about $\tilde{\alpha}^g$ which may seem folklore. We defer the rather technical proof of the next lemma to the [Appendix](#). Recall that we are assuming at all times that the given function g satisfies assumption (TI).

LEMMA 4.3. *The functional $\tilde{\alpha}^g$ is convex, lower semicontinuous with respect to weak convergence of measures on path space, and its sub-level sets are weakly compact in this topology. Furthermore, we have*

$$\tilde{\alpha}^g(Q) = \sup_{F \in B_b(\mathcal{C})} (E^Q[F] - \tilde{\rho}^g(F)) = \sup_{F \in C_b(\mathcal{C})} (E^Q[F] - \tilde{\rho}^g(F)), \quad Q \in \mathcal{P}^*.$$

In general ρ^g and $\tilde{\rho}^g$, just as α^g and $\tilde{\alpha}^g$, may differ. It is thus important to establish how ρ^g and $\tilde{\rho}^g$ are related. This is the content of the next result.

LEMMA 4.4. *If $F : \mathcal{C} \rightarrow \mathbb{R}$ is bounded and lower semicontinuous, then $\rho^g(F) = \tilde{\rho}^g(F)$.*

PROOF. Obviously $\tilde{\rho}^g \geq \rho^g$. Let $Q \in \mathcal{P}^*$ such that $\tilde{\alpha}^g(Q) < \infty$. We will exhibit a sequence Q^n of absolutely continuous measures such that

$$\liminf \{ \mathbb{E}^{Q^n}[F] - \alpha^g(Q_n) \} \geq \mathbb{E}^Q[F] - \tilde{\alpha}^g(Q),$$

which would establish the claim. Note that $\mathbb{E}^Q \int_0^1 g(t, q^Q(t)) dt = \tilde{\alpha}^g(Q) < \infty$ implies $q^Q(t) \in \text{dom}(g(t, \cdot))$, $dt \otimes dP$ -a.e. We know that $W^Q(t) := W(t) - \int_0^t q^Q(s) ds$ is a Q -Brownian motion. Define $q_n(t) = q^Q(t) 1_{\{|q^Q(t)| \leq n\}}$, and let Q_n denote the law of the process

$$X^n(t) = W^Q(t) + \int_0^t q_n(s) ds.$$

Note that q_n is uniformly bounded, and so $Q_n \in \mathcal{Q}$. Because $q_n(t) \rightarrow q^Q(t)$ for each t , it is clear that $Q_n \rightarrow Q$ weakly. Hence, by lower semicontinuity of F ,

$$\liminf_{n \rightarrow \infty} \mathbb{E}^{Q_n}[F] \geq \mathbb{E}^Q[F].$$

Finally, define $\hat{q}_n(t, X^n)$ and \widehat{W}^Q as the optional projections (under Q) of q_n and W^Q , respectively, on the filtration generated by X^n . Then \widehat{W}^Q remains a Brownian motion in this smaller filtration [39], Exercise (5.15). It follows that $q^{Q_n}(t, X^n) = \hat{q}_n(t, X^n)$, $dt \otimes dQ$ -almost surely. By convexity, we get

$$\begin{aligned} \alpha^g(Q_n) &= \mathbb{E}^{Q_n} \left[\int_0^1 g(t, q^{Q_n}(t, W)) dt \right] \\ &= \mathbb{E}^Q \left[\int_0^1 g(t, \hat{q}_n(t, X^n)) dt \right] \leq \mathbb{E}^Q \left[\int_0^1 g(t, q_n(t)) dt \right] \\ &= \mathbb{E}^Q \left[\int_0^1 (g(t, q(t)) 1_{\{|q^Q(t)| \leq n\}} + g(t, 0) 1_{\{|q^Q(t)| > n\}}) dt \right]. \end{aligned}$$

Since $\int_0^1 g(t, 0) dt < \infty$ by assumption (TI), we conclude from monotone convergence that

$$(47) \quad \limsup_n \alpha^g(Q_n) \leq \mathbb{E}^Q \left[\int_0^1 g(t, q(t)) dt \right] = \tilde{\alpha}^g(Q). \quad \square$$

Recalling the definition of the iterates ρ_n^g based on ρ^g and given in (42), we define the iterates $\tilde{\rho}_n^g$ based on $\tilde{\rho}^g$ in the same way. A simple consequence of Lemma 4.4 is that $\rho_n^g = \tilde{\rho}_n^g$ restricted to a large class of functions:

LEMMA 4.5. *Let $n \in \mathbb{N}$, and let $f : \mathcal{C}^n \rightarrow \mathbb{R}$ be lower semicontinuous and bounded. Then the functions $\mathcal{C}^{n-1} \ni (\omega_1, \dots, \omega_{n-1}) \mapsto \rho(f(\omega_1, \dots, \omega_{n-1}, \cdot))$ are lower-semicontinuous and bounded, for both $\rho = \rho^g$ and $\rho = \tilde{\rho}^g$. In particular, for such f we have $\rho_n^g(f) = \tilde{\rho}_n^g(f)$.*

PROOF. The case $n = 1$ is covered by Lemma 4.4. The general case follows by induction but for ease of presentation we consider only the case $n = 2$. Let us prove that $\omega \mapsto F(\omega) := \tilde{\rho}^g(f(\omega, \cdot))$ is lower semicontinuous. To wit, if $\omega_n \rightarrow \omega$ and $F(\omega_n) \leq c$ for all n , then by definition

$$\int f(\omega_n, \bar{\omega}) dQ(\bar{\omega}) - \tilde{\alpha}^g(Q) \leq c,$$

for all $Q \in \mathcal{P}^*$. Taking limit inferior here, and by Fatou's lemma and lower semicontinuity of f , we get

$$\int f(\omega, \bar{\omega}) dQ(\bar{\omega}) - \tilde{\alpha}^g(Q) \leq c.$$

Now taking supremum over Q we conclude $F(\omega) \leq c$. Moreover, because g is bounded from below and f is bounded, F too is bounded. The same reasoning can be applied to ρ^g . By Lemma 4.4 and the case $n = 1$ we have

$$\begin{aligned} \rho_2^g(f) &= \rho^g(\omega \mapsto \rho^g(f(\omega, \cdot))) = \tilde{\rho}^g(\omega \mapsto \rho^g(f(\omega, \cdot))) \\ &= \tilde{\rho}^g(\omega \mapsto \tilde{\rho}^g(f(\omega, \cdot))) = \tilde{\rho}_2^g(f). \end{aligned} \quad \square$$

4.3. *Proof of Theorem 2.1.* With the above machinery we can finally prove Theorem 2.1. Let us denote the empirical measure of the family $(\omega_1, \dots, \omega_n) \in \mathcal{C}^n$ by

$$L_n(\omega_1, \dots, \omega_n) := \frac{1}{n} \sum_{i \leq n} \delta_{\omega_i}$$

and recall the notation $\omega_{(n,k)}$ from (6). Apply Proposition 4.1 to get

$$\rho^{G_n} \left(F \left(\frac{1}{n} \sum_{k=1}^n \delta_{W_{(n,k)}} \right) \right) = \frac{1}{n} \rho_n^g(nF \circ L_n).$$

Since $F \circ L_n$ is clearly a continuous function on \mathcal{C}^n , Lemma 4.5 yields $\rho_n^g(nF \circ L_n) = \tilde{\rho}_n^g(nF \circ L_n)$. Now, because $\tilde{\alpha}^g$ is convex and has weakly compact sub-level sets, we may apply [23], Theorem 1.1 (taking note of the representation of [23], Proposition A.1) to get

$$\lim_{n \rightarrow \infty} \frac{1}{n} \tilde{\rho}_n^g(nF \circ L_n) = \sup_{Q \in \mathcal{P}(\mathcal{C})} (F(Q) - \tilde{\alpha}^g(Q)) = \sup_{Q \in \mathcal{P}^*} (F(Q) - \tilde{\alpha}^g(Q)).$$

To complete the proof, it remains to show that

$$(48) \quad \sup_{Q \in \mathcal{P}^*} (F(Q) - \tilde{\alpha}^g(Q)) = \sup_{Q \in \mathcal{Q}} (F(Q) - \alpha^g(Q)).$$

Indeed, this will prove the first equality of Theorem 2.1, while the second follows from Theorem 3.1. To prove (48), notice from the proof of Lemma 4.4 (specifically (47)) that the following holds: If $\tilde{\alpha}^g(Q) < \infty$, then there exist $Q_n \in \mathcal{Q}$ such that $Q_n \rightarrow Q$ weakly and $\limsup_n \alpha^g(Q_n) \leq \tilde{\alpha}^g(Q)$. From this and continuity of F we deduce (48).

5. BSDE scaling limits. This section is dedicated to the proofs of Theorems 2.7 and 2.6. We will make use of the following definitions. For a function g satisfying (TI) and for $t \in [0, 1)$, define $g^{(t)} : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ by

$$g^{(t)}(s, q) := (1 - t)g \left(t + s(1 - t), \frac{q}{\sqrt{1 - t}} \right).$$

Note that $g^{(t)}$ itself satisfies (TI), and so $\rho^{g^{(t)}}$ is well defined. Moreover, we define the operation $\otimes_t : \mathcal{C} \times \mathcal{C}_0 \rightarrow \mathcal{C}$ by

$$\omega \otimes_t \bar{\omega}(s) := \omega(s \wedge t) + \sqrt{1 - t} \bar{\omega} \left(\frac{s - t}{1 - t} \right) 1_{[t, 1]}(s).$$

We begin with the following crucial lemma, which shows how to express the (super-) solution process $Y(t)$ of a BSDE with generator g^* in terms of $\rho^{g^{(t)}}$.

LEMMA 5.1. *Let $F \in C_b(\mathcal{C})$, and let (Y, Z) be the minimal supersolution of*

$$dY(t) = -g^*(t, Z(t)) dt + Z(t) dW(t), \quad Y(1) = F(W).$$

Then, for $t \in [0, 1)$ and P -a.e. $\omega \in \mathcal{C}$, we have

$$Y(t, \omega) = \rho^{g^{(t)}}(F(\omega \otimes_t \cdot)).$$

PROOF. By Lemma A.2 (which is just a minor modification of [11], Theorem 3.4), it holds

$$Y(t) = \operatorname{ess\,sup}_{Q \in \mathcal{Q}_t} \mathbb{E}^Q \left[F(W) - \int_t^1 g(s, q^Q(s, W)) ds \mid \mathcal{F}_t \right],$$

where \mathcal{Q}_t is the set of those measures $Q \in \mathcal{Q}$ such that $Q = P$ on \mathcal{F}_t . Note that $q^Q \equiv 0$ on $[0, t]$ for $Q \in \mathcal{Q}_t$. Now, for a path $\omega \in \mathcal{C}$, define $\omega^{(t)} \in \mathcal{C}_0$ by

$$\omega^{(t)}(s) := \frac{1}{\sqrt{1-t}} (\omega(t + s(1-t)) - \omega(t)).$$

It is readily checked that $\omega \otimes_t \omega^{(t)} = \omega$ for $\omega \in \mathcal{C}$. Hence, for a.e. ω , we may write

$$(49) \quad Y(t, \omega) = \operatorname{ess\,sup}_{Q \in \mathcal{Q}_t} \mathbb{E}^Q \left[F(\omega \otimes_t W^{(t)}) - \int_t^1 g(s, q^Q(s, \omega \otimes_t W^{(t)})) ds \mid \mathcal{F}_t \right](\omega).$$

On the other hand, we can write

$$(50) \quad \rho^{g^{(t)}}(F(\omega \otimes_t \cdot)) = \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q \left[F(\omega \otimes_t W) - \int_0^1 g^{(t)}(s, q^Q(s, W)) ds \right].$$

With these preparations out of the way, we first show that

$$(51) \quad Y(t, \omega) \leq \rho^{g^{(t)}}(F(\omega \otimes_t \cdot)) \quad \text{for a.e. } \omega.$$

To see this, fix $Q \in \mathcal{Q}_t$. Define a measurable map $\mathcal{C} \ni \omega \mapsto Q_\omega \in \mathcal{P}(\mathcal{C})$ as a version of $Q(W^{(t)} \in \cdot \mid \mathcal{F}_t)(\omega)$. Recalling also that $q^Q = 0$ on $[0, t]$, and noting that $P(W^{(t)} \in \cdot \mid \mathcal{F}_t) = P$ a.s. by Brownian scaling, we have

$$\frac{dQ_\omega}{dP}(\omega^{(t)}) = \frac{dQ}{dP}(\omega) = \exp \left(\int_t^1 q^Q(s, \omega) d\omega(s) - \frac{1}{2} \int_t^1 |q^Q(s, \omega)|^2 ds \right), \quad \omega \in \mathcal{C}.$$

Now, for $\omega \in \mathcal{C}$ define $\tilde{q}_\omega : [0, 1] \times \mathcal{C}_0 \rightarrow \mathbb{R}^d$ by

$$\tilde{q}_\omega(s, \bar{\omega}) := \sqrt{1-t} q^Q(t + s(1-t), \omega \otimes_t \bar{\omega}).$$

Recalling that $\omega \otimes_t \omega^{(t)} = \omega$, by a change of variables we may write the above as

$$\frac{dQ_\omega}{dP}(\omega^{(t)}) = \exp \left(\int_0^1 \tilde{q}_\omega(s, \omega^{(t)}) d\omega^{(t)}(s) - \frac{1}{2} \int_0^1 |\tilde{q}_\omega(s, \omega^{(t)})|^2 ds \right).$$

We conclude that $Q_\omega \in \mathcal{Q}$ and $\tilde{q}_\omega = q^{Q_\omega}$ for a.e. ω . With these identifications and another change of variables in the time-integral, we can write

$$\begin{aligned} & \mathbb{E}^Q \left[F(\omega \otimes_t W^{(t)}) - \int_t^1 g(s, q^Q(s, \omega \otimes_t W^{(t)})) ds \mid \mathcal{F}_t \right](\omega) \\ &= \mathbb{E}^{Q_\omega} \left[F(\omega \otimes_t W) - \int_t^1 g(s, q^Q(s, \omega \otimes_t W)) ds \right] \\ &= \mathbb{E}^{Q_\omega} \left[F(\omega \otimes_t W) - (1-t) \int_0^1 g(t + s(1-t), q^Q(t + s(1-t), \omega \otimes_t W)) ds \right] \\ &= \mathbb{E}^{Q_\omega} \left[F(\omega \otimes_t W) - \int_0^1 g^{(t)}(s, q^{Q_\omega}(s, W)) ds \right], \end{aligned}$$

with the last line simply using the definition of $g^{(t)}$. This completes the proof of (51).

Finally, we prove the reverse, namely that

$$(52) \quad Y(t, \omega) \geq \rho^{g^{(t)}}(F(\omega \otimes_t \cdot)).$$

First, note that the definition of the operation \otimes_t entails that, for each Q , the function of ω on the right-hand side of (50) is \mathcal{F}_t -measurable. Using [3], Proposition 7.50, we may find an \mathcal{F}_t -measurable map $\mathcal{C} \ni \omega \mapsto Q_\omega \in \mathcal{Q}$ such that

$$(53) \quad \mathbb{E}^{Q_\omega} \left[F(\omega \otimes_t W) - \int_0^1 g^{(t)}(s, q^{Q_\omega}(s, W)) ds \right] \geq \rho^{g^{(t)}}(F(\omega \otimes_t \cdot)) - \epsilon,$$

for each $\omega \in \mathcal{C}$. Define Q by setting

$$\frac{dQ}{dP}(\omega) = \frac{dQ_\omega}{dP}(\omega^{(t)}).$$

The \mathcal{F}_t -measurability of $\omega \mapsto Q_\omega$ and the independence of $W^{(t)}$ and \mathcal{F}_t under P together ensure that $P(d\omega)$ indeed integrates the right-hand side to 1, so that $Q \in \mathcal{P}(\mathcal{C})$ is well defined. Using the same facts, it is straightforward to check that $Q \in \mathcal{Q}_t$; indeed, if $S \in \mathcal{F}_t$ then

$$\begin{aligned} Q(S) &= \mathbb{E} \left[\frac{dQ}{dP} 1_S(W) \right] = \mathbb{E} \left[\frac{dQ_W}{dP}(W^{(t)}) 1_S(W) \right] \\ &= \int_{\mathcal{C}} \mathbb{E} \left[\frac{dQ_\omega}{dP}(W^{(t)}) \right] 1_S(\omega) P(d\omega) = \int_{\mathcal{C}} 1_S(\omega) P(d\omega) = P(S). \end{aligned}$$

As argued in the previous paragraph, $\omega \mapsto Q_\omega$ is a version of $Q(W^{(t)} \in \cdot \mid \mathcal{F}_t)(\omega)$, and we have

$$q^{Q_\omega}(s, \bar{\omega}) = \sqrt{1-t} q^Q(t+s(1-t), \omega \otimes_t \bar{\omega}),$$

for $\omega, \bar{\omega} \in \mathcal{C}$. Using (53), the definition of $g^{(t)}$, and a change of variables, we find

$$\begin{aligned} &\rho^{g^{(t)}}(F(\omega \otimes_t \cdot)) \\ &\leq \epsilon + \mathbb{E}^Q \left[F(\omega \otimes_t W^{(t)}) - \int_0^1 g^{(t)}(s, q^{Q_\omega}(s, W^{(t)})) ds \mid \mathcal{F}_t \right](\omega) \\ &= \epsilon + \mathbb{E}^Q \left[F(\omega \otimes_t W^{(t)}) \right. \\ &\quad \left. - (1-t) \int_0^1 g(t+s(1-t), q^Q(t+s(1-t), \omega \otimes_t W^{(t)})) ds \mid \mathcal{F}_t \right](\omega) \\ &= \epsilon + \mathbb{E}^Q \left[F(\omega \otimes_t W^{(t)}) - \int_t^1 g(s, q^Q(s, \omega \otimes_t W^{(t)})) ds \mid \mathcal{F}_t \right](\omega). \end{aligned}$$

Comparing this to the expression (49), the proof of (52) is complete. \square

We now give the proof of Theorem 2.7. In the following, define $\mathcal{C}_0[t, 1]$ to be the set of continuous paths $\omega : [t, 1] \rightarrow \mathbb{R}$ with $\omega(t) = 0$. For $\omega \in \mathcal{C}$ and $\bar{\omega} \in \mathcal{C}_0[t, 1]$, define $\omega \oplus_t \bar{\omega} \in \mathcal{C}$ by

$$\omega \oplus_t \bar{\omega}(s) = \omega(s \wedge t) + \bar{\omega}(s) 1_{[t, 1]}(s).$$

Recall the notation $h_n(t, q) = h(t, q/\sqrt{n})$.

PROOF OF THEOREM 2.7. The case $t = 1$ is trivial. Indeed, then $u_n(1, \omega) = Y_n(1, \sqrt{n}\omega) = F(\omega)$ for each n , which is seen to equal $u(1, \omega) = F(\omega)$. Assume henceforth that $t \in [0, 1)$. Note first that $(g^{(t)})_n = (g_n)^{(t)} =: g_n^{(t)}$. We let

$$u_n(t, \omega) := \rho^{g_n^{(t)}} \left(F \left(\omega \otimes_t \frac{W}{\sqrt{n}} \right) \right).$$

Using Lemma 5.1, we also have almost surely $Y_n(t, \omega) = u_n(t, \omega/\sqrt{n})$. Since $F(\omega \otimes_t \cdot)$ is bounded and continuous, we may apply Theorem 2.2 to get

$$\begin{aligned} \lim_{n \rightarrow \infty} u_n(t, \omega) &= \sup_{\bar{\omega} \in \mathcal{C}_0} \left(F(\omega \otimes_t \bar{\omega}) - \int_0^1 g^{(t)}(s, \dot{\bar{\omega}}(s)) ds \right) \\ &= \sup_{\bar{\omega} \in \mathcal{C}_0} \left(F(\omega \otimes_t \bar{\omega}) - (1-t) \int_0^1 g\left(t+s(1-t), \frac{\dot{\bar{\omega}}(s)}{\sqrt{1-t}}\right) ds \right) \\ &= \sup_{\bar{\omega} \in \mathcal{C}_0} \left(F(\omega \otimes_t \bar{\omega}) - \int_t^1 g\left(s, \frac{1}{\sqrt{1-t}} \dot{\bar{\omega}}\left(\frac{s-t}{1-t}\right)\right) ds \right). \end{aligned}$$

Given $\bar{\omega} \in \mathcal{C}_0 = \mathcal{C}_0[0, 1]$, we may define $\tilde{\omega} \in \mathcal{C}_0[t, 1]$ by $\tilde{\omega}(s) := \sqrt{1-t} \bar{\omega}\left(\frac{s-t}{1-t}\right)$. Then $\omega \otimes_t \bar{\omega} = \omega \oplus_t \tilde{\omega}$, and the map $\bar{\omega} \mapsto \tilde{\omega}$ defines a bijection from \mathcal{C}_0 to $\mathcal{C}_0[t, 1]$. Hence, the above reduces to $u(t, \omega)$.

To prove the final claim, let us first assume that F is uniformly continuous. Using the fact that a convex risk measure is always 1-Lipschitz with respect to the supremum norm (e.g., [19], Lemma 4.3) we get

$$\begin{aligned} |Y_n(t, \omega) - Y_n(t, 0)| &= |u_n(t, \omega/\sqrt{n}) - u_n(t, 0)| \\ &\leq \left\| F\left(\frac{1}{\sqrt{n}}(\omega \otimes_t \cdot)\right) - F\left(\frac{1}{\sqrt{n}}(0 \otimes_t \cdot)\right) \right\|_{\infty}, \end{aligned}$$

which converges to zero by uniform continuity. This and the convergence for u_n settles the uniformly continuous case.

Now, if F is merely continuous, it is nevertheless the pointwise increasing limit of a sequence of bounded uniformly continuous (even Lipschitz) functions. Observing that both $Y_n(t)$ and $u(t, 0)$ are increasing functions of F , we easily conclude from the uniformly continuous case that

$$\liminf_{n \rightarrow \infty} Y_n(t) \geq u(t, 0) \quad \text{a.s.}$$

On the other hand, there is a uniformly bounded sequence (F_m) of uniformly continuous functions decreasing to F . This time we can conclude that

$$\limsup_{n \rightarrow \infty} Y_n(t) \leq \inf_m \sup_{\bar{\omega} \in \mathcal{C}_0[t, 1]} \left(F_m(0 \oplus_t \bar{\omega}) - \int_t^1 g(s, \dot{\bar{\omega}}(s)) ds \right) \quad \text{a.s.}$$

It remains to bound the right-hand side from above by $u(t, 0)$. For each $m \in \mathbb{N}$ find $\omega_m \in \mathcal{C}_0[t, 1]$ such that

$$\sup_{\bar{\omega} \in \mathcal{C}_0[t, 1]} \left(F_m(0 \oplus_t \bar{\omega}) - \int_t^1 g(s, \dot{\bar{\omega}}(s)) ds \right) \leq \frac{1}{m} + F_m(0 \oplus_t \omega_m) - \int_t^1 g(s, \dot{\omega}_m(s)) ds.$$

Since (F_m) is uniformly bounded, we deduce (as we did for (41) in the proof of Theorem 2.2)

$$\sup_{m \in \mathbb{N}} \int_t^1 g(s, \dot{\omega}_m(s)) ds < \infty.$$

It is a consequence of Lemma A.1 that there exists $\omega \in \mathcal{C}_0[t, 1]$ absolutely continuous and such that for a subsequence (which we do not track) $\omega_m \rightarrow \omega$ uniformly, and $\liminf_m \int_t^1 g(s, \dot{\omega}_m(s)) ds \geq \int_t^1 g(s, \dot{\omega}(s)) ds$. On the other hand, since F_m decreases pointwise to F , we have $F_m(0 \oplus_t \omega_m) \rightarrow F(0 \oplus_t \omega)$ by Dini's theorem. We conclude that

$$\inf_m \sup_{\bar{\omega} \in \mathcal{C}_0[t, 1]} \left(F_m(0 \oplus_t \bar{\omega}) - \int_t^1 g(s, \dot{\bar{\omega}}(s)) ds \right) \leq F(0 \oplus_t \omega) - \int_t^1 g(s, \dot{\omega}(s)) ds \leq u(t, 0),$$

which completes the proof. \square

PROOF OF THEOREM 2.6. The case $t = 1$ is trivial: Because $(W_{(n,k)})_{k=1}^n$ are independent Wiener processes under P , we conclude from the law of large numbers that $Y_n(1) = F(\frac{1}{n} \sum_{k=1}^n \delta_{W_{(n,k)}})$ converges a.s. to $F(P)$.

Henceforth, assume $t < 1$, so that $\lfloor nt \rfloor < n$ for all $n \in \mathbb{N}$. First notice that

$$\begin{aligned} G_n^{(t)}(s, q) &:= (G_n)^{(t)}(s, q) = (1 - t)G_n\left(t + s(1 - t), \frac{q}{\sqrt{1 - t}}\right) \\ &= (1 - t)g\left(nt + ns(1 - t) - \lfloor nt + ns(1 - t) \rfloor, \frac{q}{\sqrt{n(1 - t)}}\right). \end{aligned}$$

Plugging in $t_n := \lfloor nt \rfloor/n$, we find

$$\begin{aligned} (54) \quad G_n^{(t_n)}(s, q) &= (1 - t_n)g\left(\lfloor nt \rfloor + s(n - \lfloor nt \rfloor) - \lfloor \lfloor nt \rfloor + s(n - \lfloor nt \rfloor) \rfloor, \frac{q}{\sqrt{n - \lfloor nt \rfloor}}\right) \\ &= (1 - t_n)g\left(s(n - \lfloor nt \rfloor) - \lfloor s(n - \lfloor nt \rfloor) \rfloor, \frac{q}{\sqrt{n - \lfloor nt \rfloor}}\right) \\ &= (1 - t_n)G_{n - \lfloor nt \rfloor}(s, q), \end{aligned}$$

where the second line used the identity $\lfloor k + c \rfloor = k + \lfloor c \rfloor$, valid for any integer k and any $c \in \mathbb{R}$. Define $L_n : \mathcal{C} \rightarrow \mathcal{P}(C)$ by

$$L_n(\omega) := \frac{1}{n} \sum_{k=1}^n \delta_{\omega_{(n,k)}}.$$

Using Lemma 5.1, we write

$$Y_n(t_n, \omega) = \rho^{G_n^{(t_n)}}(F \circ L_n(\omega \otimes_{t_n} W)).$$

Note that $(\omega \otimes_{t_n} W)_{(n,k)} \equiv \omega_{(n,k)}$ if $k \leq nt_n = \lfloor nt \rfloor$, while for $k \geq \lfloor nt \rfloor + 1$ and $s \in [0, 1]$ we have

$$\begin{aligned} (\omega \otimes_{t_n} W)_{(n,k)}(s) &= \sqrt{n(1 - t_n)}\left(W\left(\frac{\frac{k-1+s}{n} - t_n}{1 - t_n}\right) - W\left(\frac{\frac{k-1}{n} - t_n}{1 - t_n}\right)\right) \\ &= \sqrt{n - \lfloor nt \rfloor}\left(W\left(\frac{k - 1 + s - \lfloor nt \rfloor}{n - \lfloor nt \rfloor}\right) - W\left(\frac{k - 1 - \lfloor nt \rfloor}{n - \lfloor nt \rfloor}\right)\right) \\ &= W_{(n - \lfloor nt \rfloor, k - \lfloor nt \rfloor)}(s). \end{aligned}$$

Hence,

$$\begin{aligned} (55) \quad L_n(\omega \otimes_{t_n} W) &= \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \delta_{\omega_{(n,k)}} + \frac{1}{n} \sum_{k=\lfloor nt \rfloor + 1}^n \delta_{W_{(n - \lfloor nt \rfloor, k - \lfloor nt \rfloor)}} \\ &= t_n \frac{1}{\lfloor nt \rfloor} \sum_{k=1}^{\lfloor nt \rfloor} \delta_{\omega_{(n,k)}} + (1 - t_n)L_{n - \lfloor nt \rfloor}(W). \end{aligned}$$

Assume first that F is uniformly continuous. Under P , $\omega_{(n,k)}$ for $k = 1, \dots, n$ are independent Brownian motions, and so as $n \rightarrow \infty$ the first term converges P -a.s. by the law of large numbers to tP . Hence, it holds for P -a.e. ω that the existence of the limit

$$\lim_{n \rightarrow \infty} Y_n(t_n, \omega) = \lim_{n \rightarrow \infty} \rho^{G_n^{(t_n)}}(F \circ L_n(\omega \otimes_{t_n} W)),$$

is equivalent to the existence of the limit

$$\lim_{n \rightarrow \infty} \rho^{G_n^{(t_n)}}(F(tP + (1 - t)L_{n - \lfloor nt \rfloor}(W))),$$

and if any of these exist, then they are equal. Indeed, from the 1-Lipschitz continuity of convex risk measures [19], Lemma 4.3, we have

$$\begin{aligned} & \left| \rho^{G_n^{(t_n)}} (F \circ L_n(\omega \otimes_{t_n} W)) - \rho^{G_n^{(t)}} (F(tP + (1-t)L_{n-\lfloor nt \rfloor}(W)) \right) \\ & \leq \| F \circ L_n(\omega \otimes_{t_n} \cdot) - F(tP + (1-t)L_{n-\lfloor nt \rfloor}(\cdot)) \|_\infty, \end{aligned}$$

with the right-hand side converging to zero thanks to the uniform continuity and boundedness of F , the law of large numbers, and the identity (55). Using this, equation (54), and Theorem 2.1 we compute the limit,

$$\begin{aligned} \lim_{n \rightarrow \infty} Y_n(t_n, \omega) &= \lim_{n \rightarrow \infty} \rho^{G_n^{(t_n)}} (F(tP + (1-t)L_{n-\lfloor nt \rfloor}(W)) \\ &= \lim_{n \rightarrow \infty} \rho^{(1-t_n)G_{n-\lfloor nt \rfloor}} (F(tP + (1-t)L_{n-\lfloor nt \rfloor}(W)) \\ &= \lim_{n \rightarrow \infty} \rho^{(1-t)G_{n-\lfloor nt \rfloor}} (F(tP + (1-t)L_{n-\lfloor nt \rfloor}(W)) \\ &= \sup_{q \in \mathcal{L}_b} \left(F(tP + (1-t)Q^q) - (1-t)\mathbb{E} \left[\int_0^1 g(s, q(s)) ds \right] \right). \end{aligned}$$

The third equality, in which $(1 - t_n)$ is replaced by $(1 - t)$ in the superscript, follows from the estimate

$$(56) \quad \left| \rho^{ag}(f) - \rho^{bg}(f) \right| \leq \left(\frac{3\|f\|_\infty}{a} + \sup g^- + g(0) \right) |b - a|,$$

valid for any g satisfying (TI), any bounded measurable f , and any $a, b \in (0, 1]$, which we justify in the next paragraph. (Here $\sup g^- := \sup_{(t,q)} \max\{0, -g(t, q)\}$.)

To prove (56) note that by monotonicity of ρ^g , it holds $\rho^g(f) \leq \rho^g(\|f\|_\infty) = \|f\|_\infty + \rho^g(0) \leq \|f\|_\infty + \sup g^-$ and $\rho^g(f) \geq \mathbb{E}[f] - g(0) \geq -\|f\|_\infty - g(0)$. Take note also of the easy identity $\rho^{cg}(f) = c\rho^g(f/c)$, valid for $c > 0$. Thus,

$$\begin{aligned} \left| \rho^{ag}(f) - \rho^{bg}(f) \right| &\leq \left| a\rho^g\left(\frac{f}{a}\right) - b\rho^g\left(\frac{f}{a}\right) \right| + \left| b\rho^g\left(\frac{f}{a}\right) - b\rho^g\left(\frac{f}{b}\right) \right| \\ &\leq \left| \rho^g\left(\frac{f}{a}\right) \right| |a - b| + b \left\| \frac{f}{a} - \frac{f}{b} \right\|_\infty \\ &\leq \left(3\frac{\|f\|_\infty}{a} + \sup g^- + g(0) \right) |a - b|. \end{aligned}$$

We have now completed the proof under the extra assumption that F is uniformly continuous. To conclude, we may drop this extra assumption by essentially the same monotone approximation arguments as in the proof of Theorem 2.7, by relying again on Lemma A.1. □

6. On the PDE connection. The goal of this section is to briefly elaborate on the PDE results of Section 2.4. The basic lemma linking the functionals ρ^g with PDEs is the following:

LEMMA 6.1. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be bounded and lower semicontinuous. Then the parabolic PDE*

$$(57) \quad \begin{cases} \partial_t v(t, x) + \frac{1}{2} \Delta v(t, x) + g(t, \nabla v(t, x)) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ v(1, x) = f(x) & \text{for } x \in \mathbb{R}^d \end{cases}$$

admits a minimal viscosity supersolution v . Moreover, $\rho^g(f(W(1))) = v(0, 0)$.

If $f \in C_b(\mathbb{R}^d)$ and $g^*(t, \cdot)$ is differentiable and there is a constant $C \geq 0$ such that

$$|g^*(t, z)| \leq C(1 + |z|^2) \quad \text{and} \quad |\partial_z g^*(t, z)| \leq C(1 + |z|), \quad z \in \mathbb{R}^d,$$

then v is the unique viscosity solution of (57).

PROOF. The existence of a minimal viscosity supersolution v is shown in [12], Theorem 5.2, where it is also shown that $v(0, 0) = Y(0)$, where (Y, Z) is the minimal supersolution of the BSDE (10). To complete the proof, simply recall from (12) that $Y(0) = \rho^g(f(W(1)))$. When g^* is of quadratic growth and $f \in C_b(\mathbb{R}^d)$, the existence of a unique viscosity solution u follows by [21], Theorems 3.2 and 3.8. By comparison, $v = u$. \square

Now, for each integer $n \geq 1$, consider the operator \mathbb{L}_n , taking bounded lower semicontinuous functions on $(\mathbb{R}^d)^n$ to bounded lower semicontinuous functions on $(\mathbb{R}^d)^{n-1}$, as follows. Given $F : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ and $(x_1, \dots, x_{n-1}) \in (\mathbb{R}^d)^{n-1}$, we define $\mathbb{L}_n F(x_1, \dots, x_{n-1}) := v(0, 0)$, where $v = v(t, x)$ is the minimal viscosity supersolution of the PDE

$$(58) \quad \begin{cases} \partial_t v(t, x) + \frac{1}{2} \Delta v(t, x) + g(t, \nabla v(t, x)) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ v(1, x) = F(x_1, \dots, x_{n-1}, x) & \text{for } x \in \mathbb{R}^d. \end{cases}$$

By Lemma 6.1, the minimal viscosity supersolution v exists, and we have

$$\rho^g(F(x_1, \dots, x_{n-1}, W(1))) = v(0, 0).$$

By definition,

$$\begin{aligned} &\rho^g(F(x_1, \dots, x_{n-1}, W(1))) \\ &= \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q \left[F\left(x_1, \dots, x_{n-1}, W(1) + \int_0^1 q^Q(t) dt\right) - \int_0^1 g(t, q^Q(t)) dt \right]. \end{aligned}$$

Because F is lower semicontinuous and bounded, this exhibits $\rho^g(F(x_1, \dots, x_{n-1}, W(1)))$ as the supremum of lower semicontinuous functions of (x_1, \dots, x_{n-1}) . Hence, \mathbb{L}_n is well defined and indeed maps bounded lower semicontinuous functions of $(\mathbb{R}^d)^n$ to bounded lower semicontinuous functions of $(\mathbb{R}^d)^{n-1}$. For $n = 1$, we interpret \mathbb{L}_1 as mapping from bounded lower semicontinuous functions of \mathbb{R}^d to real numbers. The composition $\mathbb{L}_1 \cdots \mathbb{L}_{n-1} \mathbb{L}_n$ then maps a function on $(\mathbb{R}^d)^n$ to a real number.

PROPOSITION 6.2. For a function $F \in C_b(\mathcal{P}(\mathbb{R}^d))$, define $F^n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ by

$$F^n(x_1, \dots, x_n) := nF\left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}\right).$$

Then, defining $Q_1^q := P \circ (W(1) + \int_0^1 q(t) dt)^{-1}$ for $q \in \mathcal{L}_b$ as the time-1 marginal of Q^q ,

$$(59) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{L}_1 \cdots \mathbb{L}_{n-1} \mathbb{L}_n F^n = \sup_{q \in \mathcal{L}_b} \left(F(Q_1^q) - \mathbb{E} \left[\int_0^1 g(t, q(t)) dt \right] \right).$$

PROOF. Recall the definition of ρ_n^g from Section 4.1. For a bounded lower semicontinuous function f on $(\mathbb{R}^d)^n$, define $\tilde{f} \in B_b(C^n)$ by setting $\tilde{f}(\omega_1, \dots, \omega_n) = f(\omega_1(1), \dots, \omega_n(1))$, and note that we have

$$\mathbb{L}_1 \cdots \mathbb{L}_{n-1} \mathbb{L}_n f = \rho_n^g(\tilde{f}) = n\rho^{G_n} \left(\frac{1}{n} \tilde{f}(W_{(n,1)}, \dots, W_{(n,n)}) \right).$$

Indeed, the first equality is just the definition of ρ_n^g , while the second is Proposition 4.1. In particular, we may write

$$\tilde{F}^n(\omega_1, \dots, \omega_n) = F^n(\omega_1(1), \dots, \omega_n(1)) = nF\left(\frac{1}{n} \sum_{i=1}^n \delta_{\omega_i(1)}\right),$$

and thus

$$\frac{1}{n} \mathbb{L}_1 \cdots \mathbb{L}_{n-1} \mathbb{L}_n F^n = \rho^{G_n}(\tilde{F}^n(W_{(n,1)}, \dots, W_{(n,n)})) = \rho^{G_n}\left(F\left(\frac{1}{n} \sum_{k=1}^n \delta_{W_{(n,k)}(1)}\right)\right).$$

Conclude from Theorem 2.1. \square

REMARK 6.3. The right-hand side of (59) can be further rewritten as $\sup_{\nu \in \mathcal{P}(\mathbb{R}^d)} \{F(\nu) - I(\nu)\}$, where $I(\nu) := \inf\{\mathbb{E}^Q[\int_0^1 g(t, q^Q(t)) dt] : Q \in \mathcal{P}_1^*(\mathcal{C}), Q \circ \omega(1)^{-1} = \nu\}$ is a Schrödinger-type problem under the classical observable (as discussed in Section 2.3).

We finally turn our attention to the formalization of the heuristics given in Section 2.4.1. The novelty here lies in the “stochastic” proof, involving our BSDE limit theorems which allow to bypass the regularity conditions often made on the coefficients of the PDE.

PROPOSITION 6.4. *Let $f \in C_b(\mathbb{R}^d)$. The PDE*

$$(60) \quad \begin{cases} \partial_t u_n(t, x) + \frac{1}{2n} \Delta u_n(t, x) + g^*(t, \nabla u_n(t, x)) = 0 & \text{on } [0, 1] \times \mathbb{R}^d, \\ u_n(1, x) = f(x) & \text{for } x \in \mathbb{R}^d \end{cases}$$

admits a minimal viscosity supersolution u_n . Moreover, $u_n \rightarrow u$ pointwise, where u is the function given by

$$u(t, x) = \sup_{\omega \in \mathcal{C}_0[t, 1]} \left(f(x + \omega(1)) - \int_t^1 g(s, \dot{\omega}(s)) ds \right).$$

When $g(t, q) = g(q)$ does not depend on t , then the function u reduces to the Hopf–Lax–Oleinik formula (25) and if in addition g is real-valued and f Lipschitz continuous, then u is the unique viscosity solution of (24).

PROOF. Let $(t, x) \in [0, 1] \times \mathbb{R}^d$ be fixed and put $X_n^{t,x}(s) := x + \frac{1}{\sqrt{n}}(W(s) - W(t))$, $s \geq t$. By [12] the function $u_n(t, x) := Y_n(t)$ is the minimal supersolution of the PDE (60), where (Y_n, Z_n) is the minimal supersolution of the BSDE with generator g_n^* and terminal condition $f(X_n^{t,x}(1))$. Let $F : \mathcal{C} \rightarrow \mathbb{R}$ be given by $F(\omega) = f(x + \omega(1) - \omega(t))$. By Theorem 2.7 and the fact that $Y_n(t)$ is deterministic, it holds

$$\begin{aligned} u_n(t, x) = Y_n(t) &\rightarrow \sup_{\omega \in \mathcal{C}_0[t, 1]} \left(F(0 \oplus_t \omega) - \int_t^1 g(s, \dot{\omega}(s)) ds \right) \\ &= \sup_{\omega \in \mathcal{C}_0[t, 1]} \left(f(x + \omega(1)) - \int_t^1 g(s, \dot{\omega}(s)) ds \right) = u(t, x). \end{aligned}$$

Now, when g is time-independent, the Hopf–Lax–Oleinik formula (25) follows from Jensen’s inequality. Granting the additional assumptions on g and f , it is classical that the Hopf–Lax–Oleinik formula is the unique viscosity solution of (24); see [16], Theorem 10.3. \square

7. Some extensions of the limit theorems. In this section we describe two extensions of the main theorems. First, we show how to strengthen the topology used in Theorems 2.1 and 2.6 to the 1-Wasserstein topology, which allows us to derive a Cramér-type theorem. Second, we incorporate a random initial position for $W(0)$, which has thus far been assumed to be zero.

7.1. *Extension to stronger topologies.* Recall that $\mathcal{P}(\mathcal{C})$ denotes the set of Borel probability measures on \mathcal{C} . Define \mathcal{W}_1 to be the 1-Wasserstein metric on the space

$$\mathcal{P}_1(\mathcal{C}) := \left\{ Q \in \mathcal{P}(\mathcal{C}) : \int_{\mathcal{C}} \|\omega\|_{\infty} Q(d\omega) < \infty \right\},$$

where we recall $\|\cdot\|_{\infty}$ denotes the supremum norm on \mathcal{C} . That is, $\mathcal{W}_1(Q, Q')$ is the infimum over all $\bar{Q} \in \mathcal{P}(\mathcal{C} \times \mathcal{C})$ with marginals Q and Q' of the quantity $\int \|\omega - \omega'\|_{\infty} \bar{Q}(d\omega, d\omega')$. Recall that $\tilde{\alpha}^g$ was defined in (46), and as usual we tacitly assume g satisfies (TI).

LEMMA 7.1. *The sub-level sets of $\tilde{\alpha}^g$ are \mathcal{W}_1 -compact. More precisely, for every $a \in \mathbb{R}$ the set $\Lambda_a := \{Q \in \mathcal{P}(\mathcal{C}) : \tilde{\alpha}^g(Q) \leq a\}$ is contained in $\mathcal{P}_1(\mathcal{C})$ and is compact in the \mathcal{W}_1 -topology.*

PROOF. Noting that g is bounded from below and $\tilde{\alpha}^{g+c} = \tilde{\alpha}^g + c$ for constants $c \in \mathbb{R}$, we may assume without loss of generality that $g \geq 0$. Fix $a \in \mathbb{R}$. We know from Lemma 4.3 that Λ_a is compact in the topology of weak convergence. It suffices to show (see [43], Theorem 7.12) that

$$(61) \quad \lim_{r \rightarrow \infty} \sup_{Q \in \Lambda_a} \mathbb{E}^Q[\|W\|_{\infty} 1_{\{\|W\|_{\infty} \geq 2r\}}] = 0.$$

By Assumption (TI), for each $c > 0$ we may find $N > 0$ such that $g(t, q) \geq c|q|$ whenever $|q| \geq N$. Clearly $\Lambda_a \subset \mathcal{P}^*$. For $Q \in \Lambda_a$, by definition, $W^Q(t) := W(t) - \int_0^t q^Q(s) ds$ is a Q -Brownian motion. Hence, for any $r > 1$,

$$(62) \quad \mathbb{E}^Q[\|W\|_{\infty} 1_{\{\|W\|_{\infty} \geq 2r\}}] \leq \mathbb{E}^Q[\|W\|_{\infty} 1_{\{\|W^Q\|_{\infty} \geq r\}}] + \mathbb{E}^Q[\|W\|_{\infty} 1_{\{\int_0^1 |q^Q(t)| dt \geq r\}}].$$

For the first term, we make the estimate

$$(63) \quad \begin{aligned} & \mathbb{E}^Q[\|W\|_{\infty} 1_{\{\|W^Q\|_{\infty} \geq r\}}] \\ & \leq \mathbb{E}^Q[\|W^Q\|_{\infty} 1_{\{\|W^Q\|_{\infty} \geq r\}}] + \mathbb{E}^Q\left[\int_0^1 |q^Q(t)| dt 1_{\{\|W^Q\|_{\infty} \geq r\}}\right] \\ & \leq \mathbb{E}^P[\|W\|_{\infty} 1_{\{\|W\|_{\infty} \geq r\}}] + N \mathbb{E}^Q[1_{\{\|W^Q\|_{\infty} \geq r\}}] \\ & \quad + \frac{1}{c} \mathbb{E}^Q\left[\int_0^1 g(t, q^Q(t)) dt 1_{\{\|W^Q\|_{\infty} \geq r\}}\right] \\ & \leq (1 + N) \mathbb{E}^P[\|W\|_{\infty} 1_{\{\|W\|_{\infty} \geq r\}}] + \frac{1}{c} \mathbb{E}^Q\left[\int_0^1 g(t, q^Q(t)) dt\right]. \end{aligned}$$

We bound the second term of (62) similarly:

$$(64) \quad \begin{aligned} & \mathbb{E}^Q[\|W\|_{\infty} 1_{\{\int_0^1 |q^Q(t)| dt \geq r\}}] \\ & \leq \mathbb{E}^Q[\|W^Q\|_{\infty} 1_{\{\int_0^1 |q^Q(t)| dt \geq r\}}] + \mathbb{E}^Q\left[\int_0^1 |q^Q(t)| dt 1_{\{\int_0^1 |q^Q(t)| dt \geq r\}}\right] \\ & \leq \mathbb{E}^P[\|W\|_{\infty}^2]^{1/2} Q\left(\int_0^1 |q^Q(t)| dt \geq r\right)^{1/2} \\ & \quad + \frac{1}{c} \mathbb{E}^Q\left[\int_0^1 g(t, q^Q(t)) dt\right] + N Q\left(\int_0^1 |q^Q(t)| dt \geq r\right). \end{aligned}$$

Lastly, note that the definition of Λ_a and Assumption (TI) ensure that

$$(65) \quad \lim_{r \rightarrow \infty} \sup_{Q \in \Lambda_a} Q \left(\int_0^1 |q^Q(t)| dt \geq r \right) = 0.$$

Combining this with (63)–(64) and returning to (62), we deduce (61) since c was arbitrary. \square

COROLLARY 7.2. *The conclusions of Theorems 2.1 and 2.6 hold for any $F \in C_b(\mathcal{P}_1(\mathcal{C}))$, where $\mathcal{P}_1(\mathcal{C})$ is equipped with the metric \mathcal{W}_1 , with the suprema over $Q \in \mathcal{Q}$ replaced by $Q \in \mathcal{Q} \cap \mathcal{P}_1(\mathcal{C})$.*

PROOF. The proofs are exactly the same as those of Theorems 2.1 and Theorem 2.6, with only minor points to check. In light of Lemma 7.1, we may apply the more general [23], Theorem 3.1, in place of [23], Theorem 1.1, to conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \tilde{\rho}_n^g(nF \circ L_n) = \sup_{Q \in \mathcal{P}_1(\mathcal{C})} (F(Q) - \tilde{\alpha}^g(Q)).$$

The only point worth checking is that

$$\sup_{Q \in \mathcal{P}_1(\mathcal{C})} (F(Q) - \tilde{\alpha}^g(Q)) = \sup_{Q \in \mathcal{Q} \cap \mathcal{P}_1(\mathcal{C})} (F(Q) - \alpha^g(Q))$$

holds when F is merely \mathcal{W}_1 -continuous, but the same argument as in the proof of Theorem 2.1 works: If $\tilde{\alpha}^g(Q) < \infty$, then there exists $Q_n \in \mathcal{Q}$ such that $Q_n \rightarrow Q$ weakly and $\limsup_n \alpha^g(Q_n) \leq \tilde{\alpha}^g(Q)$. Deduce from Lemma 7.1 that $\{Q_n\}$ is \mathcal{W}_1 -precompact and thus $\mathcal{W}_1(Q_n, Q) \rightarrow 0$. Hence, $F(Q_n) \rightarrow F(Q)$, and the above identity follows. \square

As a consequence of Corollary 7.2, we provide the following Crámer-type limit theorem:

COROLLARY 7.3. *For every $F \in C_b(\mathcal{C})$, we have*

$$(66) \quad \lim_{n \rightarrow \infty} \rho^{G_n} \left(F \left(\frac{1}{n} \sum_{k=1}^n W_{(n,k)} \right) \right) = \sup_{\omega \in \mathcal{C}_0} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right).$$

PROOF. Apply Corollary 7.2 to the \mathcal{W}_1 -continuous function $\mathcal{P}_1(\mathcal{C}) \ni Q \mapsto F(\int_{\mathcal{C}} \omega \times Q(d\omega))$, where the integral is understood in the Bochner sense, to get

$$\lim_{n \rightarrow \infty} \rho^{G_n} \left(F \left(\frac{1}{n} \sum_{k=1}^n W_{(n,k)} \right) \right) = \sup_{Q \in \mathcal{P}_1(\mathcal{C})} \left(F \left(\int_{\mathcal{C}} \omega Q(d\omega) \right) - \alpha^g(Q) \right).$$

By the arguments in the proof of Corollary 7.2, the above expression is equal to

$$\sup_{Q \in \mathcal{P}^*} \left(F \left(\int_{\mathcal{C}} \bar{\omega} Q(d\bar{\omega}) \right) - \tilde{\alpha}^g(Q) \right) = \sup_{\omega \in \mathcal{C}_0} (F(\omega) - I(\omega)),$$

where we define

$$I(\omega) := \inf \left\{ \tilde{\alpha}^g(Q) : Q \in \mathcal{P}^* \cap \mathcal{P}_1(\mathcal{C}), \int_{\mathcal{C}} \bar{\omega} Q(d\bar{\omega}) = \omega \right\}.$$

Indeed, we may restrict the supremum to $\mathcal{P}^* \cap \mathcal{P}_1(\mathcal{C})$ as opposed to \mathcal{P}^* because $\tilde{\alpha}^g(Q) = \infty$ for $Q \notin \mathcal{P}_1(\mathcal{C})$ by Lemma 7.1. We need only show that

$$I(\omega) = \begin{cases} \int_0^1 g(t, \dot{\omega}(t)) dt & \text{if } \omega \text{ is absolutely continuous,} \\ \infty & \text{otherwise.} \end{cases}$$

Noting that $\mathbb{E}^Q[W(t)] = \int_0^t \mathbb{E}^Q[q^Q(s)] ds$ for $Q \in \mathcal{P}^* \cap \mathcal{P}_1(\mathcal{C})$, we have

$$I(\omega) = \inf \left\{ \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] : Q \in \mathcal{P}^* \cap \mathcal{P}_1(\mathcal{C}), \int_0^t \mathbb{E}^Q[q^Q(s)] ds = \omega(t), \forall t \in [0, 1] \right\}.$$

Now, fix $\omega \in \mathcal{C}$. Jensen’s inequality yields

$$(67) \quad \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] \geq \int_0^1 g(t, \mathbb{E}^Q[q^Q(t)]) dt = \int_0^1 g(t, \dot{\omega}(t)) dt,$$

for any $Q \in \mathcal{P}^*$ for which $\int_0^t \mathbb{E}^Q[q^Q(s)] ds = \omega(t)$ for all $t \in [0, 1]$. If ω is absolutely continuous, then we can define $Q = P \circ (W + \int_0^\cdot \dot{\omega}(t) dt)^{-1}$ so that $Q \in \mathcal{P}^* \cap \mathcal{P}_1(\mathcal{C})$ with $q^Q(t) = \dot{\omega}(t)$ for all $t \in [0, 1]$. We conclude that, for ω absolutely continuous,

$$I(\omega) = \int_0^1 g(t, \dot{\omega}(t)) dt.$$

On the other hand, if ω is not absolutely continuous, then there cannot exist $Q \in \mathcal{P}^* \cap \mathcal{P}_1(\mathcal{C})$ with $\int_0^t \mathbb{E}^Q[q^Q(s)] ds = \omega(t)$ for all $t \in [0, 1]$. \square

REMARK 7.4. Comparing (8) and (66), we find that

$$\lim_{n \rightarrow \infty} \rho^{g_n} \left(F \left(\frac{W}{\sqrt{n}} \right) \right) = \lim_{n \rightarrow \infty} \rho^{G_n} \left(F \left(\frac{1}{n} \sum_{k=1}^n W_{(n,k)} \right) \right).$$

If $F(\omega) = f(\omega(1))$ depends only on the final value, then these quantities are even equal for each n , without taking a limit (by telescoping sum). This may at first seem unsurprising (at least for time-independent g) because W/\sqrt{n} and $\frac{1}{n} \sum_{k=1}^n W_{(n,k)}$ have the same law for each n . In general, however, we do not expect pre-limit equality except when ρ^g is law-invariant. By [22], the functional ρ^g is law-invariant essentially only when $g(t, q) = c|q|^2$ for $c \in (0, \infty]$, with the convention $0 \cdot \infty := 0$.

REMARK 7.5. Depending on the function g , we could conceivably generalize Lemma 7.1, and thus the rest of the results of this section, to topologies stronger than 1-Wasserstein. The choice of topology should be informed by the growth of g . In particular, suppose the assumption $\lim_{|q| \rightarrow \infty} g(t, q)/|q| = \infty$ in (TI) is strengthened to $\lim_{|q| \rightarrow \infty} g(t, q)/\psi(q) = \infty$, for some nonnegative function ψ on \mathbb{R}^d . Then we should be able to deduce that the sub-level sets of $\tilde{\alpha}^g$ are compact subsets of $\mathcal{P}_\psi(\mathcal{C}) := \{\mu \in \mathcal{P}(\mathcal{C}) : \int \psi d\mu < \infty\}$ in the topology generated by the family of linear functionals $\mu \mapsto \int \varphi d\mu$, where φ ranges over continuous functions on \mathcal{C} satisfying $|\varphi(x)| \leq c(1 + \sup_{t \in [0, 1]} \psi(x_t))$ pointwise. In particular, if $\psi(q) = |q|^p$ for $p \geq 1$, this is nothing but the p -Wasserstein topology. This is notably consistent with the classical case $g(t, q) = |q|^2$, where it is known that Sanov’s theorem holds for the empirical measure of i.i.d. Brownian motions in the p -Wasserstein topology for $p < 2$ but not for $p \geq 2$; this follows from the result of [44] and the fact that $\mathbb{E}[e^{c\|W\|_\infty^p}] < \infty$ for all $c > 0$ if and only if $0 \leq p < 2$. We refrain from pursuing this generalization because the 1-Wasserstein distance is strong enough for the purpose of the Cramér-type result Corollary 7.3.

7.2. *Extensions to nontrivial initial positions.* Preparing for our study of Schrödinger problems, we now extend some of our results to allow the Brownian motion to have a (constant) volatility different than 1 as well as a random, nonzero initial position.

We fix throughout this section the function g satisfying assumption (TI), and we will omit it from our soon-to-be cluttered superscripts. Recall that P denotes Wiener measure on \mathcal{C} and W denotes the canonical process (identity map) on \mathcal{C} . For $Q \in \mathcal{P}(\mathcal{C})$ we take a regular kernel $(Q^{\omega(0)=x})_{x \in \mathbb{R}^d}$ so by disintegration

$$Q(\cdot) = \int_{\mathbb{R}^d} Q^{\omega(0)=x}(\cdot) Q^0(dx),$$

where $Q^{\omega(0)=x} \in \mathcal{P}(\mathcal{C})$ is supported on the set $\mathcal{C}_x := \{\omega \in \mathcal{C} : \omega(0) = x\}$ and Q^0 is the time-zero marginal of Q .

We are given $\mu \in \mathcal{P}(\mathbb{R}^d)$. Recalling that $P_\epsilon = P \circ (\sqrt{\epsilon}W)^{-1}$, we define

$$P_\epsilon^{\omega(0) \sim \mu}(\cdot) := \int_{x \in \mathbb{R}^d} P_\epsilon^{\omega(0)=x}(\cdot) \mu(dx),$$

namely the law of a Brownian motion with starting distribution μ and instantaneous variance (i.e., volatility) equal to ϵ .

For $Q \in \mathcal{P}(\mathcal{C})$ with $Q \ll P_\epsilon^{\omega(0) \sim \mu}$ and $Q^0 = \mu$, we define q_ϵ^Q as the unique progressively measurable process satisfying

$$\frac{dQ}{dP_\epsilon^{\omega(0) \sim \mu}} = \exp\left(\frac{1}{\epsilon} \int_0^1 q_\epsilon^Q(t) dW(t) - \frac{1}{2\epsilon} \int_0^1 |q_\epsilon^Q(t)|^2 dt\right).$$

Then, for $Q \in \mathcal{P}(\mathcal{C})$ we define

$$\alpha_\epsilon^\mu(Q) := \begin{cases} \mathbb{E}^Q \left[\int_0^1 g(t, q_\epsilon^Q(t)) dt \right] & \text{if } Q \ll P_\epsilon^{\omega(0) \sim \mu}, Q^0 = \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

It is straightforward to check that

$$(68) \quad \alpha_\epsilon^\mu(Q) = \begin{cases} \int_{\mathbb{R}^d} \alpha_\epsilon^{\delta_x}(Q^{\omega(0)=x}) \mu(dx) & \text{if } Q \ll P_\epsilon^{\omega(0) \sim \mu}, Q^0 = \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

On the dual side, for $F \in B_b(\mathcal{C})$, we define

$$\begin{aligned} \rho_\epsilon^\mu(F) &:= \sup_{Q \in \mathcal{P}(\mathcal{C})} (\mathbb{E}^Q[F] - \alpha_\epsilon^\mu(Q)) \\ &= \sup_{Q \ll P_\epsilon^{\omega(0) \sim \mu}, Q^0 = \mu} \mathbb{E}^Q \left[F(W) - \int_0^1 g(t, q_\epsilon^Q(t)) dt \right]. \end{aligned}$$

Let us recall the notation for $\mathcal{P}_\epsilon^*(\mathcal{C})$ in Section 2.3, as well as $Q \mapsto q^Q$ defined there. Define

$$\tilde{\alpha}_\epsilon^\mu(Q) := \begin{cases} \mathbb{E}^Q \left[\int_0^1 g(t, q^Q(t)) dt \right] & \text{if } Q \in \mathcal{P}_\epsilon^*(\mathcal{C}) \text{ and } Q^0 = \mu, \\ +\infty & \text{otherwise,} \end{cases}$$

and introduce analogously

$$\tilde{\rho}_\epsilon^\mu(F) := \sup_{Q \in \mathcal{P}(\mathcal{C})} (\mathbb{E}^Q[F] - \tilde{\alpha}_\epsilon^\mu(Q)).$$

We ask the reader to bear in mind that, whenever we use g or any other function as superscript for α or ρ (resp. $\tilde{\alpha}$ or $\tilde{\rho}$), we mean it in the sense of Section 2 (resp. Section 4.2), with the starting distribution being fixed to δ_0 . On the other hand, whenever we use μ or any other measure as superscript for α , ρ , $\tilde{\alpha}$, $\tilde{\rho}$, we mean it in the sense presented in the current section (the function g being fixed).

Let us first present the analogue to Lemmas 4.3 and 4.4 (which took care of $\mu = \delta_0$ and $\epsilon = 1$) in the present setup:

LEMMA 7.6. *The functional $\tilde{\alpha}_\epsilon^\mu$ is convex and lower semicontinuous (with respect to weak convergence), and its sub-level sets are weakly compact in this topology. Furthermore, for $Q \in \mathcal{P}_\epsilon^*(\mathcal{C})$ with $Q^0 = \mu$ we have*

$$\tilde{\alpha}_\epsilon^\mu(Q) = \sup_{F \in B_b(\mathcal{C})} (\mathbb{E}^Q[F] - \tilde{\rho}_\epsilon^\mu(F)) = \sup_{F \in C_b(\mathcal{C})} (\mathbb{E}^Q[F] - \tilde{\rho}_\epsilon^\mu(F)),$$

and, on the other hand, for $F : \mathcal{C} \rightarrow \mathbb{R}$ bounded lower-semicontinuous we have

$$\rho_\epsilon^\mu(F) = \tilde{\rho}_\epsilon^\mu(F).$$

We omit the proof, since it boils down to the same arguments as for Lemmas 4.3 and 4.4. The key point of this section is the following proposition, for which we recall the notation $\mathcal{C}_x = \{\omega \in \mathcal{C} : \omega(0) = x\}$:

PROPOSITION 7.7. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\epsilon > 0$. For $F : \mathcal{C} \rightarrow \mathbb{R}$ measurable and bounded we have*

$$(69) \quad \rho_\epsilon^\mu(F) = \int_{\mathbb{R}^d} \rho_\epsilon^{\delta_x}(F) \mu(dx).$$

For $F \in C_b(\mathcal{C})$ we further have

$$(70) \quad \lim_{\epsilon \downarrow 0} \rho_\epsilon^\mu(F) = \int_{\mathbb{R}^d} \sup_{\omega \in \mathcal{C}_x} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right) \mu(dx).$$

PROOF. Using (68), we have

$$\begin{aligned} (71) \quad \rho_\epsilon^\mu(F) &= \sup_{Q \in \mathcal{P}(\mathcal{C})} (\mathbb{E}^Q[F] - \alpha_\epsilon^\mu(Q)) \\ &= \sup_{Q \ll P_\epsilon^{\omega(0) \sim \mu}, Q^0 = \mu} \left(\mathbb{E}^Q[F] - \int_{\mathbb{R}^d} \alpha_\epsilon^{\delta_x}(Q^{\omega(0)=x}) \mu(dx) \right) \\ &= \sup_{Q \ll P_\epsilon^{\omega(0) \sim \mu}, Q^0 = \mu} \int_{\mathbb{R}^d} (\mathbb{E}^{Q^{\omega(0)=x}}[F] - \alpha_\epsilon^{\delta_x}(Q^{\omega(0)=x})) \mu(dx) \\ &\leq \int_{\mathbb{R}^d} \rho_\epsilon^{\delta_x}(F) \mu(dx). \end{aligned}$$

The proof of the reverse inequality relies on a careful application of a standard measurable selection argument. A straightforward transformation of (BBD) yields

$$(72) \quad \rho_\epsilon^{\delta_x}(F) = \sup_{q \in \mathcal{L}_b} \mathbb{E}^{P^{\omega(0)=0}} \left[F \left(x + \sqrt{\epsilon} W + \int_0^\cdot q(t) dt \right) - \int_0^1 g(t, q(t)) dt \right].$$

Note that \mathcal{L}_b is a Borel subset of the (separable metric) space \mathcal{L}^2 of square-integrable progressively measurable processes, and that the map

$$\mathbb{R}^d \times \mathcal{L}_2 \ni (x, q) \mapsto \mathbb{E}^{P^{\omega(0)=0}} \left[F \left(x + \sqrt{\epsilon} W + \int_0^\cdot q(t) dt \right) - \int_0^1 g(t, q(t)) dt \right]$$

is measurable. We may apply standard analytic set theory [3], Proposition 7.47, to conclude that $x \mapsto \rho_\epsilon^{\delta_x}(F)$ is upper semianalytic and, in particular, universally measurable. The integral in the right-hand side of (69) is thus well defined, since further $\rho_\epsilon^{\delta_x}(F)$ is bounded by the bounds of F and g . By [3], Proposition 7.50, there exists a universally measurable η -approximate optimizer $q^x \in \mathcal{L}_b$ in (72), for any $\eta > 0$. Letting $Q_x = P \circ (x +$

$\sqrt{\epsilon}W + \int_0^\cdot q^x(t) dt)^{-1}$, we check that the probability measure $Q = \int_{x \in \mathbb{R}^d} Q_x \mu(dx)$ satisfies $Q \ll P_\epsilon^{\omega(0) \sim \mu}$, $Q^0 = \mu$, and $Q^{\omega(0)=x} = Q_x$. Moreover, by design,

$$\rho_\epsilon^{\delta_x}(F) - \eta \leq \mathbb{E}^{Q_x}[F] - \alpha_\epsilon^{\delta_x}(Q_x) = \mathbb{E}^{Q^{\omega(0)=x}}[F] - \alpha_\epsilon^{\delta_x}(Q^{\omega(0)=x}).$$

Hence, using the expression (71) for $\rho_\epsilon^\mu(F)$, we deduce $\int_{\mathbb{R}^d} \rho_\epsilon^{\delta_x}(F) \mu(dx) - \eta \leq \rho_\epsilon^\mu(F)$. As $\eta > 0$ was arbitrary, this proves (69).

Now we show (70). The key is to observe from (72) that

$$\rho_\epsilon^{\delta_x}(F) = \rho^{g_\epsilon}(F_\epsilon^x),$$

where $g_\epsilon(q) := g(\sqrt{\epsilon}q)$ and $F_\epsilon^x(\omega) := F(x + \sqrt{\epsilon}\omega)$. Indeed,

$$\begin{aligned} \rho^{g_\epsilon}(F_\epsilon^x) &= \sup_{q \in \mathcal{L}_b} \mathbb{E}^{P_1^{\omega(0)=0}} \left[F \left(x + \sqrt{\epsilon}W + \sqrt{\epsilon} \int_0^\cdot q(s) ds \right) - \int_0^1 g(t, \sqrt{\epsilon}q(t)) dt \right] \\ &= \sup_{q \in \mathcal{L}_b} \mathbb{E}^{P_\epsilon^{\omega(0)=0}} \left[F \left(x + W + \int_0^\cdot q(s) ds \right) - \int_0^1 g(t, q(t)) dt \right] \\ &= \sup_{Q \ll P_\epsilon^{\omega(0)=0}} \mathbb{E}^Q \left[F(x + W) - \int_0^1 g(t, q^Q(t)) dt \right] \\ &= \sup_{Q \ll P_\epsilon^{\omega(0)=x}} \mathbb{E}^Q \left[F(W) - \int_0^1 g(t, q^Q(t)) dt \right] \\ &= \rho_\epsilon^{\delta_x}(F), \end{aligned}$$

where we used (BBD) in the first and third equalities. Thus Theorem 2.2 implies

$$\lim_{\epsilon \downarrow 0} \rho_\epsilon^{\delta_x}(F) = \sup_{\omega \in \mathcal{C}_0} \left(F(x + \omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right) = \sup_{\omega \in \mathcal{C}_x} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right).$$

With this at hand, we conclude by (69) and dominated convergence. \square

8. Application to Schrödinger-type problems. Our aim is to prove the results stated in Section 2.3. We first need some preparatory lemmas. We carry on with the notation of Section 7.2, recalling the convention that $\int_0^1 g(t, \dot{\omega}(t)) dt = \infty$ if ω is not absolutely continuous. We introduce the following very important functional

$$\alpha_0^\mu(Q) := \mathbb{E}^Q \left[\int_0^1 g(t, \dot{W}(t)) dt \right].$$

We also recall that Z is a separable Banach space (of observations) and that $H : \mathcal{C} \rightarrow Z$, the observable, is a continuous linear operator.

The following Γ -convergence type result is a crucial technical step, and part (i) of it relies on our Schilder-type result (Proposition 7.7) in an essential way. Recall that $P_\epsilon = P \circ (\sqrt{\epsilon}W)^{-1}$ denotes the law of a standard Brownian motion times $\sqrt{\epsilon}$.

LEMMA 8.1. *As $\epsilon \downarrow 0$, $\tilde{\alpha}_\epsilon^\mu$ converges to the function α_0^μ in the sense of Γ -convergence. This means that for all $Q \in \mathcal{P}(\mathcal{C})$:*

(i) *Whenever $Q_\epsilon \rightarrow Q$, then*

$$\liminf_{\epsilon \downarrow 0} \tilde{\alpha}_\epsilon^\mu(Q_\epsilon) \geq \alpha_0^\mu(Q).$$

(ii) *There exists some $\tilde{Q}_\epsilon \rightarrow Q$ such that*

$$\limsup_{\epsilon \downarrow 0} \tilde{\alpha}_\epsilon^\mu(\tilde{Q}_\epsilon) \leq \alpha_0^\mu(Q).$$

Moreover, the sequence $\{\tilde{Q}_\epsilon\}$ in (ii) can be explicitly taken as $\tilde{Q}_\epsilon := Q * P_\epsilon$.

PROOF. We first show (ii). We may assume Q is such that $\alpha_0^\mu(Q) < \infty$, and take

$$\tilde{Q}_\epsilon := Q * P_\epsilon := \int_{\mathcal{C}} P \circ (\bar{\omega} + \sqrt{\epsilon}W)^{-1} Q(d\bar{\omega}).$$

To be completely clear, this means

$$\int_{\mathcal{C}} F d\tilde{Q}_\epsilon = \int_{\mathcal{C}} \int_{\mathcal{C}} F(\bar{\omega} + \sqrt{\epsilon}\omega) P(d\omega) Q(d\bar{\omega}).$$

It is readily verified, via Lebesgue dominated convergence, that $\tilde{Q}_\epsilon \rightarrow Q$ weakly. Since $\alpha_0^\mu(Q) < \infty$ it follows that Q is concentrated on absolutely continuous paths, so as a consequence $\tilde{Q}_\epsilon \in \mathcal{P}_\epsilon^*(\mathcal{C})$. Furthermore, $\tilde{Q}_\epsilon^0 = Q^0 = \mu$. As per Lemma 7.6, we know that $\tilde{\alpha}_\epsilon^\mu$ is convex. This implies

$$\tilde{\alpha}_\epsilon^\mu(\tilde{Q}_\epsilon) \leq \int_{\mathcal{C}} \tilde{\alpha}_\epsilon^\mu(P \circ (\bar{\omega} + \sqrt{\epsilon}W)^{-1}) Q(d\bar{\omega}) = \int_{\mathcal{C}} \int_0^1 g(t, \dot{\bar{\omega}}(t)) dt Q(d\bar{\omega}) = \alpha_0^\mu(Q),$$

so taking limsup we conclude.

We proceed to show (i). We take $Q_\epsilon \rightarrow Q$ and assume without loss of generality that $\tilde{\alpha}_\epsilon^\mu(Q_\epsilon) < \infty$. By the duality formula in Lemma 7.6, and by Proposition 7.7, we have for any $F \in C_b(\mathcal{C})$:

$$\begin{aligned} \liminf_{\epsilon \downarrow 0} \tilde{\alpha}_\epsilon^\mu(Q_\epsilon) &\geq \liminf_{\epsilon \downarrow 0} \{ \mathbb{E}^{Q_\epsilon}[F] - \tilde{\rho}_\epsilon^\mu(F) \} \\ &= \mathbb{E}^Q[F] - \int_{\mathbb{R}^d} \sup_{\omega \in \mathcal{C}_x} \left(F(\omega) - \int_0^1 g(t, \dot{\omega}(t)) dt \right) \mu(dx), \end{aligned}$$

where we recall the notation $\mathcal{C}_x := \{\omega \in \mathcal{C} : \omega(0) = x\}$. Now, the function

$$\mathbb{R}^d \ni x \mapsto \inf_{\omega \in \mathcal{C}_x} \left(\int_0^1 g(t, \dot{\omega}(t)) dt - F(\omega) \right),$$

is the pointwise supremum of all functions h satisfying $h(x) + F(\omega) \leq \int_0^1 g(t, \dot{\omega}(t)) dt + \Psi_{\omega(0)}(x)$ for all $x \in \mathbb{R}^d$ and all $\omega \in \mathcal{C}$, where we define $\Psi_a(x) = +\infty$ if $x \neq a$ and $\Psi_a(x) = 0$ otherwise. Hence we have

$$\begin{aligned} &\liminf_{\epsilon \downarrow 0} \tilde{\alpha}_\epsilon^\mu(Q_\epsilon) \\ &\geq \sup_{\substack{F \in C_b(\mathcal{C}) \\ h \in L^1(\mathbb{R}^d, \mu)}} \left\{ \mathbb{E}^Q[F] + \int h d\mu : h(x) + F(\omega) \leq \int_0^1 g(t, \dot{\omega}(t)) dt + \Psi_{\omega(0)}(x), \forall x, \omega \right\}. \end{aligned}$$

By Kantorovich duality [43], Theorem 1.3, the right-hand side is equal to

$$\inf_{\pi} \int_{\mathcal{C} \times \mathbb{R}^d} \left(\int_0^1 g(t, \dot{\omega}(t)) dt + \Psi_{\omega(0)}(x) \right) \pi(d\omega, dx),$$

where the infimum is over all $\pi \in \mathcal{P}(\mathcal{C} \times \mathbb{R}^d)$ with first marginal Q and second marginal μ . Unless $\mu = Q^0$, this quantity is clearly infinite, and it is then straightforward to check that the entire expression reduces to $\alpha_0^\mu(Q)$. \square

As a final preparation for the proof of Theorem 2.8, we need the following compactness lemma.

LEMMA 8.2. *The family $\{\tilde{\alpha}_\epsilon^\mu : \epsilon \leq 1\}$ is equicoercive, namely:*

$$\bigcup_{\epsilon \leq 1} \{\tilde{\alpha}_\epsilon^\mu \leq c\} \text{ is tight for each } c \in \mathbb{R}.$$

PROOF. This is the same argument as in the *inf-tightness* part of the proof of Lemma 4.3, which we provide in the Appendix below. The point is that the initial distribution of the canonical process is independent of ϵ , its quadratic variation is uniformly bounded in ϵ , and its drift is bounded in L^1 independently of ϵ thanks to Assumption (TI) and the conditions $\tilde{\alpha}_\epsilon^\mu \leq c$. \square

PROOF OF THEOREM 2.8. With the notation we have built up, equality (19) is equivalent to

$$\liminf_{\epsilon \downarrow 0} \{\tilde{\alpha}_\epsilon^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), H(Q) = \nu_\epsilon\} = \inf\{\alpha_0^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), H(Q) = \nu\}.$$

We begin by proving the upper bound,

$$\limsup_{\epsilon \downarrow 0} \inf\{\tilde{\alpha}_\epsilon^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), Q \circ H^{-1} = \nu_\epsilon\} \leq \inf\{\alpha_0^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), Q \circ H^{-1} = \nu\}.$$

If there is no $Q \in \mathcal{P}(\mathcal{C})$ with $Q \circ H^{-1} = \nu$ the right-hand side is $+\infty$. Otherwise, for each $Q \in \mathcal{P}(\mathcal{C})$ with $Q \circ H^{-1} = \nu$ we introduce $\tilde{Q}_\epsilon := Q * P_\epsilon$ as in Lemma 8.1. By linearity of H we have

$$\tilde{Q}_\epsilon \circ H^{-1} = (Q \circ H^{-1}) * (P_\epsilon \circ H^{-1}) = \nu_\epsilon.$$

By Lemma 8.1, for each $Q \in \mathcal{P}(\mathcal{C})$ we have

$$\limsup_{\epsilon \downarrow 0} \inf\{\tilde{\alpha}_\epsilon^\mu(Q') : Q' \in \mathcal{P}(\mathcal{C}), Q' \circ H^{-1} = \nu_\epsilon\} \leq \limsup_{\epsilon \downarrow 0} \tilde{\alpha}_\epsilon^\mu(\tilde{Q}_\epsilon) \leq \alpha_0^\mu(Q).$$

Infimize over $Q \in \mathcal{P}(\mathcal{C})$ satisfying $Q \circ H^{-1} = \nu$ to get the announced upper bound.

It remains to prove the lower bound,

$$\liminf_{\epsilon \downarrow 0} \inf\{\tilde{\alpha}_\epsilon^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), Q \circ H^{-1} = \nu_\epsilon\} \geq \inf\{\alpha_0^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), Q \circ H^{-1} = \nu\}.$$

If the left-hand side is infinite there is nothing to prove. Otherwise, there exist sequences $\epsilon_n \downarrow 0$ and $Q_n \in \mathcal{P}(\mathcal{C})$ with $Q_n \circ H^{-1} = \nu_{\epsilon_n}$ such that

$$\lim_{n \rightarrow \infty} \tilde{\alpha}_{\epsilon_n}^\mu(Q_n) = \liminf_{\epsilon \downarrow 0} \inf\{\tilde{\alpha}_\epsilon^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}), Q \circ H^{-1} = \nu_\epsilon\}$$

and also $\sup_n \tilde{\alpha}_{\epsilon_n}^\mu(Q_n) < \infty$. The latter property along with Lemma 8.2 ensures that we may pass to a further subsequence and assume that $Q_n \rightarrow Q$ for some $Q \in \mathcal{P}(\mathcal{C})$. Continuity of H implies $Q \circ H^{-1} = \lim_n Q_n \circ H^{-1} = \lim_n \nu_{\epsilon_n} = \nu$. Moreover, by Lemma 8.1, we have $\liminf_{n \rightarrow \infty} \tilde{\alpha}_{\epsilon_n}^\mu(Q_n) \geq \alpha_0^\mu(Q)$, and we deduce the aforementioned lower bound.

That the problems in (20) admit an optimizer, provided there exists a feasible element, follows from the compactness of the sub-level sets of $\tilde{\alpha}_\epsilon^\mu$ (see Lemma 7.6), since the constraint $Q \circ H^{-1} = \nu_\epsilon$ is closed under weak convergence of measures. The analogous result for (21) follows taking $\epsilon = 0$.

If an optimizer for Q_ϵ exists for all $\epsilon > 0$, and if \bar{Q} is an accumulation point of $\{Q_\epsilon\}_\epsilon$, then \bar{Q} must be feasible for (21). Thus there exists Q an optimizer for (21), or equivalently for

$$\inf\{\alpha_0^\mu(Q) : Q \in \mathcal{P}(\mathcal{C}) \text{ with } Q \circ H^{-1} = \nu\}.$$

Defining \tilde{Q}_ϵ as in Lemma 8.1, we have

$$\alpha^\mu(Q) = \lim \tilde{\alpha}_\epsilon^\mu(\tilde{Q}_\epsilon) \geq \liminf \tilde{\alpha}_\epsilon^\mu(Q_\epsilon) \geq \alpha_0^\mu(\bar{Q}),$$

by Lemma 8.1. So \bar{Q} is optimal for (21) as desired. \square

We now proceed to the proof of Theorem 2.9. From here on, we take

$$Z = \mathbb{R}^d \quad \text{and} \quad H(\omega) = \omega(1),$$

so we are in the classical situation. We will make use of a technical estimate for Brownian bridges. We denote by

$$P_\epsilon^{x,y}[a, b] \in \mathcal{P}(C([a, b]; \mathbb{R}^d)),$$

the Brownian bridge from “ x at time a to y at time b ” with instantaneous variance ϵ . This is the law, on the space of continuous functions on $[a, b]$, of Brownian motion with volatility ϵ conditioned to start in x and end in y . We refer to [40], Theorem 40.3, for a characterization of (multidimensional) Brownian bridges.

LEMMA 8.3. *Let $a < b$. The canonical process admits under $P_\epsilon^{x,y}[a, b]$ the decomposition*

$$W(t) = x + \int_a^t \frac{y - W(s)}{b - s} ds + \sqrt{\epsilon} B(t),$$

where B is a standard d -dimensional Brownian motion on $[a, b]$. Then, with g and $r \in (1, 2)$ as in Theorem 2.9, we have

$$\begin{aligned} (73) \quad & \mathbb{E}^{P_\epsilon^{x,y}[a,b]} \left[\int_a^b g \left(\frac{y - W(t)}{b - t} \right) dt \right] \\ & \leq K_g \left((b - a)g \left(\frac{y - x}{2(b - a)} \right) + b - a + (b - a)^{1-(r/2)} \epsilon^{r/2} \right), \end{aligned}$$

where $K_g < \infty$ is a constant depending only on g and r .

PROOF. The claimed decomposition is classical [40], Theorem 40.3. To prove (73), it suffices to consider the interval $[0, b - a]$ rather than $[a, b]$. Let $\delta = b - a$. By conditioning of Gaussian distributions, we know that $W(t)$ is Gaussian with mean $\frac{\delta - t}{\delta}x + \frac{t}{\delta}y$ and variance matrix $\frac{\epsilon}{\delta}t(\delta - t)\text{Id}$, for each $t \in (0, \delta)$, under $P_\epsilon^{x,y}[0, \delta]$. Note also that g is convex and satisfies $g(q) \leq c(1 + |q|^r)$ for some $c > 0$. From this, denoting $P_1^1 = \mathcal{N}(0, \text{Id})$, we get

$$\begin{aligned} & \mathbb{E}^{P_\epsilon^{x,y}[0,\delta]} \left[\int_0^\delta g \left(\frac{y - W(t)}{\delta - t} \right) dt \right] \\ & = \int_0^\delta \int_{\mathbb{R}^d} g \left(\frac{y - x}{\delta} + z \sqrt{\frac{\epsilon t}{\delta(\delta - t)}} \right) dP_1^1(z) dt \\ & \leq \frac{1}{2} \delta g \left(2 \frac{y - x}{\delta} \right) + \frac{1}{2} \int_0^\delta \int_{\mathbb{R}^d} g \left(2z \sqrt{\frac{\epsilon t}{\delta(\delta - t)}} \right) dP_1^1(z) dt \\ & \leq \frac{1}{2} \delta g \left(2 \frac{y - x}{\delta} \right) + \frac{c\delta}{2} + 2^{r-1}c \int_0^\delta \int_{\mathbb{R}^d} \left| z \sqrt{\frac{\epsilon t}{\delta(\delta - t)}} \right|^r dP_1^1(z) dt \\ & = \frac{1}{2} \delta g \left(2 \frac{y - x}{\delta} \right) + \frac{c\delta}{2} + cK_r \delta^{1-(r/2)} \epsilon^{r/2}, \end{aligned}$$

where $K_r = 2^{r-1} \int_{\mathbb{R}^d} |z|^r dP_1^1(z) \int_0^1 (\frac{t}{1-t})^{r/2} dt$. Note that $K_r < \infty$ for $1 < r < 2$. Finally, to correct the factor of 2 within g we simply apply the assumed Δ_2 doubling property, or rather its consequence: $g(2q) \leq C(1 + g(q/2))$ for all $q \in \mathbb{R}^d$, for some $C > 0$. \square

Note that the proof of Lemma 8.3 reveals why we need to assume $r < 2$ in Theorem 2.9. Indeed, for $r \geq 2$, the integral $\int_0^1 (\frac{t}{1-t})^{r/2} dt$ is infinite.

PROOF OF THEOREM 2.9. Because of Lemma 8.1(i), the lower bound can be established exactly as in the proof of Theorem 2.8. The delicate point is proving the upper bound

$$(74) \quad \limsup_{\epsilon \downarrow 0} \inf \{ \tilde{\alpha}_\epsilon^\mu(Q) : Q \in \mathcal{P}(C), Q^1 = \nu \} \leq \inf \{ \alpha_0^\mu(Q) : Q \in \mathcal{P}(C), Q^1 = \nu \},$$

for which we cannot rely on Lemma 8.1(ii) as we did in the proof of Theorem 2.8, because we are working now with ν instead of ν_ϵ on the left-hand side. If the right-hand side is infinite there is nothing to prove. Let us take any Q with $\alpha_0^\mu(Q) < \infty$ and $Q^1 = \nu$. We introduce the measures

$$\pi_{s,t} = Q \circ (W(s), W(t))^{-1} \quad \text{and} \quad \pi_{\epsilon,(s,t)} := \pi_{s,t} * (P_\epsilon^S \otimes \delta_0).$$

That is, $\pi_{\epsilon,(s,t)} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is the joint law of $(X(s) + \sqrt{s}\epsilon Z, X(t))$, where $X \sim Q$ and Z is an independent standard d -dimensional Gaussian. The goal is to define now \tilde{Q}_ϵ satisfying the statement in Lemma 8.1(ii), but with $\tilde{Q}_\epsilon^1 = \nu$ (and of course $\tilde{Q}_\epsilon^0 = \mu$).

Let $\delta < 1$, which we will later send to zero. We will define first $\tilde{Q}_{\epsilon,\delta}$ by convolution of Q and P_ϵ in the time interval $[0, 1 - \delta]$, and we then steer toward the appropriate marginal ν at time 1 by using a suitable mixture of Brownian bridges. Concretely, we define $\tilde{Q}_{\epsilon,\delta}$ uniquely by the four properties:

- (1) $\tilde{Q}_{\epsilon,\delta} \circ (\{W(t)\}_{t \leq 1-\delta})^{-1} = (Q * P_\epsilon) \circ (\{W(t)\}_{t \leq 1-\delta})^{-1}$,
- (2) $\tilde{Q}_{\epsilon,\delta} \circ (W(1 - \delta), W(1))^{-1} = \pi_{\epsilon,(1-\delta,1)}$,
- (3) $\tilde{Q}_{\epsilon,\delta}(W(1) \in \cdot \mid \{W(t)\}_{t \leq 1-\delta}) = \tilde{Q}_{\epsilon,\delta}(W(1) \in \cdot \mid W(1 - \delta))$, a.s.,
- (4) $\tilde{Q}_{\epsilon,\delta}(\{W(t)\}_{t \in [1-\delta,1]} \mid W(1), \{W(t)\}_{t \leq 1-\delta}) = P_\epsilon^{W(1-\delta), W(1)}[1 - \delta, 1]$, a.s.

We remark that $\tilde{Q}_{\epsilon,\delta}$ is a semimartingale law for which the martingale part is $\sqrt{\epsilon}$ times a Brownian motion and, crucially, for which the time-0 and time-1 marginals are, respectively,

$$\tilde{Q}_{\epsilon,\delta}^0 = \mu \quad \text{and} \quad \tilde{Q}_{\epsilon,\delta}^1 = \nu.$$

Because $Q_{\epsilon,\delta} = Q * P_\epsilon$ on \mathcal{F}_t , we also have

$$(75) \quad \tilde{\alpha}_\epsilon^\mu(\tilde{Q}_{\epsilon,\delta}) = \mathbb{E}^{Q * P_\epsilon} \left[\int_0^{1-\delta} g(q^{Q * P_\epsilon}(t)) dt \right] + A_{\epsilon,[1-\delta,1]}^\mu,$$

where (recalling the semimartingale decomposition of $P_\epsilon^{x,y}[1 - \delta, 1]$ stated in Lemma 8.3)

$$A_{\epsilon,[1-\delta,1]}^\mu := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}^{P_\epsilon^{x,y}[1-\delta,1]} \left[\int_{1-\delta}^1 g\left(\frac{y - W(t)}{1 - t}\right) dt \right] \pi_{\epsilon,(1-\delta,1)}(dx, dy).$$

We now use Lemma 8.3, with the constant K_g introduced therein, to bound $A_{\epsilon,[1-\delta,1]}^\mu$:

$$(76) \quad \begin{aligned} A_{\epsilon,[1-\delta,1]}^\mu &\leq K_g \delta \mathbb{E} Q \left[g\left(\frac{W(1) - W(1 - \delta) + \sqrt{(1 - \delta)\epsilon} Z}{2\delta}\right) \right] \\ &\quad + K_g \delta + K_g \delta^{1-(r/2)} \epsilon^{r/2}, \end{aligned}$$

where Z denotes a standard Gaussian, independent of W . To bound the first term, use convexity of g and Jensen’s inequality to get

$$\begin{aligned} & \delta \mathbb{E}^Q \left[g \left(\frac{W(1) - W(1 - \delta) + \sqrt{(1 - \delta)\epsilon} Z}{2\delta} \right) \right] \\ & \leq \frac{1}{2} \delta \mathbb{E}^Q \left[g \left(\frac{W(1) - W(1 - \delta)}{\delta} \right) \right] + \frac{1}{2} \delta \mathbb{E}^Q \left[g \left(\frac{\sqrt{(1 - \delta)\epsilon} Z}{\delta} \right) \right] \\ & \leq \frac{1}{2} \mathbb{E}^Q \left[\int_{1-\delta}^1 g(\dot{W}(t)) dt \right] + C\delta + C\delta^{1-r} ((1 - \delta)\epsilon)^{r/2} \mathbb{E}^Q[|Z|^r], \end{aligned}$$

with the last line using the assumption (22) in the form $g(q) \leq 2C(1 + |q|^r)$ for some $C > 0$ (note that g is convex and thus locally bounded). The first term vanishes as $\delta \rightarrow 0$ because $\mathbb{E}^Q[\int_0^1 g(\dot{W}(t)) dt] = \alpha_0^\mu(Q)$ was assumed to be finite. The final term vanishes if we take $\delta = \sqrt{\epsilon}$, as does the term $\delta^{1-(r/2)}\epsilon^{r/2}$ in (76). We conclude from (76) that $A_{\epsilon, [1-\sqrt{\epsilon}, 1]}^\mu \rightarrow 0$.

Let us finally define $\tilde{Q}_\epsilon := \tilde{Q}_{\epsilon, \sqrt{\epsilon}}$. By dominated convergence $\tilde{Q}_\epsilon \rightarrow Q$. Recalling equation (75) and that $A_{\epsilon, [1-\sqrt{\epsilon}, 1]}^\mu \rightarrow 0$, the proof of (74) would be concluded if we can show that

$$\limsup_{\epsilon \rightarrow 0} \mathbb{E}^{Q * P_\epsilon} \left[\int_0^{1-\sqrt{\epsilon}} g(q^{Q * P_\epsilon}(t)) dt \right] \leq \alpha_0^\mu(Q).$$

Let us call (X, Y) the canonical process on $\mathcal{C}_0 \times \mathcal{C}_0$ equipped with the reference measure $Q \otimes P_\epsilon$. Of course $Q * P_\epsilon = Q \otimes P_\epsilon \circ (X + Y)^{-1}$ and X has absolutely continuous trajectories. We next claim that

$$(77) \quad q^{Q * P_\epsilon}(t, \omega) = \mathbb{E}^{Q \otimes P_\epsilon} [\dot{X}(t) \mid \{X(s) + Y(s)\}_{s \leq t} = \{\omega(s)\}_{s \leq t}], \quad Q * P_\epsilon\text{-a.e. } \omega.$$

Indeed, this is a consequence of the following well-known fact [40], Theorem VI.8.4: Suppose a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ supports an \mathbb{F} -Brownian motion B as well as \mathbb{F} -progressively measurable processes b and Z such that $\mathbb{E} \int_0^1 |b(t)| dt < \infty$. Suppose also that $dZ(t) = b(t) dt + \sqrt{\epsilon} dB(t)$. If \mathbb{F}^Z denotes the complete filtration generated by Z , then there exists an \mathbb{F}^Z -Brownian motion such that $dZ(t) = \hat{b}(t, Z) dt + \sqrt{\epsilon} d\hat{B}(t)$, where \hat{b} is the optional projection of b onto \mathbb{F}^Z , that is, $\hat{b}(t, \omega) = \mathbb{E}[b(t) \mid \{Z(s)\}_{s \leq t} = \{\omega(s)\}_{s \leq t}]$. In our setting, the requisite integrability of $b(t) := \dot{X}(t)$ under $Q \otimes P_\epsilon$ follows from the assumption $\alpha_0^\mu(Q) < \infty$ and the growth assumption in (TI).

Finally, using (77), Jensen’s inequality, and dominated convergence, we conclude

$$\limsup_{\epsilon \rightarrow 0} \mathbb{E}^{Q * P_\epsilon} \left[\int_0^{1-\sqrt{\epsilon}} g(q^{Q * P_\epsilon}(t)) dt \right] \leq \limsup_{\epsilon \rightarrow 0} \mathbb{E}^{Q \otimes P_\epsilon} \left[\int_0^{1-\sqrt{\epsilon}} g(\dot{X}(t)) dt \right] = \alpha_0^\mu(Q). \quad \square$$

APPENDIX: PROOFS OF PROPERTIES OF α^g

We collect here the belated proofs of some technical results.

LEMMA A.1. *Suppose (q_n) is a sequence of $L^1([0, 1])$ -valued random variables possibly defined on different probability spaces. Let $A_n(t) := \int_0^t q_n(s) ds$ and suppose there exists $a > 0$ such that, for each n ,*

$$(78) \quad \mathbb{E} \int_0^t g(t, q_n(t)) dt \leq a.$$

Then there exist a continuous process A , a subsequence A_{n_k} which converges in law in \mathcal{C} to A , and q such that

$$(79) \quad \mathbb{E} \int_0^t g(t, q(t)) dt \leq \liminf_{k \rightarrow \infty} \mathbb{E} \int_0^t g(t, q_{n_k}(t)) dt$$

and $A(t) = \int_0^t q(s) ds$. In particular, (A_n) is tight.

PROOF. We first check tightness. By Assumption (TI), for each $c > 0$ we may find $N > 0$ such that $g(t, q) \geq c|q|$ whenever $|q| \geq N$. Moreover, there exists $b \geq 0$ such that $g(t, q) \geq -b$ for all (t, q) . In particular, for all (t, q) we have $|q| \leq N + \frac{1}{c}(g(t, q) + b)$. Hence, for $0 \leq s < t \leq 1$,

$$\begin{aligned} |A_n(t) - A_n(s)| &\leq \int_s^t |q_n(u)| du \leq \frac{1}{c} \int_s^t (g(u, q(u)) + b) du + N(t - s) \\ &\leq \frac{1}{c} \int_0^1 g(u, q(u)) du + \frac{b}{c} + N(t - s). \end{aligned}$$

Hence, for any $\delta_n \downarrow 0$, (78) yields

$$\limsup_{n \rightarrow \infty} \sup_{\tau} \mathbb{E}|A_n(\tau + \delta_n) - A_n(\tau)| \leq \limsup_{n \rightarrow \infty} \left(\frac{a + b}{c} + N\delta_n \right) = \frac{a + b}{c},$$

where the \sup_{τ} is over all stopping times with values in $[0, 1 - \delta_n]$. As $c > 0$ was arbitrary, this shows that

$$\lim_{n \rightarrow \infty} \sup_{\tau} \mathbb{E}|A_n(\tau + \delta_n) - A_n(\tau)| = 0,$$

and from Aldous' criterion for tightness [20], Theorem 16.11, we conclude that (A_n) is tight.

Passing to a subsequence and applying Skorokhod's representation, let us now assume that there exists a continuous process A such that $A_n \rightarrow A$ almost surely in \mathcal{C} , with all processes defined on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. From (78), assumption (TI), and the criterion of de la Vallée Poisson, we conclude that $\{q_n : n \in \mathbb{N}\} \subset L^1 := L^1([0, 1] \times \Omega, dt \otimes d\mathbb{P})$ is uniformly integrable and thus weakly precompact. By passing to a further subsequence, we may now assume that $q_n \rightarrow q$ weakly in L^1 . Because g is bounded from below and lower semicontinuous in its second variable, the map $q \mapsto \mathbb{E} \int_0^1 g(t, q(t)) dt$ is lower semicontinuous in the norm topology of $L^1([0, 1] \times \Omega)$ by Fatou's lemma. Because it is also convex, this map is therefore weakly lower semicontinuous on L^1 . This yields (79). Lastly, by dominated convergence, it holds for each bounded random variable Z that

$$\mathbb{E}[ZA(t)] = \lim_{n \rightarrow \infty} \mathbb{E}[ZA_n(t)] = \lim_{n \rightarrow \infty} \mathbb{E} \left[Z \int_0^t q_n(s) ds \right] = \mathbb{E} \left[Z \int_0^t q(s) ds \right].$$

Hence $A(t) = \int_0^t q(s) ds$ a.s. for each t , and by continuity we have $A = \int_0^{\cdot} q(s) ds$ a.s. \square

PROOF OF LEMMA 4.3. *Convexity:* Let $\lambda \in [0, 1]$, and fix $Q_0, Q_1 \in \mathcal{P}^*$. We work on an extended probability space $\mathcal{C} \times \{0, 1\}$, and we write (W, X) to denote the identity map on this space. We define a measure M on $\mathcal{C} \times \{0, 1\}$ by requiring that the second marginal of M be $\lambda\delta_0 + (1 - \lambda)\delta_1$, and the conditional law of W given X be Q_X . In particular, the first marginal of M is precisely $Q := \lambda Q_0 + (1 - \lambda)Q_1$. Abbreviate $q_i := q^{Q_i}$. It easily follows that the process

$$W(t) - \int_0^t q_X(s) ds$$

defines an M -Brownian motion with respect to the filtration $\bar{\mathbb{F}} = (\bar{\mathcal{F}}_t)_{t \in [0,1]}$ defined by $\bar{\mathcal{F}}_t = \mathcal{F}_t \otimes \sigma(X)$ on the product space. Now define the process $q = (q(t))_{t \in [0,1]}$ on $\mathcal{C} \times \{0, 1\}$ to be the optional projection of the process $(q_X(t))_{t \in [0,1]}$ on the filtration generated by W . In particular,

$$q(t) = \mathbb{E}^M[q_X(t) \mid (W_s)_{s \leq t}] = \mathbb{E}^M[\mathbf{1}_{\{X=0\}}q_0(t) + \mathbf{1}_{\{X=1\}}q_1(t) \mid (W_s)_{s \leq t}].$$

The process $W - \int_0^\cdot q(t) dt$ is then a Brownian motion on $(\mathcal{C}, \mathbb{F}, Q)$, where we recall that Q is the first marginal of M (e.g., by [40], Theorem VI.8.4, as in the end of the proof of Theorem 2.9). It follows that $Q \in \mathcal{P}^*$ and $q = q^Q$. Finally, using Jensen’s inequality, we compute

$$\begin{aligned} \lambda \tilde{\alpha}^g(Q_0) + (1 - \lambda) \tilde{\alpha}^g(Q_1) &= \lambda \mathbb{E}^{Q_0} \left[\int_0^1 g(t, q_0(t)) dt \right] + (1 - \lambda) \mathbb{E}^{Q_1} \left[\int_0^1 g(t, q_1(t)) dt \right] \\ &= \mathbb{E}^M \left[\int_0^1 g(t, q_X(t)) dt \right] \\ &\geq \mathbb{E}^M \left[\int_0^1 g(t, q(t)) dt \right] = \mathbb{E}^Q \left[\int_0^1 g(t, q(t)) dt \right] \\ &= \tilde{\alpha}^g(Q). \end{aligned}$$

Inf-compactness: Let $a \in \mathbb{R}$ and $\Lambda_a := \{Q : \tilde{\alpha}^g(Q) \leq a\}$. It is convenient in this step and the next to define

$$W^Q(t) := W(t) - \int_0^t q^Q(s) ds, \quad t \in [0, 1],$$

for $Q \in \mathcal{P}^*$, recalling that W^Q is a Q -Brownian motion by definition of \mathcal{P}^* . Letting $A^Q(t) := \int_0^t q^Q(s) ds$, it follows from Lemma A.1 that $\{Q \circ (A^Q)^{-1} : Q \in \Lambda_a\} \subset \mathcal{P}(\mathcal{C})$ is tight. On the other hand, $\{Q \circ (W^Q)^{-1} : Q \in \Lambda_a\} = \{P\}$ is a singleton and thus tight. Since each marginal is tight, we deduce that $\{Q \circ (W^Q, A^Q)^{-1} : Q \in \Lambda_a\} \subset \mathcal{P}(\mathcal{C} \times \mathcal{C})$ is tight. Finally, by continuous mapping, the set $\{Q \circ (W^Q + A^Q)^{-1} : Q \in \Lambda_a\} = \Lambda_a$ is tight.

Lower semicontinuity: Suppose $\{Q_n : n \in \mathbb{N}\} \subset \Lambda_a$ with $Q_n \rightarrow Q$ weakly for some $Q \in \mathcal{P}(\mathcal{C})$. We must show that Q belongs to Λ_a . Define the continuous process

$$A_n(t) = \int_0^t q^{Q_n}(s) ds = W(t) - W^{Q_n}(t),$$

for each n . Since $Q_n \circ (W^{Q_n})^{-1}$ equals Wiener measure for each n , we conclude that $\{Q_n \circ (W, W^{Q_n})^{-1} : n \in \mathbb{N}\}$ is tight, and thus $\{Q_n \circ (W, W^{Q_n}, A_n)^{-1} : n \in \mathbb{N}\}$ is tight. Relabeling a subsequence, suppose that $Q_n \circ (W, W^{Q_n}, A_n)^{-1}$ converges weakly to the law of some \mathcal{C}^3 -valued random variable (X, B, A) . Using Lemma A.1, we may assume also that $A(t) = \int_0^t q(s) ds$ for some process q satisfying

$$\mathbb{E}^Q \int_0^1 g(t, q(t)) dt \leq \liminf \mathbb{E}^{Q_n} \int_0^1 g(t, q^{Q_n}(t)) dt \leq a.$$

Clearly, the law of B is Wiener measure. Moreover, $(W^{Q_n}(s) - W^{Q_n}(t))_{s \in [t,1]}$ is independent of $(W(s), W^{Q_n}(s), A^n(s))_{s \leq t}$ for each $t \in [0, 1]$, and thus $(B(s) - B(t))_{s \in [t,1]}$ is independent of $(X(s), B(s), A(s))_{s \leq t}$. In particular, B is a Brownian motion with respect to the filtration generated by X, B , and q . Finally, notice that

$$X(t) = B(t) + A(t) = B(t) + \int_0^t q(s) ds,$$

as the same relation holds in the pre-limit. A standard argument (see [39], Exercise (5.15)) shows that $X - \int_0^\cdot \widehat{q}(s) ds$ is a Brownian motion, where \widehat{q} is the optional projection of q onto the filtration generated by X . By convexity of $g(t, \cdot)$, we have

$$\mathbb{E}^Q \int_0^1 g(t, \widehat{q}(t)) dt \leq \mathbb{E}^Q \int_0^1 g(t, q(t)) dt \leq a.$$

Recalling that Q denoted the law of X , we conclude that $Q \in \mathcal{P}^*$ and thus $Q \in \Lambda_a$.

Reverse conjugacy: By definition

$$\tilde{\alpha}^g(Q) \geq \sup_{F \in B_b(C)} \{\mathbb{E}^Q[F] - \tilde{\rho}^g(F)\} \geq \sup_{F \in C_b(C)} \{\mathbb{E}^Q[F] - \tilde{\rho}^g(F)\}.$$

Recalling the previous results showing convexity and lower semicontinuity of $\tilde{\alpha}^g$, we may apply the Fenchel–Moreau theorem with respect to the dual pairing between $C_b(C)$ and the space of measures on C to get equality above. \square

We close by elaborating slightly on the dual representation of BSDE supersolutions, which was discussed to some extent on page 1326. In particular, the following slight adaptation of results of [11] was used in Lemma 5.1, which extended equation (12) to nonzero times t .

LEMMA A.2. *Let $F \in C_b(C)$. The minimal supersolution of the BSDE*

$$dY(t) = -g^*(t, Z(t)) dt + Z(t) dW_t, \quad Y(1) = F$$

admits the representation

$$Y(t) = \operatorname{ess\,sup}_{Q \in \mathcal{Q}_t} \mathbb{E}^Q \left[F(W) - \int_t^1 g(u, q^Q(u)) du \mid \mathcal{F}_t \right] \quad P\text{-a.s. for all } t \in [0, 1],$$

where \mathcal{Q}_t is the set of $Q \in \mathcal{Q}$ such that $Q = P$ on \mathcal{F}_t .

PROOF. Since $\mathcal{Q}_t \subseteq \mathcal{Q}$, “ \geq ” follows by [11], Theorem 3.4. Reciprocally, since by (the first part of the proof of) [11], Proposition 4.2, the set $\{\mathbb{E}_Q[F(W) - \int_t^1 g(u, q^Q(u)) du \mid \mathcal{F}_t] : Q \in \mathcal{Q}\}$ of random variables is directed, it holds

$$Y_t = \lim_{n \rightarrow \infty} \mathbb{E}^{Q^n} \left[F(W) - \int_t^1 g(u, q^{Q^n}(u)) du \mid \mathcal{F}_t \right]$$

for a sequence $Q^n \in \mathcal{Q}$. Put $q^n(u) := q^{Q^n}(u)1_{[t,1]}(u)$ and let \bar{Q}^n be such that $q^{\bar{Q}^n} = q^n$. Then, $\bar{Q}^n \in \mathcal{Q}_t$ and it follows from Bayes’ rule that

$$\begin{aligned} Y(t) &= \lim_{n \rightarrow \infty} \mathbb{E} \left[e^{\int_t^1 q^{Q^n}(u) dW(u) - \frac{1}{2} \int_t^1 |q^{Q^n}(u)|^2 du} \left(F(W) - \int_t^1 g(u, q^{Q^n}(u)) du \right) \mid \mathcal{F}_t \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[e^{\int_t^1 q^n(u) dW(u) - \frac{1}{2} \int_t^1 |q^n(u)|^2 du} \left(F(W) - \int_t^1 g(u, q^n(u)) du \right) \mid \mathcal{F}_t \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E}^{\bar{Q}^n} \left[F(W) - \int_t^1 g(u, q^{\bar{Q}^n}(u)) du \mid \mathcal{F}_t \right], \end{aligned}$$

which proves “ \leq ”. \square

REFERENCES

- [1] ANDERSSON, D. and DJEHICHE, B. (2011). A maximum principle for SDEs of mean-field type. *Appl. Math. Optim.* **63** 341–356. MR2784835 <https://doi.org/10.1007/s00245-010-9123-8>
- [2] BAYRAKTAR, E. and KELLER, C. (2018). Path-dependent Hamilton–Jacobi equations in infinite dimensions. *J. Funct. Anal.* **275** 2096–2161. MR3841538 <https://doi.org/10.1016/j.jfa.2018.07.010>
- [3] BERTSEKAS, D. P. and SHREVE, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case. Mathematics in Science and Engineering* **139**. Academic Press [Harcourt Brace Jovanovich, Publishers], New York–London. MR0511544
- [4] BORELL, C. (2000). Diffusion equations and geometric inequalities. *Potential Anal.* **12** 49–71. MR1745333 <https://doi.org/10.1023/A:1008641618547>
- [5] BOUÉ, M. and DUPUIS, P. (1998). A variational representation for certain functionals of Brownian motion. *Ann. Probab.* **26** 1641–1659. MR1675051 <https://doi.org/10.1214/aop/1022855876>
- [6] CARMONA, R. and DELARUE, F. (2015). Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. *Ann. Probab.* **43** 2647–2700. MR3395471 <https://doi.org/10.1214/14-AOP946>
- [7] CHASSAGNEUX, J.-F., CRISAN, D. and DELARUE, F. (2019). Numerical method for FBSDEs of McKean–Vlasov type. *Ann. Appl. Probab.* **29** 1640–1684. MR3914553 <https://doi.org/10.1214/18-AAP1429>
- [8] DELBAEN, F., HU, Y. and BAO, X. (2011). Backward SDEs with superquadratic growth. *Probab. Theory Related Fields* **150** 145–192. MR2800907 <https://doi.org/10.1007/s00440-010-0271-1>
- [9] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. Springer, New York. MR1619036 <https://doi.org/10.1007/978-1-4612-5320-4>
- [10] DRAPEAU, S., HEYNE, G. and KUPPER, M. (2013). Minimal supersolutions of convex BSDEs. *Ann. Probab.* **41** 3973–4001. MR3161467 <https://doi.org/10.1214/13-AOP834>
- [11] DRAPEAU, S., KUPPER, M., ROSAZZA GIANIN, E. and TANGPI, L. (2016). Dual representation of minimal supersolutions of convex BSDEs. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 868–887. MR3498013 <https://doi.org/10.1214/14-AIHP664>
- [12] DRAPEAU, S. and MAINBERGER, C. (2016). Stability and Markov property of forward backward minimal supersolutions. *Electron. J. Probab.* **21** Paper No. 41, 15. MR3515571 <https://doi.org/10.1214/16-EJP4276>
- [13] DUGUNDJI, J. (1951). An extension of Tietze’s theorem. *Pacific J. Math.* **1** 353–367. MR0044116
- [14] DUPUIS, P. and ELLIS, R. S. (2011). *A Weak Convergence Approach to the Theory of Large Deviations. Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. MR1431744 <https://doi.org/10.1002/9781118165904>
- [15] EKREN, I., KELLER, C., TOUZI, N. and ZHANG, J. (2014). On viscosity solutions of path dependent PDEs. *Ann. Probab.* **42** 204–236. MR3161485 <https://doi.org/10.1214/12-AOP788>
- [16] EVANS, L. C. (1998). *Partial Differential Equations. Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR1625845
- [17] FLEMING, W. H. (1977/78). Exit probabilities and optimal stochastic control. *Appl. Math. Optim.* **4** 329–346. MR0512217 <https://doi.org/10.1007/BF01442148>
- [18] FÖLLMER, H. (1988). Random fields and diffusion processes. In *École D’Été de Probabilités de Saint-Flour XV–XVII, 1985–87. Lecture Notes in Math.* **1362** 101–203. Springer, Berlin. MR0983373 <https://doi.org/10.1007/BFb0086180>
- [19] FÖLLMER, H. and SCHIED, A. (2011). *Stochastic Finance: An Introduction in Discrete Time*, extended ed. de Gruyter, Berlin. MR2779313 <https://doi.org/10.1515/9783110218053>
- [20] KALLENBERG, O. (2006). *Foundations of Modern Probability. Probability and Its Applications (New York)*. Springer, New York. MR1464694
- [21] KOBYLANSKI, M. (2000). Backward stochastic differential equations and partial differential equations with quadratic growth. *Ann. Probab.* **28** 558–602. MR1782267 <https://doi.org/10.1214/aop/1019160253>
- [22] KUPPER, M. and SCHACHERMAYER, W. (2009). Representation results for law invariant time consistent functions. *Math. Financ. Econ.* **2** 189–210. MR2540511 <https://doi.org/10.1007/s11579-009-0019-9>
- [23] LACKER, D. (2017). A non-exponential extension of Sanov’s theorem via convex duality. Preprint.
- [24] LACKER, D. (2017). Limit theory for controlled McKean–Vlasov dynamics. *SIAM J. Control Optim.* **55** 1641–1672. MR3654119 <https://doi.org/10.1137/16M1095895>
- [25] LAURIÈRE, M. and PIRONNEAU, O. (2014). Dynamic programming for mean-field type control. *C. R. Math. Acad. Sci. Paris* **352** 707–713. MR3258261 <https://doi.org/10.1016/j.crma.2014.07.008>
- [26] LEHEC, J. (2013). Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 885–899. MR3112438 <https://doi.org/10.1214/11-aihp464>

- [27] LÉONARD, C. (2012). From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** 1879–1920. MR2873864 <https://doi.org/10.1016/j.jfa.2011.11.026>
- [28] LÉONARD, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. MR3121631 <https://doi.org/10.3934/dcds.2014.34.1533>
- [29] LÉONARD, C. (2016). Lazy random walks and optimal transport on graphs. *Ann. Probab.* **44** 1864–1915. MR3502596 <https://doi.org/10.1214/15-AOP1012>
- [30] LUKOYANOV, N. Y. (2007). Hamilton–Jacobi type equations for hereditary systems. *Proc. Steklov Inst. Math.* **259** S190–S200.
- [31] MA, J., REN, Z., TOUZI, N. and ZHANG, J. (2016). Large deviations for non-Markovian diffusions and a path-dependent Eikonal equation. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1196–1216. MR3531706 <https://doi.org/10.1214/15-AIHP678>
- [32] MIKAMI, T. (2004). Monge’s problem with a quadratic cost by the zero-noise limit of h -path processes. *Probab. Theory Related Fields* **129** 245–260. MR2063377 <https://doi.org/10.1007/s00440-004-0340-4>
- [33] MIKAMI, T. and THIEULLEN, M. (2008). Optimal transportation problem by stochastic optimal control. *SIAM J. Control Optim.* **47** 1127–1139. MR2407010 <https://doi.org/10.1137/050631264>
- [34] PAPANTOLEON, A., POSSAMAÏ, D. and SAPLAOURAS, A. (2019). Stability results for martingale representations: The general case. *Trans. Amer. Math. Soc.* **372** 5891–5946. MR4014298 <https://doi.org/10.1090/tran/7880>
- [35] PARDOUX, É. and PENG, S. (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic Partial Differential Equations and Their Applications (Charlotte, NC, 1991)*. *Lect. Notes Control Inf. Sci.* **176** 200–217. Springer, Berlin. MR1176785 <https://doi.org/10.1007/BFb0007334>
- [36] PARDOUX, E. and TANG, S. (1999). Forward–backward stochastic differential equations and quasilinear parabolic PDEs. *Probab. Theory Related Fields* **114** 123–150. MR1701517 <https://doi.org/10.1007/s004409970001>
- [37] PHAM, H. and WEI, X. (2018). Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM Control Optim. Calc. Var.* **24** 437–461. MR3843191 <https://doi.org/10.1051/cocv/2017019>
- [38] RAINERO, S. (2006). Un principe de grandes déviations pour une équation différentielle stochastique progressive rétrograde. *C. R. Math. Acad. Sci. Paris* **343** 141–144. MR2243309 <https://doi.org/10.1016/j.crma.2006.05.022>
- [39] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Springer, Berlin. MR1725357 <https://doi.org/10.1007/978-3-662-06400-9>
- [40] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol. 2: Itô Calculus*. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. MR1780932 <https://doi.org/10.1017/CBO9781107590120>
- [41] SAPLAOURAS, A. (2017). Backward stochastic differential equations with jumps are stable. Ph.D. thesis, Technische Universität Berlin.
- [42] VAN HANDEL, R. (2018). The Borell–Ehrhard game. *Probab. Theory Related Fields* **170** 555–585. MR3773794 <https://doi.org/10.1007/s00440-017-0762-4>
- [43] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 <https://doi.org/10.1090/gsm/058>
- [44] WANG, R., WANG, X. and WU, L. (2010). Sanov’s theorem in the Wasserstein distance: A necessary and sufficient condition. *Statist. Probab. Lett.* **80** 505–512. MR2593592 <https://doi.org/10.1016/j.spl.2009.12.003>
- [45] ZHANG, J. (2017). *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*. *Probability Theory and Stochastic Modelling* **86**. Springer, New York. MR3699487 <https://doi.org/10.1007/978-1-4939-7256-2>